



Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Pós-Graduação em Ciência da Computação

Tese de Doutorado

Características de Personalidade Identificadas  
via Textos aplicadas à Recomendação por  
Filtragem Colaborativa

Janderson Jason Barbosa Aguiar

Campina Grande, Paraíba, Brasil

Dezembro/2021

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Pós-Graduação em Ciência da Computação

Características de Personalidade Identificadas  
via Textos aplicadas à Recomendação por  
Filtragem Colaborativa

Janderson Jason Barbosa Aguiar

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande (Campus I) como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Área de Concentração: Ciência da Computação  
Linha de Pesquisa: Metodologia e Técnicas da Computação

Dr.<sup>a</sup> Joseana Macêdo Fachine Régis de Araújo

Dr. Evandro de Barros Costa

(Orientadores)

Campina Grande, Paraíba, Brasil

©Janderson Jason Barbosa Aguiar, Dezembro/2021

A283c Aguiar, Janderson Jason Barbosa.  
Características de personalidade identificadas via textos aplicadas à recomendação por filtragem colaborativa / Janderson Jason Barbosa Aguiar. – Campina Grande, 2021.  
204 f. : il. color.

Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2021.

“Orientação: Profa. Dra. Joseana Macêdo Fachine Régis de Araújo, Prof. Dr. Evandro de Barros Costa”.

Referências.

1. Ciência da Computação. 2. Sistemas de Recomendação. 3. Personalidade. 4. Filtragem Colaborativa. 5. Metodologia e Técnicas da Computação. I. Araújo, Joseana Macêdo Fachine Régis de. II. Costa, Evandro de Barros. III. Título.

CDU 004(043)



MINISTÉRIO DA EDUCAÇÃO  
**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE**  
POS-GRADUACAO CIENCIAS DA COMPUTACAO  
Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

## **FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES**

**JANDERSON JASON BARBOSA AGUIAR**

CARACTERÍSTICAS DE PERSONALIDADE IDENTIFICADAS VIA TEXTOS APLICADAS À  
RECOMENDAÇÃO POR FILTRAGEM COLABORATIVA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Doutor em Ciência da Computação.

Aprovada em: 17/12/2021

Profa. Dra. JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, UFCG, Orientadora

Prof. Dr. EVANDRO DE BARROS COSTA, UFAL, Orientador

Prof. Dr. JOSÉ ANTÃO BELTRÃO MOURA, UFCG, Examinador Interno

Prof. Dr. NAZARENO FERREIRA DE ANDRADE, UFCG, Examinador Interno

Prof. Dr. THALES MIRANDA DE ALMEIDA VEIRA, UFAL, Examinador Externo

Prof. Dr. FABIANO AZEVEDO DORÇA, UFU, Examinador Externo



Documento assinado eletronicamente por **JOSEANA MACEDO FECHINE, PROFESSOR DO MAGISTERIO SUPERIOR**, em 17/12/2021, às 12:22, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **Evandro de Barros Costa, Usuário Externo**, em 17/12/2021, às 13:58, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **NAZARENO FERREIRA DE ANDRADE, PROFESSOR 3 GRAU**, em 17/12/2021, às 16:20, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **JOSE ANTAO BELTRAO MOURA, PROFESSOR 3 GRAU**, em 18/12/2021, às 11:29, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **Fabiano Azevedo Dorça, Usuário Externo**, em 26/01/2022, às 07:35, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **2008943** e o código CRC **B91125C2**.

*Dedico esta tese à minha família,  
que está sempre vibrando  
por minhas conquistas.*

## Agradecimentos

Primeiramente, agradeço a Deus por inúmeros aspectos — pelos momentos de inspiração, pelas oportunidades, por me ajudar a ter foco em momentos de desânimo, pela saúde física e mental, entre outras bênçãos.

À minha família, especialmente aos meus pais, por toda a educação, compreensão e paciência durante esses anos. Aos meus avós maternos (*in memoriam*), que, mesmo sem entenderem muito, torciam por meus estudos. Aos meus avós paternos, por sempre demonstrarem carinho e expectativa de ter um “neto doutor”.

À minha orientadora Joseana Fechine e ao meu coorientador Evandro Costa, agradeço imensamente pelas contribuições e, principalmente, pela paciência comigo ao longo desse tempo. Agradeço também aos membros da banca examinadora, pela disponibilidade para contribuir com a pesquisa realizada.

À Universidade Federal de Campina Grande (UFCG), que me acolheu nesse longo tempo, com funcionários que me ajudaram em muitos momentos (em especial, Paloma e Lyana), além dos muitos professores excelentes. Agradeço também à Universidade Estadual da Paraíba (UEPB), instituição onde iniciei minha formação em Computação, e à qual tive o prazer de voltar atuando como professor substituto.

Aos colegas de graduação e pós-graduação, aos alunos e aos amigos (muitos do EJC e das EJNS), pelos momentos de escuta e incentivo. À Tia Marlene (*in memoriam*), pelo seu exemplo de dedicação, tanto no âmbito escolar (querida professora do jardim da infância) quanto no âmbito religioso (querida catequista); ao mencioná-la, estendo minha gratidão aos valiosos educadores que tive desde criança.

Aos pesquisadores que contatei ao longo do desenvolvimento da pesquisa, pelas contribuições diretas ou indiretas.

Por fim, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro, e a todos que de alguma forma ajudaram na realização desta pesquisa.

*O encontro de duas personalidades  
assemelha-se ao contato de  
duas substâncias químicas:  
se alguma reação ocorre,  
ambos sofrem uma transformação.  
(Carl Jung)*

## Resumo

A quantidade de recursos digitais tende a aumentar com o crescente uso das tecnologias de informação e comunicação em diversos domínios, como comércio eletrônico, *e-learning* e turismo. Nesse contexto, há os Sistemas de Recomendação (SR), visando a, por exemplo, direcionar os recursos mais adequados aos usuários. Há pesquisadores que consideram a personalidade dos usuários ao elaborar estratégias para SR. Muitos deles partem da premissa “pessoas similares quanto a personalidade tendem a preferir itens similares” e aplicam a personalidade em SR via Filtragem Colaborativa (FC); porém, não foram encontradas pesquisas investigando se, com a colaboração enfocada na opinião de usuários com personalidade similar, a acurácia de quaisquer algoritmos de FC seria, no mínimo, mantida. Além disso, a maioria das pesquisas não emprega meios automáticos para reconhecimento de personalidade (APR — *Automatic Personality Recognition*) e desconsidera que os diferentes componentes representativos da personalidade podem ter diferentes influências no processo de recomendação. Ademais, essas pesquisas comumente se concentram em cenários específicos (a exemplo de situações de *cold-start*). Destarte, nesta tese, é apresentada uma investigação, em cenários gerais, com o objetivo de melhorar a acurácia em SR ao aplicar informações referentes à personalidade humana, obtidas sem o uso dos tradicionais questionários de personalidade. Com o estudo experimental realizado, (i) os resultados de estudos iniciais que utilizaram questionários para identificar a personalidade foram reforçados; (ii) verificou-se que a acurácia de um algoritmo de FC pode ser afetada quando o algoritmo desconsidera a opinião de usuários menos similares quanto à personalidade (tendendo a melhorar a acurácia de algoritmos baseados em memória); (iii) averiguou-se que os componentes empregados para definir a personalidade influenciam de modo diferente na acurácia das recomendações, embora não seja adequado focar em componentes específicos e ignorar outros; e (iv) foram propostas e avaliadas novas estratégias para SR aplicando a personalidade dos usuários. Considerando-se os resultados obtidos, esta tese destaca ser pertinente o uso de características de personalidade dos usuários em SR baseados em FC (sem a necessidade de preenchimento de questionários específicos para identificar tais características), especialmente em estratégias híbridas de recomendação, dado que, embora seja válido considerar a personalidade em um processo de recomendação personalizada, tal fator não é o único aspecto importante.

**Palavras-chave:** Sistemas de Recomendação; Personalidade; Filtragem Colaborativa.

## Abstract

The amount of digital resources tends to increase with the growing use of information and communication technologies in several domains, such as e-commerce, e-learning, and tourism. In this context, there are Recommender Systems (RS) aiming, for example, to direct the most appropriate resources to users. Some researchers consider the user's personality when designing RS strategies. Many of these researchers consider the premise that "people with similar personalities tend to prefer similar items", and apply personality in RS via Collaborative Filtering (CF). However, we have not found research investigating whether, with collaboration focused on the opinion of users with a similar personality, the accuracy of any CF algorithms would, at a minimum, be maintained. Besides, most researchers do not apply APR (Automatic Personality Recognition) and do not consider that the different personality components can influence the recommendation process in different ways. Furthermore, these researchers commonly focus on specific scenarios (such as cold-start situations). Therefore, this thesis presents an investigation to improve RS accuracy in general scenarios by applying information related to human personality obtained without traditional personality questionnaires. With the experimental study conducted, (i) we reinforced results from initial studies that used questionnaires to identify personality; (ii) we realized that the accuracy of a CF algorithm could be affected when the algorithm disregards the opinion of less similar users in terms of personality (tending to improve the accuracy of memory-based algorithms); (iii) we analyzed that the components used to define personality influence the recommendation accuracy differently, although it is inappropriate to focus on specific components and ignore others; and (iv) we proposed and evaluated new strategies for RS applying the personality of users. The results obtained in this thesis highlight the relevance of the use of personality characteristics of CF-based RS users (without the need to complete specific questionnaires to identify such characteristics), especially in hybrid recommendation strategies, since, although it is valid to consider personality in a personalized recommendation process, this factor is not the only important aspect.

**Keywords:** Recommender Systems; Personality; Collaborative Filtering.

# Lista de Figuras

2.1	Representação gráfica de um perfil de personalidade criado pelo <i>Watson Personality Insights</i> . . . . .	20
2.2	Classificação dos sistemas de recomendação baseados em personalidade. . .	27
4.1	Abordagens de recomendação consideradas na investigação referente à QP1.	51
4.2	Representação geral da abordagem elaborada referente à QP2 (“tps”). . . .	52
4.3	Representação geral da estratégia “itps”. . . . .	57
4.4	Representação geral da estratégia <i>weighted</i> . . . . .	58
4.5	Representação geral da estratégia <i>re-ranking</i> . . . . .	60
5.1	<i>Boxplots</i> referentes à QP1 com o conjunto de dados AmazonFoods. . . . .	69
5.2	<i>Boxplots</i> referentes à QP1 com o conjunto de dados AmazonFoodsCompleto.	70
5.3	<i>Boxplots</i> referentes à QP4 (estratégia “itps”) com o conjunto de dados TripAdvisor. . . . .	80
5.4	<i>Boxplots</i> referentes à QP4 (estratégia “itps”) com o conjunto de dados AmazonFoods. . . . .	81
5.5	<i>Boxplots</i> referentes à QP4 (estratégia “itps”) com o conjunto de dados AmazonFoodsCompleto. . . . .	82
5.6	<i>Boxplots</i> referentes à QP4 (estratégia <i>weighted</i> ) com os conjuntos de dados TripAdvisor e TripAdvisorCompleto. . . . .	85
5.7	<i>Boxplots</i> referentes à QP4 (estratégia <i>weighted</i> ) com os conjuntos de dados AmazonFoods e AmazonFoodsCompleto. . . . .	86
5.8	<i>Boxplots</i> referentes à QP4 (estratégia <i>re-ranking</i> ) com o conjunto de dados TripAdvisor. . . . .	89

---

5.9	<i>Boxplots</i> referentes à QP4 (estratégia <i>re-ranking</i> ) com o conjunto de dados TripAdvisorCompleto. . . . .	90
5.10	<i>Boxplots</i> referentes à QP4 (estratégia <i>re-ranking</i> ) com o conjunto de dados AmazonFoods. . . . .	91
5.11	<i>Boxplots</i> referentes à QP4 (estratégia <i>re-ranking</i> ) com o conjunto de dados AmazonFoodsCompleto. . . . .	92
5.12	<i>Boxplots</i> referentes à QP4 (estratégia <i>re-ranking</i> ) com o conjunto de dados AmazonFoods, destacando avaliação por <i>cluster</i> , com o traço Amabilidade (t4). . . . .	93
5.13	<i>Boxplots</i> referentes à QP4 (estratégia <i>re-ranking</i> ) com o conjunto de dados AmazonFoodsCompleto, destacando avaliação por <i>cluster</i> , com o traço Amabilidade (t4). . . . .	94
E.1	<i>Boxplots</i> referentes à QP2 (modelo <i>Big Five</i> ) com o conjunto de dados TripAdvisor. . . . .	164
E.2	<i>Boxplots</i> referentes à QP2 (modelo <i>Big Five</i> ) com o conjunto de dados TripAdvisorCompleto. . . . .	165
E.3	<i>Boxplots</i> referentes à QP2 (modelo <i>Big Five</i> ) com o conjunto de dados AmazonFoods. . . . .	166
E.4	<i>Boxplots</i> referentes à QP2 (modelo <i>Big Five</i> ) com o conjunto de dados AmazonFoodsCompleto. . . . .	167
E.5	<i>Boxplots</i> referentes à QP2 (modelo <i>Values</i> ) com o conjunto de dados TripAdvisor. . . . .	169
E.6	<i>Boxplots</i> referentes à QP2 (modelo <i>Values</i> ) com o conjunto de dados TripAdvisorCompleto. . . . .	170
E.7	<i>Boxplots</i> referentes à QP2 (modelo <i>Values</i> ) com o conjunto de dados AmazonFoods. . . . .	171
E.8	<i>Boxplots</i> referentes à QP2 (modelo <i>Values</i> ) com o conjunto de dados AmazonFoodsCompleto. . . . .	172
E.9	<i>Boxplots</i> referentes à QP2 (modelo <i>Needs</i> ) com o conjunto de dados TripAdvisor. . . . .	174

---

E.10	<i>Boxplots</i> referentes à QP2 (modelo <i>Needs</i> ) com o conjunto de dados TripAdvisorCompleto. . . . .	175
E.11	<i>Boxplots</i> referentes à QP2 (modelo <i>Needs</i> ) com o conjunto de dados AmazonFoods. . . . .	176
E.12	<i>Boxplots</i> referentes à QP2 (modelo <i>Needs</i> ) com o conjunto de dados AmazonFoodsCompleto. . . . .	177
F.1	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo UserKNN e o conjunto de dados TripAdvisorCompleto. . . . .	180
F.2	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo ItemKNN e o conjunto de dados TripAdvisorCompleto. . . . .	181
F.3	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SMF e o conjunto de dados TripAdvisorCompleto. . . . .	182
F.4	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo UserKNN e o conjunto de dados AmazonFoods. . . . .	183
F.5	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo ItemKNN e o conjunto de dados AmazonFoods. . . . .	184
F.6	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SMF e o conjunto de dados AmazonFoods. . . . .	185
F.7	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo BMF e o conjunto de dados AmazonFoods. . . . .	186
F.8	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SVD++ e o conjunto de dados AmazonFoods. . . . .	187
F.9	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo UserKNN e o conjunto de dados AmazonFoodsCompleto. . . . .	188
F.10	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo ItemKNN e o conjunto de dados AmazonFoodsCompleto. . . . .	189
F.11	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SMF e o conjunto de dados AmazonFoodsCompleto. . . . .	190
F.12	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo BMF e o conjunto de dados AmazonFoodsCompleto. . . . .	191

---

F.13	Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SVD++ e o conjunto de dados AmazonFoodsCompleto. . . . .	192
F.14	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo UserKNN e o conjunto de dados AmazonFoods. . . . .	193
F.15	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo ItemKNN e o conjunto de dados AmazonFoods. . . . .	194
F.16	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SMF e o conjunto de dados AmazonFoods. . . . .	195
F.17	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo BMF e o conjunto de dados AmazonFoods. . . . .	196
F.18	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SVD++ e o conjunto de dados AmazonFoods. . . . .	197
F.19	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo UserKNN e o conjunto de dados AmazonFoodsCompleto. . . . .	198
F.20	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo ItemKNN e o conjunto de dados AmazonFoodsCompleto. . . . .	199
F.21	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SMF e o conjunto de dados AmazonFoodsCompleto. . . . .	200
F.22	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo BMF e o conjunto de dados AmazonFoodsCompleto. . . . .	201
F.23	Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SVD++ e o conjunto de dados AmazonFoodsCompleto. . . . .	202
F.24	Intervalos de confiança referentes à QP3 (análise complementar) com o algoritmo SMF e o conjunto de dados Personality2018. . . . .	203
F.25	Intervalos de confiança referentes à QP3 (análise complementar) com o algoritmo BMF e o conjunto de dados Personality2018. . . . .	204

# Lista de Quadros

2.1	Características referentes aos traços do modelo <i>Big Five</i> . . . . .	13
2.2	Características referentes às dimensões do modelo <i>Needs</i> . . . . .	15
2.3	Características referentes às dimensões do modelo <i>Values</i> . . . . .	15
3.1	Comparação entre pesquisas relacionadas. . . . .	37
5.1	Síntese dos resultados referentes à QP2. . . . .	72
5.2	Síntese dos resultados referentes à QP4 (Estratégia 1). . . . .	83
5.3	Síntese da investigação experimental realizada. . . . .	96
C.1	Textos dos <i>reviews</i> de um usuário Amazon (A11LNPG39A2ZV4). . . . .	153

# Lista de Tabelas

5.1	Características referentes aos conjuntos de dados. . . . .	67
5.2	Características referentes ao conjunto de dados Personality2018. . . . .	76
D.1	Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP1. . . . .	155
D.2	P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP2 (para o modelo <i>Big Five</i> ). . . . .	156
D.3	P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP2 (para o modelo <i>Values</i> ). . . . .	157
D.4	P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP2 (para o modelo <i>Needs</i> ). . . . .	158
D.5	Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP3 (para variações omitindo traços). . . . .	159
D.6	Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP3 (para variações sem omitir traços). . . . .	160
D.7	P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP4 (para estratégia “itps”). . . . .	161
D.8	Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP4 (para estratégia <i>weighted</i> ). . . . .	162
D.9	Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP4 (para estratégia <i>re-ranking</i> ). . . . .	162

# Lista de Abreviaturas e Siglas

<b>APR</b>	<i>Automatic Personality Recognition</i> (Reconhecimento Automático de Personalidade)
<b>BMF</b>	<i>Biased Matrix Factorization</i> (Fatoração de Matriz Enviesada)
<b>BPRMF</b>	<i>Bayesian Personalized Ranking Matrix Factorization</i> (Fatoração de Matriz baseada em Ranking Personalizado Bayesiano)
<b>CNN</b>	<i>Convolutional Neural Network</i> (Rede Neural Convolutacional)
<b>FBC</b>	Filtragem Baseada no Conteúdo
<b>FC</b>	Filtragem Colaborativa
<b>FFM</b>	<i>Five Factor Model</i> (Modelo dos Cinco Grandes Fatores)
<b>FIPI</b>	<i>Five-Item Personality Inventory</i> (Inventário de Personalidade de Cinco Itens)
<b>GA</b>	<i>Genetic Algorithm(s)</i> (Algoritmo(s) Genético(s))
<b>IBM</b>	<i>International Business Machines Corporation</i>
<b>ItemKNN</b>	<i>Item based K-Nearest Neighbors</i> (K-Vizinhos Mais Próximos baseados no Item)
<b>itps</b>	Estratégia de recomendação com foco em itens com traços de personalidade similares
<b>KNN</b>	<i>K-Nearest Neighbors</i> (K-Vizinhos Mais Próximos)
<b>LIWC</b>	<i>Linguistic Inquiry and Word Count</i> (Dicionário de Linguística e Contagem de Palavras)
<b>MAP</b>	<i>Mean Average Precision</i> (Média das Precisas Médias)
<b>MBTI</b>	<i>Myers-Briggs Type Indicator</i> (Indicador de Tipos de Myers-Briggs)
<b>ML</b>	<i>Machine Learning</i> (Aprendizado de Máquina)
<b>MRR</b>	<i>Mean Reciprocal Rank</i> (Posição Recíproca Média)
<b>MSI</b>	<i>Mean Self-Information</i>
<b>NDCG</b>	<i>Normalized Discounted Cumulative Gain</i> (Ganho Cumulativo Descontado Normalizado)

---

<b>OCEAN</b>	Acrônimo para <i>Openness–Conscientiousness–Extraversion–Agreeableness–Neuroticism</i> (Abertura–Conscienciosidade–Extroversão–Amabilidade–Neuroticismo)
<b>PB</b>	<i>Personality-Based collaborative filtering approach</i> (abordagem de FC baseada em personalidade)
<b>PVQ</b>	<i>Portrait Values Questionnaire</i> (Questionário de Perfis de Valores)
<b>QP</b>	Questão de Pesquisa
<b>RB</b>	<i>Rating-Based collaborative filtering approach</i> (abordagem de FC baseada em ratings)
<b>RIASEC</b>	Acrônimo para <i>Realistic–Investigative–Artistic–Social–Enterprising–Conventional</i> (Realista–Investigativo–Artístico–Social–Empreendedor–Convencional)
<b>RMSE</b>	<i>Root Mean Square Error</i> (Raiz do Erro Quadrático Médio)
<b>RPBC</b>	<i>Rating-Personality Based Cascade hybrid approach</i> (abordagem híbrida cascata baseada em ratings e personalidade)
<b>RPBL</b>	<i>Rating-Personality Based Linear hybrid approach</i> (abordagem híbrida linear baseada em ratings e personalidade)
<b>SBIE</b>	Simpósio Brasileiro de Informática na Educação
<b>SMF</b>	<i>Simple Matrix Factorization</i> (Fatoração de Matriz Simples)
<b>SR</b>	Sistema(s) de Recomendação
<b>SVD++</b>	<i>Singular Value Decomposition Plus Plus</i> (Decomposição em Valores Singulares Mais Mais)
<b>TIPI</b>	<i>Ten-Item Personality Inventory</i> (Inventário de Personalidade de Dez Itens)
<b>TKI</b>	<i>Thomas-Kilmann Conflict Mode Instrument</i> (Instrumento de Modo de Conflito Thomas-Kilmann)
<b>tps</b>	Estratégia de recomendação com foco em traços de personalidade similares dos usuários
<b>UserKNN</b>	<i>User based K-Nearest Neighbors</i> (K-Vizinhos Mais Próximos baseados no Usuário)
<b>VAE</b>	<i>Variational AutoEncoders</i> (Codificadores Automáticos Variacionais)
<b>WebMedia</b>	Simpósio Brasileiro de Sistemas Multimídia e Web
<b>XML</b>	<i>eXtensible Markup Language</i> (Linguagem de Marcação Extensível)

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivos e Questões de Pesquisa . . . . .	4
1.3	Percurso da Pesquisa e Contribuições . . . . .	6
1.4	Estrutura do documento . . . . .	9
<b>2</b>	<b>Fundamentação Teórica</b>	<b>11</b>
2.1	Personalidade . . . . .	11
2.1.1	Modelo <i>Big Five</i> . . . . .	12
2.1.2	Outros Modelos com Características de Personalidade . . . . .	14
2.1.3	Meios para Identificação de Personalidade . . . . .	16
2.2	Sistemas de Recomendação . . . . .	21
2.2.1	Recomendação via Filtragem Colaborativa . . . . .	24
2.2.2	Recomendação Baseada em Personalidade . . . . .	25
2.3	Outros Fundamentos . . . . .	28
2.3.1	<i>K-means</i> . . . . .	28
2.3.2	Algoritmos Genéticos . . . . .	29
2.4	Considerações Finais do Capítulo . . . . .	29
<b>3</b>	<b>Pesquisas Relacionadas</b>	<b>31</b>
3.1	Pesquisas sobre Recomendação Baseada em Personalidade . . . . .	32
3.2	Considerações Finais do Capítulo . . . . .	47
<b>4</b>	<b>Abordagens Elaboradas</b>	<b>48</b>
4.1	Abordagens referentes à QP1 . . . . .	50

4.2	Abordagens referentes à QP2 . . . . .	52
4.3	Abordagens referentes à QP3 . . . . .	54
4.4	Abordagens referentes à QP4 . . . . .	55
4.4.1	Estratégia 1 . . . . .	56
4.4.2	Estratégia 2 . . . . .	58
4.4.3	Estratégia 3 . . . . .	59
4.5	Considerações Finais do Capítulo . . . . .	61
<b>5</b>	<b>Procedimentos Metodológicos e Resultados</b>	<b>63</b>
5.1	Conjuntos de Dados . . . . .	65
5.2	Análise Experimental . . . . .	67
5.2.1	Experimento referente à QP1 . . . . .	67
5.2.2	Experimento referente à QP2 . . . . .	71
5.2.3	Experimento referente à QP3 . . . . .	74
5.2.4	Experimento referente à QP4 . . . . .	78
5.3	Considerações Finais do Capítulo . . . . .	95
<b>6</b>	<b>Considerações Finais</b>	<b>98</b>
6.1	Conclusões . . . . .	98
6.2	Limitações e Sugestões para Pesquisas Futuras . . . . .	101
<b>A</b>	<b>Publicações</b>	<b>135</b>
<b>B</b>	<b>Slides</b>	<b>140</b>
<b>C</b>	<b>Exemplo referente a APR baseado em textos de um usuário</b>	<b>152</b>
<b>D</b>	<b>Resultados de testes estatísticos</b>	<b>154</b>
<b>E</b>	<b>Boxplots e análises referentes à QP2</b>	<b>163</b>
<b>F</b>	<b>Intervalos de confiança referentes à QP3</b>	<b>179</b>

# Capítulo 1

## Introdução

Com o aumento no volume de recursos digitais disponíveis na Web, o cenário de sobrecarga de conteúdo disponível pode dificultar a tarefa de acesso aos recursos relevantes. É conveniente, pois, auxiliar as pessoas na escolha dos diversos recursos digitais que vão surgindo.

Os Sistemas de Recomendação (SR) são capazes de facilitar a descoberta dos conteúdos mais adequados ao perfil de cada indivíduo, baseando-se em suas características e/ou preferências (ADOMAVICIUS; TUZHILIN, 2005) (BOBADILLA et al., 2013) (AGGARWAL et al., 2016) (WANG et al., 2021).

A utilização de um SR consiste em uma solução visando a contribuir com a experiência do usuário em determinado domínio, recomendando recursos personalizados e relevantes, além de contribuir para explorar o acervo de itens existente.

### 1.1 Motivação

A personalidade é um fator crítico influente no comportamento e nos interesses das pessoas (NUNES; HU, 2012) (KARUMUR; NGUYEN; KONSTAN, 2018). Ao compreender os SR como uma forma de facilitar aos usuários a descoberta de recursos compatíveis com seu perfil, torna-se motivador estudar como considerar, em SR, os aspectos psicológicos desses usuários — tais como traços de personalidade, que influenciam processos de tomada de decisão (NUNES, 2012).

Embora os aspectos psicológicos humanos sejam difíceis de identificar intencionalmente, a aplicação da “Computação da Personalidade” pode agregar valor às recomendações.

Tal termo é uma tradução livre para *Personality Computing*, apresentado por Vinciarelli e Mohammadi (2014) para indicar o campo de estudo, de natureza interdisciplinar, que se concentra na integração das teorias da psicologia da personalidade com os sistemas computacionais. Segundo Dhelim et al. (2021), há evidências de que alavancar as teorias da personalidade pode ajudar a superar desafios de modelagem de usuário.

A questão geral motivadora desta tese é similar à questão de pesquisa da tese de Nunes (2008): “Como melhorar as recomendações geradas pelos SR para oferecer informações, produtos ou serviços mais personalizados para os usuários?”. O objetivo descrito em sua tese — intitulada *Recommender System based on Personality Traits* — consistiu em encontrar evidências de que a personalidade humana pode influenciar o processo de tomada de decisão em computadores.

Após o estudo de Nunes (2008), outros pesquisadores propuseram estratégias de recomendação baseadas em personalidade. Muitas delas consideram que pessoas com personalidades similares tendem a preferir (“dar *likes*” ou avaliar com notas altas) itens similares, ou seja, aplicam Filtragem Colaborativa (FC) baseando-se, além das notas (*ratings*), também em informações explícitas da personalidade dos usuários.

Segundo Hu e Pu (2011), infere-se que pessoas com personalidade semelhante terão interesses e padrões comportamentais semelhantes. Se as preferências (representadas por *likes* e/ou notas de 1 a 5, por exemplo) de usuários com personalidade similar tendem a ser semelhantes, uma hipótese é que a personalidade dos usuários é um fator que já influencia os algoritmos de FC de maneira implícita (pois utilizam os *likes* e/ou notas fornecidos por usuários, influenciados implicitamente pelos seus traços de personalidade). Em outras palavras, a personalidade dos usuários acaba influenciando todos os algoritmos de FC, independentemente da técnica de FC utilizada.

Partindo da premissa “pessoas similares quanto à personalidade tendem a preferir itens similares”, se, em um algoritmo de FC, o “treinamento” para se definir uma nota (*rating*) for realizado desconsiderando a “opinião” (*ratings*) de usuários não tão similares quanto à personalidade, outra hipótese é que isso não tenderia a diminuir significativamente a qualidade da predição (isto é: permanecendo-se igual ou melhorando-a). Em outras palavras, a hipótese levantada é que, na execução de determinado algoritmo de FC, ao considerar apenas a “opinião” dos usuários mais próximos, em cada traço de personalidade, a predição pode

melhorar.

Outra hipótese é que alguns traços poderiam ter mais influência no aspecto de “preferências similares”, sendo adequado analisar quais traços podem impactar mais fortemente na melhoria da predição de uma “opinião” (nota), realizando inclusive ponderações se for o caso.

A maioria das pesquisas propondo estratégias de recomendação baseadas em traços de personalidade utiliza questionários para identificar personalidade. Nunes (2008), também usando questionários, expõe, nas conclusões de sua tese, este dilema: se for usado um questionário refinado (com muitas questões, para, por exemplo, conseguir identificar as facetas do modelo de personalidade), as pessoas não o responderão; se não for usado um questionário refinado (como o uso de questionários com apenas 10 itens, utilizado em muitas pesquisas), os traços de personalidade serão bastante abstratos para serem usados com o intuito de melhorar efetivamente as recomendações.

Além disso, há o aspecto de que, ao utilizar um questionário, o respondente pode não preenchê-lo com cuidado (por considerar como algo delongado e/ou entediante), acarretando facilmente a produção de recomendações irrelevantes (DHELIM et al., 2021). Explicitar numericamente a personalidade de uma pessoa não é uma tarefa trivial, mas há avanços referentes à identificação da personalidade dos usuários sem preenchimento de questionários (APR — *Automatic Personality Recognition*). Por exemplo, com o *Watson Personality Insights*, da IBM (*International Business Machines Corporation*), tal identificação é facilitada: é empregada análise linguística para induzir características de personalidade individuais (IBM, 2021).

Com base nisso, outro elemento motivador da pesquisa foi a utilização do *Watson Personality Insights*, para analisar, primeiramente, algumas estratégias propostas avaliadas anteriormente com a identificação via questionários e, principalmente, analisar as hipóteses levantadas nesta seção com a identificação via uma ferramenta para possibilitar facilmente o uso posterior (por exemplo, por parte de empresas de comércio eletrônico), sem precisar que os usuários (clientes) preencham questionários de personalidade.

A partir do exposto nesta seção, a realização desta pesquisa, em síntese, foi motivada pela possibilidade de se inspirar nas ideias da tese de Nunes (2008) — pioneira na recomendação baseada em traços de personalidade — considerando os avanços na identificação da perso-

nalidade humana e nos algoritmos de FC. Laconicamente, partindo-se do pressuposto que usuários similares quanto à personalidade tendem a preferir itens similares, buscou-se reunir informações com o propósito de responder ao seguinte problema: “De que forma o conhecimento das características de personalidade dos usuários de SR, identificadas automaticamente via textos, pode contribuir para melhorar, em cenários gerais, a acurácia dos algoritmos de recomendação?”.

Partindo-se da motivação apresentada, foi formulada a principal hipótese desta tese: é possível obter recomendações de itens mais acuradas para os usuários, em cenários gerais, ao incorporar informações referentes à personalidade, obtidas via APR baseado em textos (linguagem natural) desses usuários.

Com isso, foram elencados as questões de pesquisa e os objetivos descritos na Seção 1.2.

## 1.2 Objetivos e Questões de Pesquisa

Considerando-se a motivação apresentada na Seção 1.1, o objetivo geral desta pesquisa de doutorado é avançar o campo referente à personalização das recomendações ao investigar estratégias para a melhoria dos SR baseados em FC, utilizando dados da personalidade dos usuários identificada via análise linguística (APR baseado em texto).

Em outras palavras, com esta pesquisa, objetiva-se, ao rever a ideia da recomendação via FC baseada em personalidade, investigar se o uso de dados explícitos da personalidade dos usuários (identificada via ferramenta *Watson Personality Insights*, e não via questionários) contribui para melhorar a acurácia dos algoritmos de FC (independentemente da abordagem de FC), enfocando na “colaboração” (dados de opinião) dos usuários similares quanto à personalidade.

A partir de tal objetivo geral, o desenvolvimento deste estudo foi guiado pelas seguintes questões de pesquisa.

- **QP1:** Os resultados obtidos pelos primeiros estudos relevantes que propuseram estratégias de recomendação via FC baseadas em traços de personalidade (identificados via questionários) são similares aos obtidos utilizando-se APR e conjuntos de dados maiores?

- **QP2:** A acurácia das predições de um algoritmo de FC é afetada se o algoritmo for executado focando na opinião de usuários similares quanto à personalidade (identificada via APR)?
- **QP3:** Os traços de personalidade (identificados via APR) influenciam de maneira diferente no resultado da recomendação? É possível obter melhores resultados ao ignorar algum(ns) traço(s) ou ao ponderá-los?
- **QP4:** Baseando-se nas análises referentes às questões de pesquisa QP1, QP2 e QP3, é possível incorporar estratégias aos algoritmos de FC (utilizando explicitamente dados de personalidade) e obter recomendações mais relevantes?

Orientados por essas questões de pesquisa, foram definidos os seguintes objetivos específicos.

- **O1:** Rever pesquisas iniciais de referência para verificar se os resultados se mantêm ao obter dados de personalidade via APR (baseado em texto) e conjuntos de dados com maior número de usuários e itens. Este objetivo relaciona-se à questão de pesquisa QP1.
- **O2:** Analisar se a acurácia das predições de um algoritmo de FC se mantêm (ou tende a melhorar) ao desconsiderar os “dados de opinião” (*ratings*) de usuários menos similares quanto à personalidade (identificada automaticamente via textos) de um usuário-alvo de recomendação. Este objetivo relaciona-se à questão de pesquisa QP2.
- **O3:** Analisar a influência dos traços de personalidade (identificados via APR baseado em texto) visando a destacar se algum(ns) deles influencia(m) mais fortemente na melhoria dos resultados dos algoritmos de FC baseada em personalidade. Este objetivo relaciona-se à questão de pesquisa QP3.
- **O4:** Propor e avaliar estratégias para a geração de recomendações personalizadas baseadas nas características de personalidade dos usuários e suas preferências anteriores, indicando se algoritmos de FC geram resultados mais relevantes ao considerar explicitamente a personalidade dos usuários, identificada automaticamente via textos. Este objetivo relaciona-se à questão de pesquisa QP4.

### 1.3 Percurso da Pesquisa e Contribuições

Esta pesquisa de doutorado se iniciou com foco no domínio educacional, visando a trazer contribuições referentes especificamente à recomendação personalizada de objetos de aprendizagem, com base em características da personalidade dos estudantes. Entretanto, ao analisar a literatura, a motivação mudou ao longo da pesquisa e, assim, seu objetivo foi alterado, ficando menos dependente de um domínio específico de aplicação.

Como contribuição inicial da pesquisa, destacam-se as análises referentes ao levantamento do estado-da-arte. A revisão bibliográfica inicial, focando no domínio educacional, sobre aspectos que influenciam na aquisição de conhecimento de um indivíduo, culminou na publicação de um artigo no periódico *Informática na Educação: Teoria & Prática (IETP)*, intitulado “Considerando Estilos de Aprendizagem, Emoções e Personalidade em Informática na Educação” (AGUIAR, 2017). Além disso, o levantamento apresentado no Capítulo 3, sumarizado no quadro comparativo de pesquisas, complementa estudos secundários recentes sobre o estado-da-arte referente a SR baseados em personalidade.

Ainda sobre a realização do levantamento bibliográfico inicial, uma contribuição indireta foi, a partir do minicurso denominado “Sistemas de Recomendação: Conceitos, Técnicas e Aplicações”, ministrado no III Encontro de Educação, Ciência e Tecnologia (ENECT)<sup>1</sup> e no Sábado Temático<sup>2</sup>, a divulgação dessa área de pesquisa aos participantes desses eventos — alunos da Universidade Estadual da Paraíba (UEPB) e do Instituto Federal da Paraíba (IFPB), que mostraram interesse pela área de Sistemas de Recomendação.

Destacam-se, também, como contribuição, os estudos sobre influência dos componentes dos modelos teóricos analisados inicialmente nesta pesquisa, no domínio educacional. Foi publicada, no XXVIII Simpósio Brasileiro de Informática na Educação (SBIE), uma análise referente à influência das dimensões de Estilos de Aprendizagem do modelo de Felder–Silverman (AGUIAR et al., 2017). Em relação à influência dos traços de personalidade do modelo *Big Five*, uma análise foi publicada no XXX SBIE (AGUIAR; ARAÚJO; COSTA, 2019): em síntese, os resultados indicaram que os traços de personalidade influenciam de maneira diferente na recomendação personalizada de recursos educacionais para estudantes,

<sup>1</sup>Minicurso ministrado no dia 27 de março de 2018, na UEPB, campus Campina Grande — (<https://www.even3.com.br/enect2018>)

<sup>2</sup>Minicurso ministrado no dia 21 de julho de 2018, no IFPB, campus Campina Grande — (<https://sites.google.com/view/sabtematicoifpbeg>)

não havendo um destaque expressivo de um dos traços, para, por exemplo, desconsiderar os outros na utilização do modelo *Big Five* em Sistemas de Recomendação Educacionais. Essas contribuições relacionam-se à questão de pesquisa QP3.

Além disso, foi realizado um estudo, denominado “Recomendação de Objetos de Aprendizagem utilizando Filtragem Colaborativa baseada em Tendências e em Estilos de Aprendizagem” (AGUIAR; ARAÚJO; COSTA, 2018b), propondo uma estratégia de recomendação, no domínio educacional, baseada em FC e nos Estilos de Aprendizagem dos estudantes.

Destaca-se, ainda, que, em um projeto desta tese, foi realizado um estudo intitulado “Estratégia baseada em Personalidade e Tendências para Recomendação de Objetos de Aprendizagem usando Algoritmo Genético” (AGUIAR; ARAÚJO; COSTA, 2018a), publicado no I Workshop Latino-Americano de Trabalhos em Andamento em Computação — evento satélite do CLEI–LACLO 2018 (XLIV Conferência Latino-americana de Informática & XIII Conferência Latino-americana de Tecnologias de Aprendizagem). Tal workshop, dentre seus objetivos, oportuniza o compartilhamento de ideias referentes a teses de doutorado em andamento.

Relacionado às questões de pesquisa QP1 e QP4, foi publicado, na Revista Ibérica de Sistemas e Tecnologias de Informação (RISTI), o artigo intitulado “Estudo Comparativo de Abordagens para Sistemas de Recomendação baseados em Personalidade com uso do serviço IBM Watson Personality Insights” (AGUIAR; ARAÚJO; COSTA, 2020b). Além disso, no Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), foi apresentado o artigo “*Collaborative Filtering Strategy for Product Recommendation Using Personality Characteristics of Customers*” (AGUIAR; ARAÚJO; COSTA, 2020a), propondo uma estratégia conexa ao objetivo orientado pela questão de pesquisa QP4.

Em outro artigo apresentado — este para o SBIE 2020 —, intitulado “Estudo de Desempenho de Algoritmos de Filtragem Colaborativa para Sistemas de Recomendação Educacionais aplicando Agrupamento Prévio de Usuários com Traços de Personalidade Similares” (AGUIAR; ARAÚJO; COSTA, 2020c), foi analisada a questão de pesquisa QP2, considerando-se o domínio educacional, e sem uso do *Watson Personality Insights*.

Com esta pesquisa, pretende-se também disseminar a ideia de que, embora existam SR baseados em traços de personalidade (com uso explícito de dados de personalidade dos usuários), a personalidade dos usuários, por influenciar seu processo de tomada de decisão,

está sendo considerada implicitamente desde os primeiros algoritmos de FC que realizaram predições/recomendações considerando as notas (ou *likes*) anteriores de seus usuários. Durante a pesquisa, investigou-se a contribuição do uso explícito de dados de personalidade para a melhoria dos algoritmos de FC.

Além de propor novas estratégias de FC para contribuir com o estado-da-arte, outra contribuição da pesquisa consiste na avaliação, com APR baseado em texto (via *Watson Personality Insights*), de alguns algoritmos relevantes na literatura — avaliados anteriormente com menos usuários e itens, com a identificação de personalidade via questionários, e em domínio de aplicação diferente. Destaca-se o uso do *Watson Personality Insights* por tal ferramenta realizar a identificação via análise textual considerando as facetas do *Big Five*, além de usar os modelos *Needs* e *Values* (não foi encontrada outra ferramenta de APR com tais funcionalidades). Além disso, o *Watson Personality Insights* facilitaria, por parte das empresas, incorporar nos seus sistemas as características de personalidade dos usuários. A elaboração de uma ferramenta similar, especialmente incorporando o português brasileiro, embora bastante útil, não era o foco para esta pesquisa.

Em síntese, é possível elencar estes feitos ao longo da pesquisa descrita nesta tese: levantamento bibliográfico sobre SR baseados em personalidade [Capítulo 3]; elaboração de uma estratégia de recomendação com uso específico do modelo *Needs* (na literatura, há com *Values* e *Big Five*) [Seção 4.1]; mescla de estratégias de FC, baseadas em personalidade, oriundas de grupos relevantes da área [Seção 4.1]; criação de uma estratégia genérica de agrupamento baseada em características de personalidade [Seção 4.2]; análise investigativa indicando que os traços de personalidade podem influenciar diferentemente na recomendação de itens [Seção 4.3 e Subseção 5.2.3]; elaboração de uma estratégia aplicando algoritmo genético para abranger características diversas referentes a distintos modelos de personalidade [Seção 4.3]; proposta de uma estratégia de recomendação referente à “personalidade de itens” (aspecto pouco explorado na literatura) [Subseção 4.4.1]; proposta de estratégias híbridas (*weighted* e *re-ranking*) considerando fatoração de matriz e personalidade dos usuários [Subseções 4.4.2 e 4.4.3]; avaliação de estratégias de recomendação em cenários com identificação de características de personalidade dos usuários sem uso de questionários (via APR baseado em texto) [Capítulo 5].

Além de proporcionar à comunidade algumas estratégias de recomendação de itens (con-

siderando características da personalidade do usuário na construção do seu perfil), que podem ser utilizadas em variados domínios, esta pesquisa também visa a difundir o entendimento de que os usuários possuem particularidades, com diferentes características. Também pretende-se contribuir com a difusão do termo “Computação da Personalidade”, ainda pouco empregado em pesquisas em português.

Considerando-se os comentários anteriores desta seção e os objetivos específicos visados pela pesquisa, é possível destacar os seguintes pontos como as principais contribuições:

- Análise referente à influência das características de personalidade humana (componentes de alguns modelos definidos na literatura) no processo de recomendação de itens via FC.
- Estratégias de recomendação personalizada de itens, independentes de domínio, empregando características da personalidade dos usuários.
- Análise de abordagens de recomendação baseada em personalidade (comparando abordagens do estado-da-arte, variações de abordagens existentes e estratégias propostas) aplicando APR baseado em texto.

Por fim, almeja-se com esta pesquisa e seus resultados — a exemplo das contribuições bibliográficas (publicações citadas nesta seção e detalhadas no Apêndice A) — continuar incentivando estudos sobre SR considerando aspectos psicológicos.

## 1.4 Estrutura do documento

Além deste capítulo introdutório, este documento contempla mais cinco capítulos, estruturados como informado a seguir.

No Capítulo 2, é apresentado o embasamento teórico relativo à Personalidade e a Sistemas de Recomendação, destacando-se os conceitos necessários para entendimento dos capítulos seguintes.

No Capítulo 3, é apresentada uma revisão da literatura, na qual são apresentadas pesquisas desenvolvidas no âmbito de recomendação de recursos, especialmente aquelas envolvendo os conceitos empregados nesta tese.

No Capítulo 4, são comentadas as abordagens elaboradas considerando-se as ideias durante o desenvolvimento da pesquisa.

No Capítulo 5, são abordados os aspectos referentes à realização das análises experimentais, sendo também apresentados os resultados obtidos.

No Capítulo 6, são apresentadas as considerações finais acerca da pesquisa, indicando-se sugestões para pesquisas futuras.

Além dos capítulos, há informações adicionais em apêndices (a exemplo do Apêndice B, que contém os slides elaborados para a apresentação oral desta pesquisa).

# Capítulo 2

## Fundamentação Teórica

Este capítulo destina-se a apresentar os fundamentos teóricos referentes às áreas do conhecimento relacionadas ao objetivo da pesquisa. Na Seção 2.1 são abordados conceitos fundamentais sobre Personalidade. A fundamentação sobre Sistemas de Recomendação é contemplada na Seção 2.2. Outros fundamentos são apresentados na Seção 2.3.

### 2.1 Personalidade

Apesar de ser uma das linhas de pesquisa mais antigas da Psicologia, não há um consenso sobre a definição do termo Personalidade. Allport (1937), ao examinar a literatura, extraiu aproximadamente cinquenta definições diferentes. Embora não haja uma única definição empregada por todos os teóricos da personalidade, é possível considerá-la como um padrão consistente de traços de comportamento e processos intrapessoais originados dentro do indivíduo — características que dão consistência e individualidade ao comportamento de uma pessoa nas diversas situações ao longo do tempo (BURGER, 2010) (FEIST; FEIST; ROBERTS, 2015).

Há várias perspectivas referentes à personalidade dos indivíduos — biológica, cognitiva, psicanalítica, dos traços, entre outras. A perspectiva dos traços de personalidade tem sido a mais comumente aplicada à representação de personalidade, tanto na academia quanto na indústria, dado que é a mais facilmente mensurável de forma explícita em humanos por meio de computadores (NUNES, 2009) (NUNES, 2012) (WANG et al., 2021).

Historicamente, diferentes psicólogos da personalidade tiveram sua lista particular de

traços em que se focaram e, até pelo menos a década de 1980, houve pouco consenso em relação a definir as dimensões principais da personalidade. Atualmente, há uma tendência consensual para considerar a existência de cinco dimensões principais da personalidade, os “cinco grandes” — *Big Five* — traços da personalidade (FEIST; FEIST; ROBERTS, 2015).

### 2.1.1 Modelo *Big Five*

Os estudos de personalidade baseados em traços contribuíram para o desenvolvimento gradual do Modelo dos Cinco Grandes Fatores, conhecido internacionalmente como *Five Factor Model* (FFM) ou *Big Five*, consistindo em um modelo de ampla adoção e aceitação, especialmente devido às pesquisas de Robert McCrae e Paul Costa (JOHN; SRIVASTAVA, 1999) (MCCRAE; COSTA, 1997) (FEIST; FEIST; ROBERTS, 2015).

Conforme apresentado por Feist, Feist e Roberts (2015), McCrae e Costa focaram, até o início da década de 1980, em duas dimensões (Neuroticismo e Extroversão); em seguida, encontraram um terceiro fator (Abertura) e defenderam um modelo da personalidade com três fatores; em meados da década de 1980, eles começaram a tratar os cinco fatores. A partir do final da década de 1980, a maioria dos psicólogos da personalidade optou pelo *Big Five*, sendo os cinco fatores encontrados em culturas variadas. Segundo McCrae e Costa (2003), os adultos tendem a manter a estrutura (fatores) de personalidade conforme envelhecem.

Segundo a IBM (2021), o *Big Five* “é o modelo de personalidade mais amplamente usado para descrever como uma pessoa geralmente se envolve com o mundo”. Segundo Jaques e Nunes (2021), o *Big Five* é o modelo de personalidade tradicionalmente utilizado na computação, pois é um dos mais aptos a se modelar e implementar em computadores. Recentemente, o modelo foi usado em pesquisas no contexto da pandemia ocasionada pela COVID-19, a exemplo da pesquisa de Carvalho, Pianowski e Gonçalves (2020), com o objetivo de investigar a associação entre dois traços do *Big Five* e o engajamento com as medidas de distanciamento social e lavagem das mãos.

Os cinco fatores do *Big Five* podem ser definidos, em português, como: Abertura, Conscienciosidade, Extroversão, Amabilidade e Neuroticismo. No Quadro 2.1, são apresentadas considerações relativas a cada um dos traços desse modelo.

Quadro 2.1: Características referentes aos traços do modelo *Big Five*.

Traços	Características
Abertura ou Abertura à Mudança ou Abertura à Experiência ( <i>Openness</i> )	<b>Significado:</b> refere-se aos comportamentos exploratórios, ao reconhecimento da importância de ter novas experiências, à imaginação ativa, à sensibilidade estética, à curiosidade intelectual e à preferência pela variedade.
	<b>Índices altos:</b> tendem a representar pessoas curiosas, criativas, imaginativas, que se divertem com novas ideias, que costumam apresentar valores não convencionais, que experimentam intensamente uma gama ampla de emoções.
	<b>Índices baixos:</b> tendem a representar pessoas convencionais, dogmáticas, rígidas nas suas crenças e atitudes, conservadoras nas suas preferências e menos responsivas emocionalmente.
	<b>Facetas:</b> Aventura, Interesses artísticos, Emotividade, Imaginação, Intelecto, Liberalismo.
Conscienciosidade ou Conscientização ou Realização ( <i>Conscientiousness</i> )	<b>Significado:</b> refere-se ao grau de organização e condução de tarefas, à persistência, à motivação, ao controle e ao autocontrole, em termos de planejamento, apresentados pelos indivíduos.
	<b>Índices altos:</b> tendem a representar pessoas mais organizadas, confiáveis, trabalhadoras, determinadas, pontuais, meticolosas, ambiciosas e persistentes.
	<b>Índices baixos:</b> tendem a representar pessoas preguiçosas, descuidadas, desorganizadas, negligentes, que não têm objetivos claros, e que tendem a ter pouco comprometimento e responsabilidade diante de tarefas.
	<b>Facetas:</b> Esforço por realizações, Cautela, Senso de dever, Ordem, Autodisciplina, Competência.
Extroversão ( <i>Extraversion</i> )	<b>Significado:</b> refere-se às formas como os indivíduos interagem com os outros, indicando o quanto são comunicativos, falantes, ativos, assertivos, responsivos e gregários.
	<b>Índices altos:</b> tendem a representar pessoas mais sociáveis, ativas, falantes, otimistas, alegres, bem dispostas e cordiais.
	<b>Índices baixos:</b> não representam necessariamente pessoas infelizes, não amistosas ou pessimistas, mas tendem a representar pessoas reservadas, introvertidas, sóbrias, indiferentes, independentes e quietas.
	<b>Facetas:</b> Atividade, Assertividade, Emoções positivas, Busca por entusiasmo, Simpatia, Gregarismo.
Amabilidade ou Socialização ( <i>Agreeableness</i> )	<b>Significado:</b> refere-se à qualidade das relações interpessoais dos indivíduos e aos tipos de interações que um indivíduo apresenta — predisposição para ser amistoso, para se sensibilizar com a situação alheia, para se colocar no lugar do outro, para demonstrar gentileza, complacência, compaixão.
	<b>Índices altos:</b> tendem a representar pessoas generosas, amáveis, afetuosas, prestativas e altruístas, com tendência à responsabilidade e empatia, e que acreditam que a maioria das outras pessoas agirá da mesma forma.
	<b>Índices baixos:</b> tendem a representar pessoas petulantes, não cooperativas, facilmente irritáveis, e que podem apresentar-se de maneira manipuladora, vingativa e insensível.
	<b>Facetas:</b> Altruísmo, Cooperação, Modéstia, Franqueza, Complacência, Confiança.
Neuroticismo ou Instabilidade Emocional ( <i>Neuroticism</i> )	<b>Significado:</b> refere-se à tendência cotidiana dos indivíduos a experimentar padrões emocionais vinculados a desconforto psicológico (aflição, angústia, culpa, medo, nojo, raiva, sofrimento, tristeza, vergonha, entre outros), além de referir-se à ansiedade excessiva, à pouca habilidade para conter impulsos e para lidar com o estresse, à propensão a ideias dissociadas da realidade (irracionais), e à dificuldade para tolerar a frustração causada pela não realização de desejos.
	<b>Índices altos:</b> tendem a representar pessoas predispostas a vivenciar mais profundamente o seu sofrimento emocional.
	<b>Índices baixos:</b> tendem a representar pessoas mais tranquilas, relaxadas, constantes, menos agitadas, com habilidade maior em lidar com situações estressantes e, em síntese, com estabilidade emocional.
	<b>Facetas:</b> Hostilidade, Ansiedade, Depressão, Impulsividade, Autoconsciência, Vulnerabilidade ao estresse.

Fonte: adaptação de Aguiar, Fechine e Costa (2015, p. 158).

### 2.1.2 Outros Modelos com Características de Personalidade

Além do *Big Five*, há outros modelos tratando características referentes à personalidade dos indivíduos. Nesta subseção, são comentados os outros modelos empregados na pesquisa: o modelo *Needs* (Necessidades) e o modelo *Values* (Valores).

As pessoas têm necessidades que, segundo Maslow (1943), podem ser categorizadas como: realização pessoal (incluindo, por exemplo, criatividade e espontaneidade); estima (incluindo, por exemplo, respeito aos outros e autoestima); afeto (a exemplo de amizade e intimidade sexual); segurança (incluindo, por exemplo, segurança do emprego e da saúde); e as necessidades fisiológicas (a exemplo de respiração e sono).

Na literatura de marketing (FORD, 2005), as necessidades humanas — referentes aos desejos que um indivíduo espera serem atendidos quando analisa um produto/serviço — são descritas em doze categorias: Desafio, Proximidade, Curiosidade, Entusiasmo, Harmonia, Ideal, Liberdade, Amor, Praticidade, Autoexpressão, Estabilidade e Estrutura. Tais categorias são a base do modelo *Needs*, e são resumidamente caracterizadas no Quadro 2.2 — elaborado com base nos quadros apresentados por Srivastava, Bala e Kumar (2020) e IBM (2021).

Valores significam a importância de coisas diferentes (por exemplo, poder, segurança, tradição, sucesso, felicidade, status social) na vida das pessoas. Os valores influenciam as ações e os comportamentos dos indivíduos. As prioridades dos valores humanos diferem de um indivíduo para outro, resultando em diversidades de ações de um indivíduo em diferentes situações (KHAN et al., 2020).

De acordo com Srivastava, Bala e Kumar (2020), várias teorias e modelos foram sugeridos ao longo do tempo para medir os valores humanos, sendo a teoria de Schwartz reconhecida em todas as principais culturas. Segundo Schwartz (1992), os valores humanos básicos são objetivos transsituacionais desejáveis, de importância variável, e servem como princípios orientadores na vida das pessoas.

O modelo *Values* descreve os fatores de motivação que influenciam a tomada de decisão de um indivíduo, considerando as cinco dimensões dos valores humanos básicos propostos por Schwartz (1992): Conservação, Entusiasmo, Hedonismo, Autoaprimoramento e Auto-transcendência. Tais dimensões são resumidamente caracterizadas no Quadro 2.3 — elaborado com base nos quadros apresentados por Srivastava, Bala e Kumar (2020) e IBM (2021).

Quadro 2.2: Características referentes às dimensões do modelo *Needs*.

<i>Needs</i>	<b>Características</b>
Desafio ( <i>Challenge</i> )	Indivíduos que pontuam alto anseiam por realização (obter sucesso) ao enfrentar desafios, aceitando-os voluntariamente. Relaciona-se ao desejo de vitória sobre um oponente e ao desejo de alcançar autoridade e status ou prestígio.
Proximidade ( <i>Closeness</i> )	Indivíduos que pontuam alto gostam da ligação familiar, de formar um lar, do sentimento de pertença.
Curiosidade ( <i>Curiosity</i> )	Indivíduos que pontuam alto têm o desejo de desvendar, descobrir e aprender.
Entusiasmo ( <i>Excitement</i> )	Indivíduos que pontuam alto demonstram empolgação, desejam diversão, sair e aproveitar a vida, têm emoções positivas e alta energia.
Harmonia ( <i>Harmony</i> )	Indivíduos que pontuam alto valorizam e respeitam as outras pessoas, seus pontos de vista e seus sentimentos.
Ideal ( <i>Ideal</i> )	Indivíduos que pontuam alto desejam a perfeição (associada à sofisticação e também ao desenvolvimento espiritual refinado) e têm um senso de comunidade e progresso responsável.
Liberdade ( <i>Liberty</i> )	Indivíduos que pontuam alto têm sede de ter controle sobre as próprias ações sem restrições, a necessidade de sair e fazer coisas novas (senso de possibilidade total).
Amor ( <i>Love</i> )	Indivíduos que pontuam alto desfrutam do contato social, seja em uma relação de um para um ou de um para muitos.
Praticidade ( <i>Practicality</i> )	Indivíduos que pontuam alto têm o desejo de concluir tarefas (realização de trabalhos), de ter habilidade, qualificação e eficiência, o que pode incluir a expressão física e a experiência.
Autoexpressão ( <i>Self-expression</i> )	Indivíduos que pontuam alto desfrutam da descoberta e da afirmação de suas próprias identidades, expressando seu ego, sua força e sua independência por meio de ações externas.
Estabilidade ( <i>Stability</i> )	Indivíduos que pontuam alto buscam segurança física, controle, confiabilidade, e favorecem o que é sensato, que foi experimentado e testado, com bom histórico.
Estrutura ( <i>Structure</i> )	Indivíduos que pontuam alto precisam que as coisas estejam bem organizadas e sob controle (desejo de manter as coisas em ordem). Relaciona-se à rotina, consistência, sistematicidade e previsibilidade.

Fonte: autoria própria — baseando-se em Srivastava, Bala e Kumar (2020) e IBM (2021).

Quadro 2.3: Características referentes às dimensões do modelo *Values*.

<i>Values</i>	<b>Características</b>
Conservação ( <i>Conservation</i> )	Indivíduos que pontuam alto enfatizam a tradição, a autorrestrrição, a ordem e a resistência à mudança. Ênfase em segurança, harmonia e estabilidade da sociedade, dos relacionamentos e de si mesmo. Relaciona-se à restrição de ações, inclinações e impulsos que possam perturbar ou prejudicar outras pessoas e violar expectativas ou normas sociais. Demonstração de respeito, compromisso e aceitação dos costumes e ideias que a cultura ou religião de uma pessoa oferece.
Entusiasmo ( <i>Openness-to-change</i> )	Indivíduos que pontuam alto demonstram empolgação, enfatizam pensamento e ação independentes, escolhendo, criando, explorando; assim como uma prontidão para novas experiências (desafios na vida).
Hedonismo ( <i>Hedonism</i> )	Indivíduos que pontuam alto buscam ter prazer na vida e a gratificação sensual para si.
Autoaprimoramento ( <i>Self-enhancement</i> )	Indivíduos que pontuam alto buscam alcançar o sucesso pessoal demonstrando competência de acordo com os padrões sociais. Relaciona-se ao prestígio e status social, controle/domínio sobre pessoas e recursos.
Autotranscendência ( <i>Self-transcendence</i> )	Indivíduos que pontuam alto demonstram preocupação pelo bem-estar dos outros (buscando preservar e melhorar o bem-estar daqueles com quem mantém contato pessoal frequente; demonstrando compreensão, apreço, tolerância e proteção para o bem-estar de todas as pessoas e para a natureza).

Fonte: autoria própria — baseando-se em Srivastava, Bala e Kumar (2020) e IBM (2021).

### 2.1.3 Meios para Identificação de Personalidade

Para delinear um perfil da personalidade de um indivíduo, há diversos questionários disponíveis. Os questionários (com respostas fornecidas normalmente no formato de escala de Likert) são o meio mais comumente usado para avaliação de personalidade (DHELIM et al., 2021).

Antes da maioria dos psicólogos da personalidade passar a optar pelo *Big Five*, McCrae e Costa analisaram a maioria dos principais inventários da personalidade, a exemplo do Indicador de Tipos de Myers-Briggs (MBTI — *Myers-Briggs Type Indicator*) (FEIST; FEIST; ROBERTS, 2015). Outros exemplos de questionários sem usar o modelo *Big Five* são o Instrumento de Modo de Conflito Thomas-Kilmann (TKI — *Thomas-Kilmann Conflict Mode Instrument*) e o *Holland Code (RIASEC) Test* (teste referente ao Código de Holland, também chamado Modelo RIASEC — Realista–Investigativo–Artístico–Social–Empreendedor–Convencional). Esses três instrumentos foram utilizados em algumas pesquisas relacionadas citadas no Capítulo 3.

Em relação ao modelo *Values*, há o Questionário de Perfis de Valores (PVQ — *Portrait Values Questionnaire*), um questionário com 21 itens a serem avaliados em uma escala de Likert de 6 pontos (SCHWARTZ, 2003). Em relação ao modelo *Big Five*, há vários questionários validados, com variações no número de itens.

Um dos principais deles é um inventário comercial, com direitos autorais: o NEO-PI-R — *Revised NEO Personality Inventory* (COSTA; MCCRAE, 1992). Este instrumento possui 240 questões de múltipla escolha (por meio das quais é possível avaliar os 5 fatores do *Big Five* e as 6 facetas de cada fator) e foi validado para corroborar a ideia da universalidade da perspectiva dos traços de personalidade (MCCRAE; COSTA, 1997).

O NEO-IPIP — *NEO International Personality Item Pool* (JOHNSON, 2000) — é uma alternativa de custo livre ao NEO-PI-R, com 300 itens (60 para cada um dos fatores do *Big Five*, 10 para cada faceta). Tais itens correspondem a afirmativas, às quais o respondente atribui um valor de concordância (em uma escala de Likert de cinco pontos) do quanto cada uma delas o representa.

Além desses, há outros questionários, tais como o NEO-FFI — *NEO Five-Factor Inventory* (COSTA; MCCRAE, 1992) —, versão reduzida do NEO-PI-R, com 60 itens, e o BFI — *Big-Five Inventory*, com 44 itens (JOHN; DONAHUE; KENTLE, 1991). Há também ques-

tionários curtos, tais como o BFI-10, uma versão curta do BFI com 10 itens (RAMMSTEDT; JOHN, 2007), e o TIPI (*Ten-Item Personality Inventory*), uma versão bastante reduzida do NEO-IPIP, com 10 itens (GOSLING; RENTFROW; SWANN, 2003).

Embora o uso dos questionários tradicionais seja a prática mais comum para identificar informações sobre a personalidade humana, geralmente isso é cansativo para os respondentes, sobrecarregando-os. Segundo alguns pesquisadores, a melhor forma de obter os traços de personalidade é por meio de abordagens sem considerável esforço cognitivo por parte dos usuários (JAQUES; NUNES, 2021).

Além disso, Dhelim et al. (2021) destacam duas desvantagens principais nos questionários de autoavaliação: (i) o problema do autopreenchimento (quando o sujeito tende a fornecer respostas erradas a algumas das características sociais indesejadas em determinadas circunstâncias — a exemplo de não confirmar “tendência a ser preguiçoso”) e (ii) o efeito de grupo de referência (quando as respostas fornecidas pelo sujeito são relativas ao ambiente ao seu redor — a exemplo de um indivíduo introvertido julgar-se extrovertido por estar rodeado por um grupo de amigos ainda mais introvertidos).

A integração dos traços de personalidade dos usuários em sistemas computacionais despertou novas direções de pesquisa, como o Reconhecimento Automático de Personalidade (APR — *Automatic Personality Recognition*). Embora o uso de questionários seja, em geral, mais preciso que APR, por vezes o preenchimento de um questionário é inconveniente (DHELMIM et al., 2021).

Mairesse et al. (2007) e Tausczik e Pennebaker (2010) mostraram haver correlação entre o *Big Five* e as características linguísticas encontradas em textos. Bachrach et al. (2012), Markovikj et al. (2021) e Schwartz et al. (2013) indicaram haver correlação entre traços de personalidade e padrões de uso da rede social Facebook, podendo facilitar a identificação de personalidade por meio da mineração dos dados dos perfis de usuário. Ferwerda e Tkalčič (2018) apresentaram um estudo referente a prever a personalidade dos usuários da rede social Instagram com base no conteúdo de suas fotos e nos recursos visuais (saturação, etc.).

Considerando-se APR como o processo de mapear os dados (de redes sociais, por exemplo) relacionados a um indivíduo para uma pontuação representativa de seu tipo de personalidade, é possível realizar esta classificação: (i) APR baseado em comportamento, em que a personalidade é inferida ao associar traços dominantes relevantes a padrões comportamentais

do indivíduo (em um jogo, por exemplo); (ii) APR baseado em multimídia, em que a personalidade é inferida ao associar traços de personalidade a recursos de fotos/vídeos do perfil de mídia social do indivíduo (por exemplo, relacionando determinado traço a usuários que postam frequentemente determinado tipo de foto); e (iii) APR baseado em texto, em que a personalidade é inferida ao associar traços de personalidade a mensagens ou postagens (em formato de texto) em mídias sociais, baseando-se geralmente na frequência de escolha de palavras do indivíduo (DHELIM et al., 2021).

Em geral, APR baseado em texto tem maior precisão que APR baseado em multimídia e APR baseado em comportamento. Por exemplo, um usuário pode compartilhar frequentemente fotos da natureza, como parte de seu trabalho como fotógrafo, em suas redes sociais, e isso pode não refletir sua personalidade. Neste caso, um APR baseado em imagem poderia facilmente prever erroneamente a personalidade de tal usuário analisando suas fotos compartilhadas sem considerar o contexto de tais fotos (DHELIM et al., 2021).

Segundo Huang et al. (2020), a frequência com que os indivíduos usam certas categorias de palavras pode fornecer pistas para características psicológicas. Há teorias da psicologia da linguagem afirmando que padrões de uso das palavras relacionam-se forte e significativamente aos traços de personalidade do indivíduo (HIRSH; PETERSON, 2009). Isso inspirou a categoria de APR baseado em texto.

Uma das técnicas proeminentes mais comuns e com precisão aceitável é o LIWC (*Linguistic Inquiry and Word Count*), que categoriza o texto analisado em vários conjuntos (“*buckets*”) psicologicamente relevantes (tais como ‘*social processes*’ e ‘*affective processes*’), e mede a frequência das palavras em cada um desses “*buckets*” para prever os traços de personalidade do indivíduo (TAUSCZIK; PENNEBAKER, 2010) (MEHTA et al., 2020).

Recentemente, Yang, Lee e Yeh (2020) propuseram um esquema automático para identificar os traços do *Big Five* usando técnicas de mineração de texto. Nesse sentido, um estudo bastante referenciado é o de Schwartz et al. (2013), que analisaram 700 milhões de palavras, frases e tópicos coletados de 75 mil usuários voluntários do Facebook. Esses voluntários também responderam ao NEO-IPIP e a acurácia foi de 91,9%.

O estudo de Schwartz et al. (2013) serviu de base para o desenvolvimento de uma antiga ferramenta on-line de análise semântica das postagens do Facebook, a Five Labs, criada pela startup americana Five (FIVE, 2015). Ademais, a pesquisa de Schwartz et al. (2013) é

referência para o *Watson Personality Insights*, da IBM (2021), comentado a seguir.

### **O método de APR via *Watson Personality Insights***

O *Watson Personality Insights* infere características de personalidade (dimensões e aspectos dos modelos *Big Five*, *Needs* e *Values*) a partir de informações textuais (BADENES et al., 2014), refletindo tendências recentes na pesquisa sobre inferência de personalidade, baseada na psicologia, psicolinguística e marketing. A IBM (2021) afirmou que “descobriu que as características que são inferidas do texto geralmente podem prever com segurança o comportamento real”.

Há pesquisadores, tais como Liu et al. (2016) e Liu et al. (2017), que empregaram o *Watson Personality Insights* e ratificaram que a personalidade inferida pode prever o comportamento no mundo real. Arnoux et al. (2017), por exemplo, mostraram o *Watson Personality Insights* superando outras técnicas para predição dos modelos *Big Five* e *Values*.

Com o serviço *Watson Personality Insights*, são disponibilizadas pontuações (resultados numéricos no intervalo de 0 a 1) para cada uma das características de personalidade (*Big Five*, *Needs* e *Values*). O serviço funciona analisando informações textuais com base em uma abordagem de vocabulário aberto (*open-vocabulary approach*) — converte o texto de entrada em *tokens*, usa uma técnica de *word-embedding* (GloVe<sup>1</sup>) para obter uma representação vetorial para as palavras no texto, e alimenta essa representação em um algoritmo de aprendizado de máquina<sup>2</sup> (IBM, 2021).

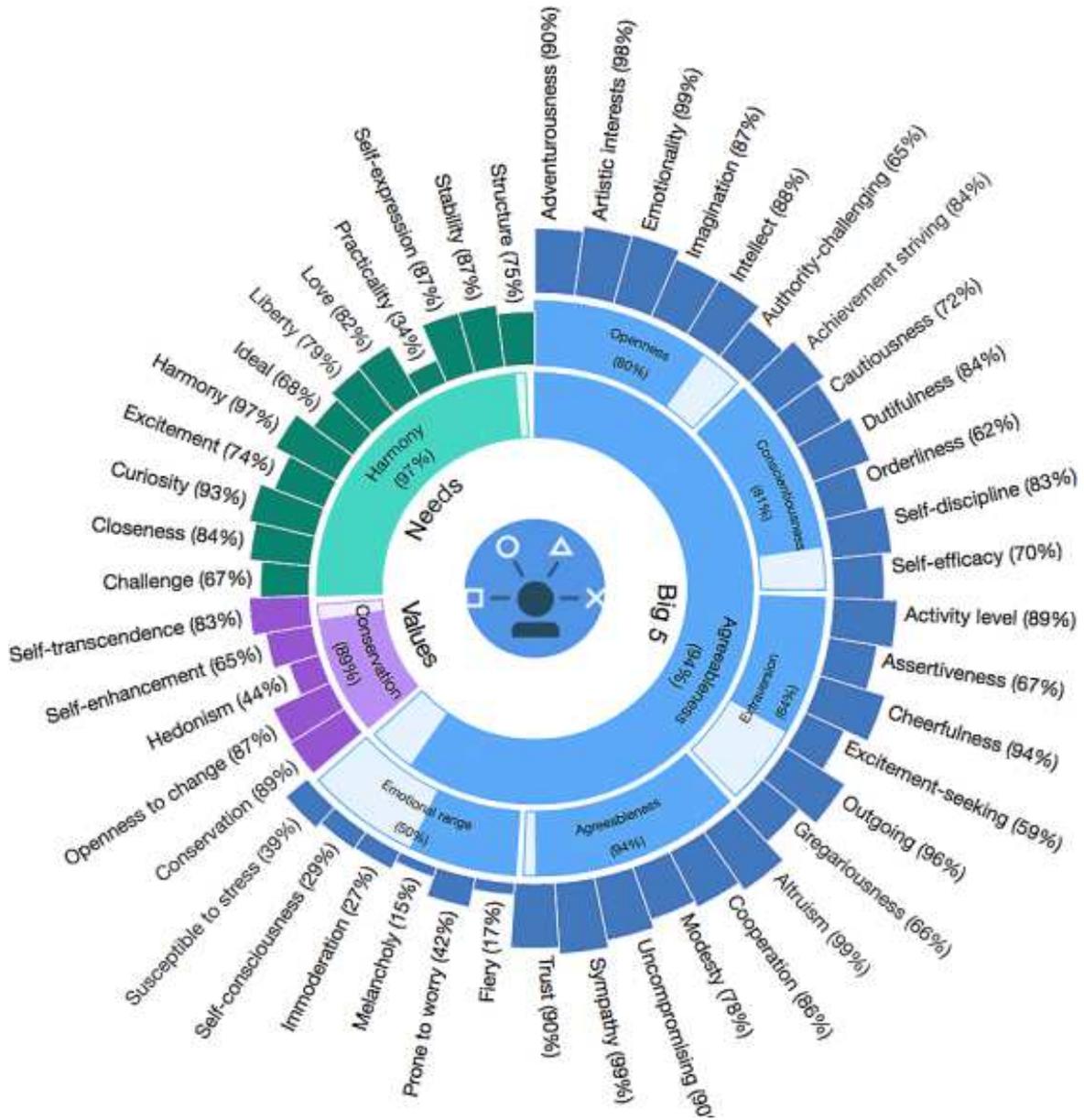
Na Figura 2.1, é apresentado um exemplo de perfil de personalidade, representado graficamente, criado pelo serviço *Watson Personality Insights*. Nesse exemplo, o indivíduo apresenta, dentre outras características, 97% para a dimensão *Harmony* (Harmonia) do modelo *Needs*, 89% para a dimensão *Conservation* (Conservação) do modelo *Values*, 94% para o traço *Agreeableness* (Amabilidade) e 83% para a faceta *Self-discipline* (Autodisciplina) — faceta do traço *Conscientiousness* (Conscienciosidade) — do modelo *Big Five*.

---

<sup>1</sup>GloVe — *Global Vectors for Word Representation* (PENNINGTON; SOCHER; MANNING, 2014) consiste em um algoritmo (software livre) para a obtenção de representações vetoriais de palavras. Disponível em: <<https://nlp.stanford.edu/projects/glove/>>. Acesso em: 30 de out. de 2021.

<sup>2</sup>Na página da IBM (2021), não há detalhes sobre o algoritmo utilizado, mas há informações relevantes neste artigo publicado pelos pesquisadores da IBM: Arnoux et al. (2017). Em síntese, eles usam GloVe (*word-embedding*) combinado com Processos Gaussianos (RASMUSSEN, 2003).

Figura 2.1: Representação gráfica de um perfil de personalidade criado pelo *Watson Personality Insights*.



Fonte: Lamprecht, Ahmad e Robra-Bissantz (2018, p. 547).

## 2.2 Sistemas de Recomendação

Em meio à sobrecarga de informação aliada à evolução da Web, os Sistemas de Recomendação (SR) surgiram e se popularizaram, filtrando os variados conteúdos disponíveis por meio da coleta de informações sobre as preferências de seus usuários para um conjunto de itens (a exemplo de filmes e objetos de aprendizagem), objetivando auxiliar tais usuários no processo de busca e acesso a conteúdo relevante (RICCI; ROKACH; SHAPIRA, 2015) (AGGARWAL et al., 2016) (WANG et al., 2021).

Segundo Bobadilla et al. (2013), o processo para gerar recomendações em SR baseia-se em definir o algoritmo de filtragem a ser utilizado, as técnicas empregadas, o modelo escolhido (com base, por exemplo, na utilização direta de dados — *memory-based* — ou em um modelo gerado usando tais dados — *model-based*), entre outros aspectos. As técnicas mais conhecidas em SR são a Filtragem Baseada no Conteúdo (FBC) e a Filtragem Colaborativa (FC).

O processo de recomendação que emprega FBC utiliza informações prévias do usuário em relação a um item para recomendar itens semelhantes. Tais itens são comparados de acordo com características definidas a eles associadas. Em um SR baseado em FBC, os produtos mais similares àqueles previamente avaliados de maneira positiva serão recomendados (BURKE, 2002) (COSTA; AGUIAR; MAGALHÃES, 2013).

A eficácia dos SR baseados em FBC depende de quão bem os itens são descritos. Portanto, há algumas limitações referentes, por exemplo, à análise de conteúdo de dados pouco estruturados, à complexidade existente na extração/análise de conteúdo não textual (multimídia, por exemplo), e à análise da semântica de textos que, mesmo contendo termos semelhantes, podem ter qualidades distintas (ADOMAVICIUS; TUZHILIN, 2005) (COSTA; AGUIAR; MAGALHÃES, 2013) (PARK, 2019).

O processo de recomendação que emprega FC baseia-se em como os seres humanos tomam decisões, ao considerar o julgamento de usuários com interesses em comum. Na FC, com as avaliações dos itens pelos usuários do sistema, é possível calcular médias para os itens e, assim, sugerir os itens mais bem avaliados pelos usuários com preferências semelhantes, sem precisar entender ou reconhecer as características dos itens (BOBADILLA et al., 2013) (COSTA; AGUIAR; MAGALHÃES, 2013).

Vários pesquisadores comentam a importância de técnicas híbridas de recomendação, que combinam duas ou mais técnicas de recomendação diferentes, explorando as vantagens de cada uma (BOBADILLA et al., 2013) (AGGARWAL et al., 2016) (DHELMIM et al., 2021).

No artigo clássico de Burke (2002), são apresentados os seguintes métodos de hibridização: *Weighted* (ponderado) — as pontuações de várias técnicas são combinadas para produzir uma única recomendação; *Switching* (alternado) — o sistema alterna entre técnicas dependendo da situação atual; *Mixed* (misto) — as recomendações de vários recomendadores diferentes são apresentadas simultaneamente; *Feature combination* (combinação de características) — características de fontes de dados de recomendações diferentes são unidas e dadas como entrada para um único algoritmo de recomendação; *Cascade* (cascata) — um recomendador refina as recomendações dadas por outro, reordenando os itens recomendados; *Feature augmentation* (aumento de características) — a saída de uma técnica é utilizada como uma característica de entrada para outra; e *Meta-level* (Meta-nível) — o modelo aprendido por um recomendador é usado como entrada para outro.

Para avaliar as recomendações/predições, visando a melhorar os métodos e tecnologias em SR, é comum os pesquisadores usarem métricas de Recuperação de Informação (BOBADILLA et al., 2013) (AGGARWAL et al., 2016). No artigo clássico de Herlocker et al. (2004), são apresentadas categorias de métricas de avaliação em SR.

Com as métricas de acurácia de predição (*prediction accuracy metrics*) — a exemplo do RMSE (*Root Mean Square Error*) —, é possível analisar as diferenças entre as notas (*ratings*) reais e as estimadas pelo SR para os itens. Nesse sentido, as abordagens de recomendação com os menores valores de RMSE, por exemplo, representam as abordagens mais próximas à “opinião” dos usuários-alvos das recomendações. A métrica RMSE é comumente empregada em pesquisas que avaliam acurácia em SR — como, por exemplo, Khan et al. (2020) e Wang et al. (2021). A métrica RMSE é calculada como indicado na Equação 2.1.

$$RMSE = \sqrt{\frac{\sum_{(u,i) \in S} (r_{u,i} - \tilde{r}_{u,i})^2}{|S|}}. \quad (2.1)$$

Na Equação 2.1 (e ao longo deste documento),  $r_{u,i}$  representa a nota real que um usuário  $u$  atribuiu a um item  $i$ ,  $\tilde{r}_{u,i}$  representa a nota estimada pelo SR,  $S$  representa o conjunto de todos os pares usuário–item  $(u, i)$  do conjunto de teste e  $|S|$  representa a quantidade desses

pares (tamanho do conjunto  $S$ ).

Com as métricas de acurácia de classificação (*classification accuracy metrics*) — a exemplo de *Precision* (Precisão), *Recall* (Cobertura) e *F-Measure* (Medida F) —, é possível mensurar quão relevante para o usuário é o conjunto de itens recomendados. Nessa categoria, em geral, as abordagens de recomendação com os maiores valores (de *F-Measure*, por exemplo) representam as abordagens mais próximas à “opinião” dos usuários-alvos das recomendações.

Se uma nota  $r_{u,i}$  é menor que um limiar (*threshold*)  $t$ , é possível considerar que o item  $i$  é desinteressante para o usuário  $u$  (caso contrário,  $i$  é interessante para  $u$ ). A partir disso, há estas quatro possibilidades considerando a nota real ( $r_{u,i}$ ) e a nota estimada ( $\tilde{r}_{u,i}$ ):

- Verdadeiro Positivo (TP — *True Positive*);

$$TP = \{(u, i) | (u, i) \in S \wedge r_{u,i} \geq t \wedge \tilde{r}_{u,i} \geq t\}.$$

- Verdadeiro Negativo (TN — *True Negative*);

$$TN = \{(u, i) | (u, i) \in S \wedge r_{u,i} < t \wedge \tilde{r}_{u,i} < t\}.$$

- Falso Positivo (FP — *False Positive*); e

$$FP = \{(u, i) | (u, i) \in S \wedge r_{u,i} < t \wedge \tilde{r}_{u,i} \geq t\}.$$

- Falso Negativo (FN — *False Negative*).

$$FN = \{(u, i) | (u, i) \in S \wedge r_{u,i} \geq t \wedge \tilde{r}_{u,i} < t\}.$$

Com isso, as métricas *Precision*, *Recall* e *F-Measure* podem ser definidas como indicado nas Equações 2.2, 2.3 e 2.4, respectivamente. *Precision* consiste na proporção dos itens recuperados que são relevantes em relação ao total de itens recuperados. *Recall* consiste na razão entre o número de itens recuperados que são relevantes e o total dos itens na coleção que são relevantes. *F-Measure* consiste na média harmônica entre *Precision* e *Recall*.

$$precision = \frac{|TP|}{|TP| + |FP|} = \frac{|\{itens\_recuperados\} \cap \{itens\_relevantes\}|}{|\{itens\_recuperados\}|}. \quad (2.2)$$

$$recall = \frac{|TP|}{|TP| + |FN|} = \frac{|\{itens\_recuperados\} \cap \{itens\_relevantes\}|}{|\{itens\_relevantes\}|}. \quad (2.3)$$

$$F\text{-Measure} = 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}. \quad (2.4)$$

Convém mencionar que, além do uso de métricas de acurácia, há outras maneiras de avaliar as recomendações, como, por exemplo, avaliar a novidade das listas dos itens recomendados. Nessa perspectiva, Zhou et al. (2010) propuseram a métrica MSI (*Mean Self-information*), usada em outras pesquisas (VARGAS; CASTELLS, 2011) (SUÁREZ-GARCÍA et al., 2018), considerando como novidade os itens acessados por menos usuários.

A métrica MSI pode ser calculada como indicado na Equação 2.5 — quanto maior o valor de MSI, mais novidade há nas listas avaliadas. Na Equação 2.5,  $U$  representa o conjunto de todos os usuários,  $L_u$  representa a lista de itens recomendados para determinado usuário  $u$  e  $\textit{count}(i)$  representa o número de usuários que acessaram determinado item  $i$  (em outras palavras, representa a quantidade de interações relacionadas a  $i$ ).

$$MSI = \frac{1}{|U|} \sum_{u \in U} \left( -\frac{1}{|L_u|} \sum_{i \in L_u} \log_2 \frac{\textit{count}(i)}{|U|} \right). \quad (2.5)$$

Apesar da variedade de aspectos a se comentar sobre SR, são destacados nesta seção os tópicos mais importantes para o entendimento desta pesquisa. Nas seguintes subseções, são abordadas a Filtragem Colaborativa e a Recomendação Baseada em Personalidade.

### 2.2.1 Recomendação via Filtragem Colaborativa

A Filtragem Colaborativa (FC) é uma das técnicas mais utilizadas em SR. Em linhas gerais, caracteriza-se por gerar recomendações/predições de itens desconhecidos por um usuário-alvo, mas conhecidos e considerados relevantes por usuários com preferências similares àquele. Os algoritmos de FC podem ser baseados em vizinhança (baseados em memória) ou baseados em modelo (BOBADILLA et al., 2013) (AGGARWAL et al., 2016) (LEX et al., 2021).

Os algoritmos de FC baseados em vizinhança (*neighborhood-based* ou *memory-based*) são bastante populares. Nas abordagens com foco no usuário, uma etapa básica dessa categoria de FC é identificar os usuários mais semelhantes a um usuário-alvo (empregando-se alguma métrica de similaridade). Usualmente, define-se um valor  $K$  para o tamanho da vizinhança, como no popular algoritmo UserKNN (*User-based K-Nearest Neighbors*). Alter-

nativamente, há abordagens com vizinhança enfocada nos itens, como o algoritmo ItemKNN (*Item-based K-Nearest Neighbors*).

Explicando o KNN com base no usuário (UserKNN), inicialmente, calcula-se a similaridade do usuário-alvo em relação aos outros usuários (usando uma métrica como o Coeficiente de Correlação de Pearson, a Distância Euclidiana ou a Similaridade dos Cossenos), em seguida, seleciona-se a vizinhança (grupo de K usuários com as maiores similaridades), e, por fim, são normalizadas as avaliações e computadas as previsões, ponderando as avaliações de tais usuários vizinhos (COSTA; AGUIAR; MAGALHÃES, 2013).

Os algoritmos de FC baseados em modelo (*model-based*) estão sendo cada vez mais utilizados e consistem em aprender um modelo a partir do conjunto de *ratings*, empregando-se métodos de aprendizado de máquina e mineração de dados. Um destaque no estado-da-arte em FC tem sido a utilização de algoritmos baseados em fatoração de matriz (*matrix factorization*). Exemplos desses algoritmos são: SVD++ (*Singular Value Decomposition Plus Plus*), SMF (*Simple Matrix Factorization*) e BMF (*Biased Matrix Factorization*) — com implementações disponíveis na biblioteca MyMediaLite (GANTNER et al., 2011).

Com a fatoração de matriz, em linhas gerais, a matriz de interação usuários-itens é decomposta no produto de duas matrizes retangulares representando os usuários e os itens em um espaço latente de dimensão inferior, e isso é feito aplicando um algoritmo de redução de dimensionalidade, como o *Singular Value Decomposition* (AGGARWAL et al., 2016).

A eficácia dos SR baseados em FC depende do *feedback* dos usuários em relação aos itens. Portanto, destacam-se algumas limitações: primeiro avaliador (uma vez que um novo item não é recomendado até usuários o avaliarem), pontuações esparsas (comum em sistemas com pequeno número de usuários e grande número de itens) e usuário com preferências muito diferentes (uma vez que as recomendações para ele podem ser fracas pela escassez de usuários com preferências similares) (ADOMAVICIUS; TUZHILIN, 2005) (COSTA; AGUIAR; MAGALHÃES, 2013) (PARK, 2019).

### 2.2.2 Recomendação Baseada em Personalidade

Por estar relacionada ao processo de tomada de decisões, alguns pesquisadores — a exemplo de Nunes e Hu (2012) e Tkalčič (2018) — defendem que a personalidade deve ser considerada ao projetar SR. Visto que alguns itens, como filmes ou músicas, estão fortemente

relacionados à psique humana, alguns pesquisadores de SR têm considerado a personalidade do usuário no processo de geração de recomendações, sendo o modelo *Big Five* o mais usado (CHEN; WU; HE, 2016) (DHELMIM et al., 2021) (LEX et al., 2021).

Além de abordar a Filtragem Híbrida e as técnicas de FBC e FC, Burke (2002) definiu a Filtragem Demográfica, a Filtragem Baseada em Conhecimento e a Filtragem Baseada em Utilidade. González et al. (2007), ao expandir a categorização de Burke (2002), propuseram uma filtragem envolvendo aspectos psicológicos do usuário, tais como Inteligência Emocional e Interação Social, combinados aos aspectos caracterizadores dos itens. Nunes (2008), por sua vez, expandiu este modelo, apresentando a Personalidade do usuário como um fator relevante no processo de recomendação; este estudo é considerado o primeiro a usar traços de personalidade em SR, revelando indícios de que o uso de traços de personalidade proporciona melhoria na recomendação (NUNES, 2008) (NUNES, 2009).

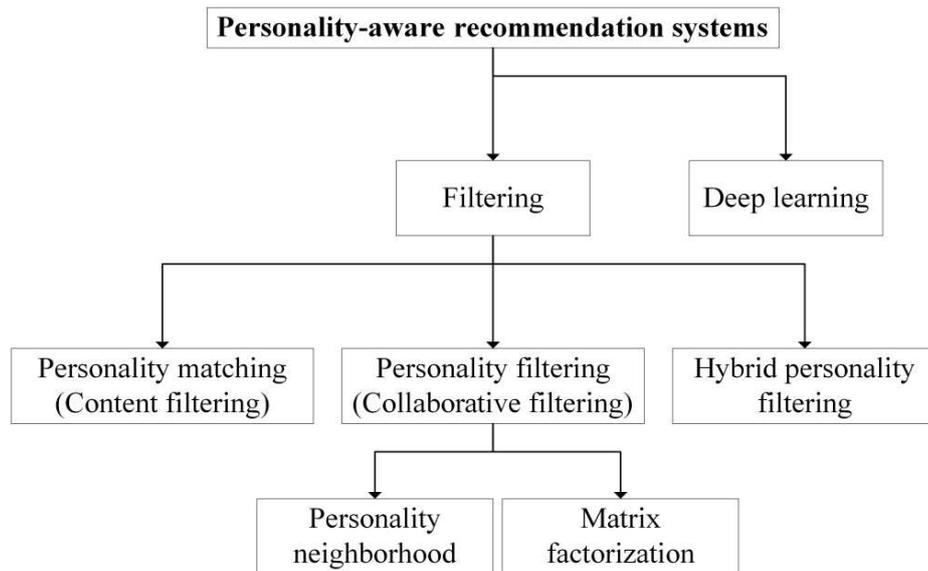
Para viabilizar a recomendação baseada em personalidade, é aplicada alguma técnica (como a FC e/ou FBC) aliada à adição de dados da personalidade do usuário no processo, sendo preciso definir como representá-la (a exemplo do uso do modelo *Big Five*) e como identificá-la (via questionário ou APR), como comentado na Seção 2.1.

É possível classificar os SR baseados em personalidade (*personality-based recommender systems* ou *personality-aware recommendation systems*) a partir da técnica de recomendação, como definido por Dhelim et al. (2021): sistemas empregando *deep learning* e sistemas baseados em filtragem (*filtering*), sendo estes subdivididos, como ilustrado na Figura 2.2, em filtragem colaborativa (*personality filtering*), filtragem baseada no conteúdo (*personality matching*) ou filtragem híbrida (*hybrid personality filtering*).

Recentemente, *deep learning* tem sido usado em SR, com modelos como, por exemplo, Rede Neural Convolucional (*Convolutional Neural Network* — CNN) e Codificadores Automáticos Variacionais (*Variational AutoEncoders* — VAE). Em SR baseados em personalidade, *deep learning* pode ser usado no próprio processo de recomendação ou no processo de identificação da personalidade dos usuários.

A categoria *personality matching* é semelhante à FBC. O SR calcula a correspondência entre determinado usuário e alguns itens. Tal correspondência é calculada com base nos dados de personalidade do usuário e algumas características do item, aplicando um reconhecimento automático de personalidade sobre o conteúdo desse item, como, por exemplo, o

Figura 2.2: Classificação dos sistemas de recomendação baseados em personalidade.



Fonte: Dhelim et al. (2021, p. 14).

tipo de personalidade dos atores (no caso de recomendação de filmes), a descrição textual do item, seus rótulos e/ou categorias. *Personality matching* é geralmente aplicada se for possível observar um critério de correspondência comum entre o conteúdo recomendado e o usuário-alvo.

A categoria *personality filtering* compreende os métodos que aliam os dados de personalidade à FC, podendo ser divididos em métodos de fatoração de matriz e métodos baseados em vizinhança. Nos métodos de fatoração de matriz aprimorada com dados de personalidade (*personality-enhanced matrix factorization*), estende-se a fatoração de matriz convencional para incorporar, com seus *ratings*, os traços de personalidade do usuário. Por outro lado, a filtragem via vizinhança baseada em personalidade (*personality neighborhood filtering*) é a técnica de recomendação mais comum em SR baseados em personalidade, incorporando-se os dados de personalidade no cálculo de similaridade para determinar os vizinhos de cada usuário. A estratégia para definir a vizinhança (usuários mais similares) concernente à personalidade baseia-se na inferência de que indivíduos com personalidade similar terão interesses e padrões comportamentais similares (HU; PU, 2011).

Por fim, a categoria *hybrid personality filtering* envolve os métodos que combinam *personality filtering* e *personality matching*, mostrando-se útil ao aproveitar as vantagens de ambos os métodos.

## 2.3 Outros Fundamentos

Considerando-se os conceitos apresentados nas seções anteriores, compreende-se que estratégias de aprendizado de máquina (*Machine Learning* — ML) são úteis na recomendação de itens.

ML contempla a questão de como construir sistemas computacionais que melhoram automaticamente com a experiência. Como um campo de estudo, ML situa-se na interseção da ciência da computação, estatística e uma variedade de outras disciplinas (incluindo o estudo psicológico da aprendizagem humana, o estudo da evolução, neurociência e economia) (JORDAN; MITCHELL, 2015).

Um conjunto diversificado de algoritmos de ML foi desenvolvido para cobrir a ampla variedade de dados e tipos de problemas. Os SR (comentados na Seção 2.2) são sistemas de ML baseados em dados que indicam *links* entre um conjunto de usuários e um conjunto de itens; no caso, o problema consiste em sugerir itens de interesse para determinado usuário, baseando-se nos dados de todos os usuários (JORDAN; MITCHELL, 2015).

Além dos aspectos relacionados a ML apresentados nas seções anteriores, nesta pesquisa foram empregados os algoritmos comentados nas Subseções 2.3.1 e 2.3.2.

### 2.3.1 *K-means*

O *k-means* é um dos algoritmos mais utilizados para agrupamento (*clustering*). *Clustering* é um método de ML não supervisionado que realiza a atribuição de um conjunto de observações a subconjuntos (*clusters*), de forma que as observações dentro de um mesmo *cluster* são similares de acordo com algum(ns) critério(s) predefinido(s), enquanto observações feitas em *clusters* diferentes não são similares (HAN; KAMBER; PEI, 2012).

Dado um conjunto de observações, o objetivo do algoritmo *k-means* é agrupá-las em *k clusters* (grupos disjuntos), de modo que seja minimizada a soma das distâncias entre os elementos pertencentes a um *cluster* e seu respectivo centro<sup>3</sup> (HAN; KAMBER; PEI, 2012).

Dentre as implementações possíveis do *k-means*, foi considerado nesta pesquisa o algoritmo de Hartigan e Wong (1979) — via R Core Team (2021). Em síntese, o algoritmo *k-means* começa com a inicialização de um conjunto de *k* centroides, um para cada *cluster*,

---

<sup>3</sup>O centro de um *cluster* representa o ponto médio dos pontos pertencentes a determinado agrupamento.

sendo comumente utilizada, para tal inicialização, uma seleção aleatória entre os pontos da base de dados. Após a seleção inicial dos centroides, para formar os agrupamentos, cada ponto é associado ao centroide mais próximo com base em uma medida de similaridade, a exemplo da Distância Euclidiana. Em seguida, os centroides são recalculados — o novo centroide é a média dos antigos pontos do *cluster*. Tal processo é repetido até que, ao finalizar uma iteração, nenhum centroide seja alterado.

### 2.3.2 Algoritmos Genéticos

Algoritmos Genéticos (GA — *Genetic Algorithms*) são técnicas de busca, baseadas em heurísticas, inspiradas pelo processo biológico de evolução natural, e visam a resolver problemas por meio da criação/evolução de populações de indivíduos, que vão se reproduzindo e sendo avaliados. Cada indivíduo (cromossomo) em um GA consiste em uma possível solução para o problema, sendo comumente representado (representação cromossomial) por cadeias de bits (GOLDBERG, 1989) (LINDEN, 2012).

De acordo com Linden (2012), o funcionamento de um GA pode ser resumido por meio das seguintes instruções: **(i)** inicialize a população de cromossomos (geração inicial, criada aleatoriamente); **(ii)** avalie cada cromossomo na população usando uma função de *fitness*, útil para selecionar os indivíduos melhor adaptados, isto é, as melhores soluções até o momento; **(iii)** selecione os pais para gerar novos cromossomos; **(iv)** aplique os operadores de *crossover* e mutação aos pais para gerar os indivíduos da nova geração; **(v)** apague os velhos membros da população, sendo também possível manter, para a próxima geração, aqueles indivíduos que correspondem à melhor solução para o problema até o momento (seleção elitista); **(vi)** avalie todos os novos cromossomos e insira-os na população; **(vii)** se um limite pré-definido de gerações foi atingido ou o melhor cromossomo satisfaz os requerimentos e desempenho, retorne tal cromossomo, caso contrário, volte à instrução **iii**.

## 2.4 Considerações Finais do Capítulo

Uma vez que a personalidade pode ser considerada um fator duradouro, previsível e estável que determina os comportamentos, preferências e interesses humanos, as características de personalidade são frequentemente empregadas em SR com base na suposição de que in-

divíduos com traços de personalidade semelhantes terão interesses e padrões de comportamento semelhantes (ATAS et al., 2021).

Apesar da variedade de tópicos a se explicar sobre ML, SR e características de personalidade, foram enfocados, nas seções deste capítulo, os conceitos fulcrais para o entendimento desta tese.

No capítulo seguinte, é apresentada uma revisão referente a pesquisas relacionadas a esta tese. Com isso, são realizados mais comentários concernentes a SR baseados em personalidade.

# Capítulo 3

## Pesquisas Relacionadas

Neste capítulo, é apresentado um levantamento bibliográfico do estado-da-arte, destacando-se as pesquisas relacionadas à tese ora descrita.

O número de publicações científicas no campo da Computação da Personalidade (*Personality Computing*) aumentou consideravelmente na última década. Há evidências, nessas publicações, que alavancar as teorias da personalidade pode ajudar a resolver alguns dos problemas de sistemas de recomendação (SR). Por exemplo, a personalidade pode ser aproveitada para resolver o problema de partida a frio (*cold-start*), além de aumentar a percepção de confiança dos usuários nesses sistemas — em relação aos SR não baseados em personalidade —, e também contribuir para explorar a diversidade dos itens dentro de um conjunto de recomendações (HU; PU, 2009) (HU; PU, 2011) (TKALČIČ et al., 2011) (NUNES; HU, 2012) (CHEN; WU; HE, 2016) (NGUYEN et al., 2018) (DHELIM et al., 2021) (LEX et al., 2021).

O termo *Personality Computing* foi cunhado em um *survey* referente a sistemas computacionais que usam a personalidade do usuário (VINCIARELLI; MOHAMMADI, 2014). Entretanto, não há muitos *surveys*, nesse campo, principalmente no contexto de recomendações. Sobre o estado-da-arte em SR baseados na personalidade, Nunes e Hu (2012) e Cantador e Fernández-Tobías (2014) apresentaram visões gerais (publicações curtas), e Tkalčič e Chen (2015) apresentaram um capítulo no livro *Recommender Systems Handbook* (RICCI; ROKACH; SHAPIRA, 2015). Nessas três publicações, os autores comentaram também desafios e questões em aberto.

Em um recente *survey*, Lex et al. (2021) objetivaram apresentar uma revisão completa

do estado-da-arte dos SR que consideram construtos e teorias da psicologia para modelar e prever o comportamento dos usuários (denominando-os *psychology-informed recommender systems*), incluindo os SR baseados em personalidade.

Tran, Felfernig e Tintarev (2021) realizaram uma análise baseada em uma revisão sistemática da literatura sobre fatores psicológicos em SR. Eles forneceram exemplos e apontaram questões abertas de pesquisa referentes à inclusão de fatores psicológicos em SR.

A pesquisa de Dhelim et al. (2021) consiste em uma revisão abrangente da literatura de SR baseados em personalidade. Eles visaram a estudar e classificar sistematicamente esses sistemas, e comentaram que sua pesquisa, recentemente publicada, é o primeiro *survey* com foco nesse tipo de SR.

Na Seção 3.1, são elencadas pesquisas referenciadas nesses estudos citados nos parágrafos anteriores, além de outras pesquisas encontradas consideradas relevantes, tanto em âmbito internacional quanto nacional.

### **3.1 Pesquisas sobre Recomendação Baseada em Personalidade**

Na literatura, há diversas pesquisas relativas a SR. Dentre esses sistemas, há aqueles considerando a personalidade dos usuários no processo de recomendação, em diferentes domínios de aplicação.

Em relação a redes sociais, por exemplo, um domínio expressivo é a recomendação de amigos (TOMMASEL et al., 2015) (TOMMASEL et al., 2015) (TOMMASEL et al., 2016) (MUKTA; ALI; MAHMUD, 2016) (XIAO; FAN; DU, 2018) (NEEHAL; MOTTALIB, 2019) (NING; DHELIM; AUNG, 2019) (CHAKRABARTY et al., 2020) (SHAYEGAN; VALIZADEH, 2020).

Bastante populares na área de SR de modo geral, estes dois domínios também recebem destaque por parte dos pesquisadores em SR baseados em personalidade: o domínio de filmes/vídeos (RECIO-GARCIA et al., 2009) (QUIJANO-SANCHEZ; RECIO-GARCIA; DIAZ-AGUDO, 2010) (QUIJANO-SANCHEZ; RECIO-GARCIA; DIAZ-AGUDO, 2011) (WU; CHEN; HE, 2013) (WU; CHEN, 2015) (WU, 2017) (POTASH; RUMSHISKY, 2016) (YI; LEE; JUNG, 2016) (NALMPANTIS; TJORTJIS, 2017) (BALAKRISHNAN;

ARABI, 2018) (WU; CHEN; ZHAO, 2018) (ASABERE; ACAKPOVI, 2020) (BOLOCK et al., 2020) (KHAN et al., 2020); e o domínio de músicas (HU, 2010) (HU; PU, 2010a) (CHENG; TANG, 2016) (FERWERDA et al., 2016) (FERWERDA; TKALČIČ; SCHEDL, 2017a) (FERWERDA; TKALČIČ; SCHEDL, 2017b) (ONORI; MICARELLI; SANSONETTI, 2016) (SCHEDL et al., 2016) (LU; TINTAREV, 2018) (FERWERDA et al., 2019) (GUPTA et al., 2020) (LIU; HU, 2020) (MOSCATO; PICARIELLO; SPERLI, 2020).

Quanto a traços de personalidade e características de imagens, a associação é dupla: conhecer as características da imagem pode ajudar a inferir a personalidade dos usuários que interagem com a imagem, e conhecer os traços de personalidade dos usuários pode ajudar a recomendar imagens relevantes. Assim, é conveniente destacar também o domínio de recomendação de imagens (GUNTUKU; ROY; WEISI, 2015) (GELLI et al., 2017) (LI et al., 2019) (LI et al., 2020).

Considerando-se a importância de SR em comércio eletrônico, há propostas aplicando personalidade na recomendação de produtos (BOLOGNA et al., 2013) (HU; PU, 2014) (ADAMOPOULOS; TODRI, 2015) (ROFFO; VINCIARELLI, 2016) (ROFFO, 2016) (BUETTNER, 2017) (PAIVA; COSTA; SILVA, 2017) (LIU et al., 2019) (BEHESHTI et al., 2020) (HUANG et al., 2020) (KHODABANDEHLOU; GOLPAYEGANI; RAHMAN, 2020) (PATRO et al., 2020) (SRIVASTAVA; BALA; KUMAR, 2020).

Há também o uso de traços de personalidade para SR educacionais, a exemplo de recomendações de cursos, de participantes de conferências, de artigos de pesquisa, de estratégias pedagógicas e de objetos de aprendizagem (NUNES et al., 2010) (TREVISAN; CAZELLA; NUNES, 2011) (SANTOS et al., 2014) (UDDIN; BANERJEE; LEE, 2016) (BARVINSKI et al., 2017) (XIA et al., 2017) (ASABERE; ACAKPOVI; MICHAEL, 2018) (GIANOTTI; CAZELLA; BEHAR, 2019) (ZHENG; SUBRAMANIYAN, 2020) (QAMHIEH; SAMMANEH; DEMAIDI, 2020).

Outro domínio com SR baseados em personalidade é o de turismo — pontos de interesse (TANASESCU et al., 2013) (BRAUNHOFER; ELAHI; RICCI, 2014) (BRAUNHOFER et al., 2014) (BRAUNHOFER; ELAHI; RICCI, 2015) (ROSHCHINA; CARDIFF; ROSSO, 2015) (HAFSHEJANI; KAEDI; FATEMI, 2018) (ALVES et al., 2020a) (JEONG; LEE; JUNG, 2020).

Além dos citados anteriormente, há outros domínios que receberam contribuições refe-

rentes à incorporação de aspectos da personalidade no processo de recomendação — jogos (YANG et al., 2017) (CHAN; ARYA; WHITEHEAD, 2018) (YANG; HUANG, 2019); artigos de notícias (DHELIM; AUNG; NING, 2020) (DHELIM et al., 2021); empregos (TING; VARATHAN, 2018); receitas (ADAJI et al., 2018); e atividades diárias (KHWAJA et al., 2019).

Muitas dessas pesquisas envolvendo SR baseados em personalidade foram propostas com uso de algum questionário para identificar características da personalidade dos usuários (NUNES, 2008) (NUNES, 2009) (RECIO-GARCIA et al., 2009) (TKALČIČ et al., 2009) (TKALČIČ et al., 2011) (HU; PU, 2010b) (HU; PU, 2011) (HU, 2010) (HU; PU, 2010a) (NUNES et al., 2010) (QUIJANO-SANCHEZ; RECIO-GARCIA; DIAZ-AGUDO, 2010) (QUIJANO-SANCHEZ; RECIO-GARCIA; DIAZ-AGUDO, 2011) (TREVISAN; CAZELLA; NUNES, 2011) (ELAHI et al., 2013) (QUAN, 2013) (WU; CHEN; HE, 2013) (ALHARTHI; TRAN, 2014) (ALHARTHI, 2015) (BRAUNHOFER; ELAHI; RICCI, 2014) (BRAUNHOFER et al., 2014) (BRAUNHOFER; ELAHI; RICCI, 2015) (FERNÁNDEZ-TOBÍAS; CANTADOR, 2014) (FERNÁNDEZ-TOBÍAS; CANTADOR, 2015) (HU; PU, 2014) (SANTOS et al., 2014) (WANG, 2015) (CHENG; TANG, 2016) (FERNÁNDEZ-TOBÍAS et al., 2016) (FERWERDA et al., 2016) (FERWERDA; TKALČIČ; SCHEDL, 2017a) (FERWERDA; TKALČIČ; SCHEDL, 2017b) (MUKTA; ALI; MAHMUD, 2016) (ROFFO; VINCIARELLI, 2016) (ROFFO, 2016) (SCHEDL et al., 2016) (YI; LEE; JUNG, 2016) (BARVINSKI et al., 2017) (NALMPANTIS; TJORTJIS, 2017) (PAIVA; COSTA; SILVA, 2017) (BALAKRISHNAN; ARABI, 2018) (CHAN; ARYA; WHITEHEAD, 2018) (HAFSHEJANI; KAEDI; FATEMI, 2018) (LU; TINTAREV, 2018) (SUN; REN; XU, 2018) (WU; CHEN; ZHAO, 2018) (FERWERDA et al., 2019) (GIANOTTI; CAZELLA; BEHAR, 2019) (KHWAJA et al., 2019) (NEEHAL; MOTTALIB, 2019) (NING; DHELIM; AUNG, 2019) (ZHENG; SUBRAMANIYAN, 2020) (ALVES et al., 2020a) (BOLOCK et al., 2020) (DHELIM; AUNG; NING, 2020) (JEONG; LEE; JUNG, 2020) (KHODABANDEHLOU; GOLPAYEGANI; RAHMAN, 2020) (LIU; HU, 2020) (PATRO et al., 2020) (QAMHIEH; SAMMANEH; DEMAIDI, 2020) (DHELIM et al., 2021).

Devido à necessidade de se ter dados da personalidade dos usuários, não há muitos conjuntos de dados disponíveis publicamente, com muitos usuários, para avaliar propostas de recomendação baseadas em personalidade. Lex et al. (2021), em sua visão geral sobre con-

juntos de dados em SR, citam o Personality2018 (único conjunto de dados citado em tal estudo que contém dados de personalidade). Tal conjunto de dados também é citado por Dhelim et al. (2021), que também destacam o myPersonality. O Personality2018 (com dados dos usuários do MovieLens) e o myPersonality (com dados dos usuários do Facebook), possuem dados de personalidade segundo o modelo *Big Five* — o Personality2018 baseou-se no questionário TIPI e o myPersonality no questionário IPIP 336-*version*. Em 2018, os criadores do myPersonality decidiram interromper o projeto. Os avanços em APR facilitam a obtenção de dados da personalidade dos usuários, em conjuntos de dados públicos, para realizar experimentos.

Dentre as pesquisas que propõem estratégias de recomendação baseadas em personalidade, muitas aplicam FC, principalmente baseada em vizinhança — *personality neighborhood* (NUNES, 2008) (NUNES, 2009) (RECIO-GARCIA et al., 2009) (TKALČIČ et al., 2009) (TKALČIČ et al., 2011) (HU; PU, 2010b) (HU; PU, 2011) (HU, 2010) (HU; PU, 2010a) (NUNES et al., 2010) (QUIJANO-SANCHEZ; RECIO-GARCIA; DIAZ-AGUDO, 2010) (QUIJANO-SANCHEZ; RECIO-GARCIA; DIAZ-AGUDO, 2011) (TREVISAN; CAZELLA; NUNES, 2011) (QUAN, 2013) (ALHARTHI; TRAN, 2014) (ALHARTHI, 2015) (FERNÁNDEZ-TOBÍAS; CANTADOR, 2014) (FERNÁNDEZ-TOBÍAS; CANTADOR, 2015) (GUNTUKU; ROY; WEISI, 2015) (ROSHCHINA; CARDIFF; ROSSO, 2015) (TOMMASEL et al., 2015) (TOMMASEL et al., 2015) (TOMMASEL et al., 2016) (WANG, 2015) (WU; CHEN, 2015) (WU, 2017) (SCHEDL et al., 2016) (YI; LEE; JUNG, 2016) (NALMPANTIS; TJORTJIS, 2017) (SRIVASTAVA; BALA; KUMAR, 2017) (XIA et al., 2017) (ADAJI et al., 2018) (ASABERE; ACAKPOVI; MICHAEL, 2018) (BALAKRISHNAN; ARABI, 2018) (HAFSHEJANI; KAEDI; FATEMI, 2018) (SUN; REN; XU, 2018) (WU; CHEN; ZHAO, 2018) (ASABERE; ACAKPOVI, 2020) (BOLOCK et al., 2020) (CHAKRABARTY et al., 2020) (DHELM; AUNG; NING, 2020) (KHODABANDEH-LOU; GOLPAYEGANI; RAHMAN, 2020) (PATRO et al., 2020) (SRIVASTAVA; BALA; KUMAR, 2020).

Convém citar que, embora a aplicação de *personality neighborhood* considere geralmente a vizinhança entre usuários, algumas pesquisas consideraram o conceito de personalidade focando-o para os itens em vez dos usuários — Tanasescu et al. (2013), Alharthi (2015) e Wang (2015).

Dentre os grupos de pesquisa que iniciaram os estudos em SR baseados em personalidade, é possível destacar Tkalčič et al. e Hu e Pu, bastante referenciados, e destacados por Chen, Wu e He (2016) — ao abordarem as pesquisas que usam personalidade na recomendação baseada em FC —, além de destacados por Alves et al. (2020b) — ao comentarem os SR baseados em personalidade em uma revisão referente à incorporação da personalidade no projeto de interface do usuário. Tkalčič et al. (2009) e Hu e Pu (2011), em suas conclusões, abordaram que os métodos baseados em personalidade (FC usando o *Big Five* para o cálculo da similaridade entre usuários) foram melhores (ou no mínimo equivalentes) quando comparados à abordagem tradicional (*rating-based*).

Analogamente a Tkalčič et al. (2009), Hu e Pu (2010b), Tkalčič et al. (2011) e Hu e Pu (2011) — que usaram o *Big Five* para definir a vizinhança em FC —, Srivastava, Bala e Kumar (2017) utilizaram o modelo *Values*, e os resultados indicaram desempenho melhor que a abordagem tradicional (sem uso de personalidade). Entretanto, eles não compararam sua abordagem com outra abordagem baseada em personalidade. Tais autores, recentemente, apresentaram abordagens, baseando-se nos modelos *Big Five*, *Values* e *Needs*, para identificação de usuários com “preferências excêntricas” (*gray sheep*), um dos problemas em SR (SRIVASTAVA; BALA; KUMAR, 2020). Em relação à forma de identificação da personalidade, Tkalčič et al. e Hu e Pu, em suas pesquisas, usaram questionários; Srivastava, Bala e Kumar usaram o *Watson Personality Insights*.

Destaca-se que Srivastava, Bala e Kumar (2017), ao direcionar estudos futuros para a melhoria da eficiência dos algoritmos de FC, indicaram que os componentes do modelo *Values* podem ser combinados com outros modelos, além de incorporar os *ratings* como em algoritmos tradicionais de FC. Nesse sentido, a investigação descrita nesta tese também visou a analisar se a influência de outros modelos (*Values* e *Needs*), aliados ao *Big Five* (modelo geralmente empregado), tendem a melhorar as previsões/recomendações, em cenários gerais.

Embora seja mais comum a realização de estudos incorporando FC e traços de personalidade envolvendo algoritmos baseados em vizinhança, há também os que consideram algoritmos baseados em fatoração de matriz (ELAHI et al., 2013) (BRAUNHOFER; ELAHI; RICCI, 2014) (BRAUNHOFER et al., 2014) (BRAUNHOFER; ELAHI; RICCI, 2015) (ADAMOPOULOS; TODRI, 2015) (FERNÁNDEZ-TOBÍAS et al., 2016) (POTASH; RUMSHISKY, 2016) (ROFFO, 2016) (ZHENG; SUBRAMANIYAN, 2020) (YAKHCHI et

al., 2020) (WANG et al., 2021).

Alguns pesquisadores não focaram especificamente em realizar uma proposta para recomendação baseada em personalidade, mas realizaram algumas análises sobre personalidade visando à aplicação e a melhoramentos de estudos futuros em SR. É o caso de Cantador, Fernández-Tobías e Bellogín (2013), que usaram o conjunto de dados myPersonality, Golbeck e Norris (2013), que usaram o questionário BFI, e Karumur e Konstan (2016), Karumur, Nguyen e Konstan (2016) e Karumur, Nguyen e Konstan (2018), que usaram o questionário TIPI.

É conveniente comentar que, ao realizar a revisão da literatura, foram encontradas pesquisas que usaram o termo *personality* em SR, mas não da forma empregada nesta pesquisa (baseado em algum modelo específico, como o *Big Five*). É o caso da pesquisa de Bian et al. (2012), que consideraram características gerais de personagens de TV como personalidade, e também é o caso das pesquisas de Zhou et al. (2011), Feng e Qian (2013), Hariadi e Nurjanah (2017), Zhang et al. (2018) e Wang et al. (2020), que consideraram como personalidade determinado comportamento dos usuários no sistema, sem explicitar o uso de algum modelo embasado na psicologia.

No Quadro 3.1, em caráter comparativo, são resumidamente apresentadas, além da pesquisa proposta nesta tese, as pesquisas relacionadas (incluindo as anteriormente mencionadas), que apresentaram propostas de recomendação baseada em personalidade (empregando algum modelo embasado na psicologia).

Quadro 3.1: Comparação entre pesquisas relacionadas.

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Nunes (2008), Nunes (2009)	<i>Big Five</i>	NEO-IPIP; TIPI	<i>Personality neighborhood</i>	Não específico
Recio-Garcia et al. (2009)	TKI	TKI <i>test</i>	<i>Personality neighborhood</i>	Filmes
Tkalčič et al. (2009), Tkalčič et al. (2011)	<i>Big Five</i>	IPIP 50- <i>version</i>	<i>Personality neighborhood</i>	Não específico

(*Continua na próxima página...*)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Hu e Pu (2010b), Hu e Pu (2011)	<i>Big Five</i>	TIPI	<i>Personality neighborhood</i>	Não específico
Hu (2010), Hu e Pu (2010a)	<i>Big Five</i>	TIPI	<i>Personality neighborhood</i>	Músicas
Nunes et al. (2010)	<i>Big Five</i>	NEO-IPIP	<i>Personality neighborhood</i>	Educacional
Quijano-Sanchez, Recio-Garcia e Diaz-Agudo (2010), Quijano-Sanchez, Recio-Garcia e Diaz-Agudo (2011)	TKI	TKI test	<i>Personality neighborhood</i>	Filmes
Trevisan, Cazella e Nunes (2011)	<i>Big Five</i>	NEO-IPIP; TIPI	<i>Personality neighborhood</i>	Programas de TV / Educacional
Bologna et al. (2013)	RIASEC	Mapeamento de profissões à personalidade	<i>Personality matching</i>	Produtos
Elahi et al. (2013)	<i>Big Five</i>	TIPI	<i>Matrix factorization</i>	Não específico
Quan (2013)	RIASEC	<i>Holland Code Test</i>	<i>Personality neighborhood</i>	Não específico
Tanasescu et al. (2013)	<i>Big Five</i>	<i>Text-based APR</i> (método próprio)	<i>Personality matching</i>	Pontos de Interesse (turismo)
Wu, Chen e He (2013)	<i>Big Five</i>	NEO-IPIP	<i>Personality matching</i>	Filmes
Alharthi e Tran (2014), Alharthi (2015)	<i>Big Five</i>	BFI-44	<i>Personality neighborhood</i>	Não específico

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Braunhofer, Elahi e Ricci (2014), Braunhofer et al. (2014), Braunhofer, Elahi e Ricci (2015)	<i>Big Five</i>	TIPI; FIPI	<i>Matrix factorization</i>	Pontos de Interesse (turismo)
Fernández-Tobías e Cantador (2014), Fernández-Tobías e Cantador (2015)	<i>Big Five</i>	IPIP	<i>Personality neighborhood</i>	Não específico
Hu e Pu (2014)	<i>Big Five</i>	<i>California Q-Set</i>	<i>Personality matching</i>	Produtos
Santos et al. (2014)	<i>Big Five</i>	BFI	<i>Personality matching</i>	Educacional
Adamopoulos e Todri (2015)	<i>Big Five, Values, Needs</i>	<i>Text-based APR (Watson Personality Insights)</i>	<i>Matrix factorization</i>	Produtos
Guntuku, Roy e Weisi (2015)	<i>Big Five</i>	BFI-10; <i>Multimedia-based APR (imagens)</i>	<i>Personality neighborhood</i>	Imagens
Roshchina, Cardiff e Rosso (2015)	<i>Big Five</i>	<i>Text-based APR (Mairesse et al. (2007))</i>	<i>Personality neighborhood</i>	Pontos de Interesse (turismo)
Tommasel et al. (2015), Tommasel et al. (2015), Tommasel et al. (2016)	<i>Big Five</i>	<i>Text-based APR (Mairesse et al. (2007) e SMOreg)</i>	<i>Personality neighborhood</i>	Amigos
Wang (2015)	<i>Big Five</i>	TIPI	<i>Personality neighborhood</i>	Não específico

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Wu e Chen (2015), Wu (2017)	<i>Big Five</i>	<i>Behavior-based APR</i> (método próprio)	<i>Personality neighborhood</i>	Filmes
Cheng e Tang (2016)	<i>Big Five</i>	BFI-44	<i>Personality matching</i>	Músicas
Fernández-Tobías et al. (2016)	<i>Big Five</i>	IPIP	<i>Matrix factorization</i>	Não específico
Ferwerda et al. (2016), Ferwerda, Tkalčič e Schedl (2017a), Ferwerda, Tkalčič e Schedl (2017b)	<i>Big Five</i>	BFI-44	<i>Personality matching</i>	Músicas
Mukta, Ali e Mahmud (2016)	<i>Big Five</i>	IPIP 44-version	<i>Deep learning</i>	Amigos
Onori, Micarelli e Sansonetti (2016)	<i>Big Five</i>	BFI-44; <i>Behavior-based APR (Apply Magic Sauce)</i>	<i>Personality neighborhood; Personality matching</i>	Músicas
Potash e Rumshisky (2016)	<i>Big Five</i>	<i>Text-based APR</i> (método próprio)	<i>Matrix factorization</i>	Filmes
Roffo e Vinciarelli (2016)	<i>Big Five</i>	BFI-10	<i>Personality matching</i>	Produtos
Roffo (2016)	<i>Big Five</i>	BFI-10	<i>Matrix factorization</i>	Produtos
Schedl et al. (2016)	<i>Big Five</i>	BFI-44	<i>Personality neighborhood</i>	Músicas
Uddin, Banerjee e Lee (2016)	<i>Big Five</i>	<i>Behavior-based APR</i>	<i>Personality matching</i>	Educacional

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Yi, Lee e Jung (2016)	MBTI	MBTI	<i>Personality neighborhood</i>	Filmes
Barvinski et al. (2017)	<i>Big Five</i>	NEO-IPIP; TIPI	<i>Personality matching</i>	Educacional
Buettner (2017)	<i>Big Five</i>	TIPI; BFI-S; <i>Behavior-based APR (Personality Prediction Engine)</i>	<i>Personality matching</i>	Produtos
Gelli et al. (2017)	<i>Big Five</i>	<i>Text-based APR (Apply Magic Sauce)</i>	<i>Personality matching</i>	Imagens
Nalmpantis e Tjortjis (2017)	<i>Big Five</i>	IPIP 50-version	<i>Personality neighborhood</i>	Filmes
Paiva, Costa e Silva (2017)	<i>Big Five</i>	BFI-10	<i>Hybrid personality filtering</i>	Produtos
Srivastava, Bala e Kumar (2017)	<i>Values</i>	<i>Text-based APR (Watson Personality Insights)</i>	<i>Personality neighborhood</i>	Não específico
Xia et al. (2017)	<i>Big Five</i>	Instrumento próprio (notas 1–5)	<i>Personality neighborhood</i>	Educacional
Yang et al. (2017)	<i>Big Five</i>	<i>Text-based APR (Mairesse et al. (2007))</i>	<i>Personality matching</i>	Jogos
Adaji et al. (2018)	<i>Big Five</i>	<i>Text-based APR (LIWC)</i>	<i>Personality neighborhood</i>	Receitas

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Asabere, Acakpovi e Michael (2018)	<i>Big Five</i>	Instrumento próprio (notas 1–5)	<i>Personality neighborhood</i>	Educacional
Balakrishnan e Arabi (2018)	<i>Big Five</i>	BFI-44	<i>Personality neighborhood</i>	Filmes
Chan, Arya e Whitehead (2018)	<i>Big Five</i>	TIPI	<i>Personality matching</i>	Jogos
Hafshejani, Kaedi e Fatiemi (2018)	<i>Big Five</i>	NEO-FFI-60	<i>Personality neighborhood</i>	Pontos de Interesse (turismo)
Lu e Tintarev (2018)	<i>Big Five</i>	TIPI	<i>Personality matching</i>	Músicas
Sun, Ren e Xu (2018)	<i>Big Five</i>	TIPI	<i>Personality neighborhood</i>	Não específico
Ting e Varathan (2018)	<i>Big Five</i>	BFI-44; <i>Behavior-based APR (Facebook personality scores)</i>	<i>Personality matching</i>	Empregos
Wu, Chen e Zhao (2018)	<i>Big Five</i>	BFI-44	<i>Personality neighborhood</i>	Filmes
Xiao, Fan e Du (2018)	<i>Big Five</i>	<i>Text-based APR (método próprio baseado em Mairesse et al. (2007) e TextMind)</i>	<i>Personality matching</i>	Amigos
Ferwerda et al. (2019)	<i>Big Five</i>	BFI-44	<i>Personality matching</i>	Músicas
Gianotti, Cazella e Behar (2019)	<i>Big Five</i>	NEO-IPIP	<i>Hybrid personality filtering</i>	Educacional

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Khwaja et al. (2019)	<i>Big Five</i>	IPIP	<i>Personality matching</i>	Atividades diárias
Li et al. (2019), Li et al. (2020)	<i>Big Five</i>	BFI-10; <i>Multimedia-based APR</i> (imagens)	<i>Deep learning</i>	Imagens
Liu et al. (2019)	<i>Big Five, Values</i>	<i>Text-based APR</i> (método próprio)	<i>Deep learning</i>	Produtos
Neehal e Mottalib (2019)	<i>Big Five</i>	IPIP 50-version	<i>Deep learning</i>	Amigos
Ning, Dhelim e Aung (2019)	<i>Big Five</i>	NEO-PI-R-60	<i>Hybrid personality filtering</i>	Amigos
Yang e Huang (2019)	<i>Big Five</i>	<i>Text-based APR</i> (Mairesse et al. (2007) e método próprio)	<i>Hybrid personality filtering</i>	Jogos
Alves et al. (2020a)	<i>Big Five</i>	BFI-44	<i>Personality matching</i>	Pontos de Interesse (turismo)
Asabere e Acakpovi (2020)	<i>Big Five</i>	Instrumento próprio (notas 1–5)	<i>Personality neighborhood</i>	Programas de TV
Beheshti et al. (2020)	<i>Big Five</i>	<i>Text-based APR</i> (LIWC)	<i>Deep learning</i>	Produtos
Bolock et al. (2020)	<i>Big Five</i>	NEO-PI-R	<i>Personality neighborhood</i>	Filmes

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Chakrabarty et al. (2020)	<i>Big Five</i>	Instrumento próprio (cada usuário informa um valor para si em cada traço)	<i>Personality neighborhood</i>	Amigos
Dhelim, Aung e Ning (2020)	<i>Big Five</i>	TIPI	<i>Personality neighborhood</i>	Artigos de notícias
Gupta et al. (2020)	<i>Big Five</i>	<i>Behavior-based APR</i> (método próprio)	<i>Personality matching</i>	Músicas
Huang et al. (2020)	<i>Big Five, Lifestyles</i>	<i>Text-based APR</i> (método próprio)	<i>Personality matching</i>	Produtos
Jeong, Lee e Jung (2020)	MBTI ( <i>Introversion/Extraversion</i> )	MBTI	<i>Deep learning</i>	Pontos de Interesse (turismo)
Khan et al. (2020)	<i>Big Five, Values</i>	<i>Text-based APR</i> ( <i>Watson Personality Insights</i> )	<i>Personality matching</i>	Filmes
Khodabandehlou, Golpayegani e Rahman (2020)	<i>Big Five</i>	NEO-PI-R-60	<i>Personality neighborhood</i>	Produtos
Liu e Hu (2020)	<i>Big Five</i>	TIPI	<i>Personality matching;</i> <i>Deep learning</i>	Músicas
Moscato, Picariello e Sperli (2020)	<i>Big Five</i>	BFI-10; <i>Behavior-based APR</i> (método próprio)	<i>Personality matching;</i> <i>Deep learning</i>	Músicas
Patro et al. (2020)	TKI	TKI <i>test</i>	<i>Personality neighborhood</i>	Produtos

(Continua na próxima página...)

Quadro 3.1: Comparação entre pesquisas relacionadas. (Continuação do quadro...)

<b>Pesquisa</b>	<b>Modelo de Personalidade</b>	<b>Identificação da Personalidade</b>	<b>Método de Recomendação</b>	<b>Domínio</b>
Qamhieh, Sammaneh e Demaidi (2020)	MBTI	MBTI	<i>Deep learning</i>	Educacional
Shayegan e Valizadeh (2020)	<i>Big Five</i>	NEO-FFI-60; <i>Text-based APR</i> (LIWC)	<i>Deep learning</i>	Amigos (Canais do Telegram)
Srivastava, Bala e Kumar (2020)	<i>Big Five, Values, Needs</i>	<i>Text-based APR</i> ( <i>Watson Personality Insights</i> )	<i>Personality neighborhood</i>	Produtos
Yakhchi et al. (2020)	<i>Big Five</i>	<i>Text-based APR</i> (LIWC)	<i>Matrix factorization</i>	Não específico
Zheng e Subramaniyan (2020)	<i>Big Five</i>	TIPI	<i>Matrix factorization</i>	Educacional
Dhelim et al. (2021)	<i>Big Five</i>	TIPI	<i>Hybrid personality filtering</i>	Artigos de notícias
Wang et al. (2021)	<i>Big Five</i>	<i>Text-based APR</i> (método próprio)	<i>Matrix factorization</i>	Não específico
ESTA PESQUISA	<i>Big Five, Values, Needs</i>	<i>Text-based APR</i> ( <i>Watson Personality Insights</i> )	<i>Personality neighborhood;</i> <i>Hybrid personality filtering</i>	Não específico

(Fim do quadro.)

Fonte: autoria própria.

Todas as pesquisas elencadas no Quadro 3.1, direta ou indiretamente, têm como motivação a questão que também motiva esta pesquisa — “Como melhorar as recomendações geradas pelos SR para oferecer informações, produtos ou serviços mais personalizados para os usuários?”. Nesse sentido, tais pesquisas são relacionadas a esta tese no que se refere à QP4, pois, em geral, apresentam e geralmente avaliam alguma estratégia visando à melhoria das recomendações, embora nem todas com foco em FC e no uso de APR baseado em texto.

Em relação à QP1, poucas das pesquisas empregando APR mostraram comparações com as estratégias iniciais da área, para analisar o desempenho de tais propostas pioneiras.

Quanto à QP2, não foram encontradas estratégias aplicando aspectos de personalidade via FC de modo genérico (em que fosse possível substituir facilmente o algoritmo de FC por algum mais recente do estado-da-arte). Além disso, não foram encontradas pesquisas conjecturando que todos os algoritmos de FC já consideram implicitamente a personalidade (sobretudo utilizando algum modelo consolidado na literatura, como o *Big Five*).

Em referência à QP3, pelo conhecimento do autor desta tese, a partir do levantamento bibliográfico realizado, este é o primeiro estudo com foco na análise da influência — individual, agrupada e ponderada — dos traços de personalidade do modelo *Big Five*, sobretudo abrangendo, na investigação, as dimensões dos modelos *Needs* e *Values*, além das facetas por traço do *Big Five*.

É pertinente destacar que, apesar dos avanços em SR envolvendo aprendizado de máquina (*Machine Learning* — ML) e *deep learning*, continuam sendo realizados estudos referentes a abordagens mais clássicas de FC considerando a personalidade dos usuários. É possível compreender *deep learning* como um subgrupo específico de técnicas de ML, e comumente ML é empregado considerando muitos dados. Nesse sentido, abordagens baseadas em vizinhança definida via personalidade tornam-se relevantes para empregar no processo de recomendação em cenários bastante comuns (a exemplo de sistemas com quantidade de usuários não muito expressiva e entrada de novos usuários no sistema). Além disso, destaca-se que Dacrema, Cremonesi e Jannach (2019) indicaram que algoritmos de recomendação baseados em vizinhança superaram, em vários casos analisados, algoritmos de *deep learning* (*neural algorithms*) recentemente publicados em conferências de alto nível.

Por fim, analisando os estudos listados nas referências das pesquisas de Dhelim et al. (2021), Lex et al. (2021) e Tran, Felfernig e Tintarev (2021) — além de outros estudos encontrados referentes à recomendação baseada em personalidade (sumarizados no Quadro 3.1) —, é possível dizer que não foram encontradas investigações com os mesmos propósitos da investigação realizada nesta tese, reforçando-se seu diferencial. Esta pesquisa contribui com o avanço do estado-da-arte especialmente por não focar em questionários de personalidade nem focar na melhoria de apenas cenários específicos (melhoria em um domínio particu-

lar, ou melhoria da acurácia em situações de *cold-start*, ou melhoria da diversidade)<sup>1</sup>. Uma vez que há pesquisas em SR indicando o *trade-off* entre acurácia e diversidade/novidade (por exemplo, melhorias na diversidade das recomendações desencadeando decréscimo na acurácia), é pertinente obter resultados, como realizado nesta tese, referentes a manter — e até aumentar — a acurácia das recomendações, em cenários gerais, ao considerar características de personalidade dos usuários (reconhecidas automaticamente com base em textos), dado que isto possibilita melhorias em aspectos como a diversidade<sup>2</sup> das recomendações.

## 3.2 Considerações Finais do Capítulo

Estratégias de recomendação de produtos/serviços baseadas na personalidade dos usuários têm sido propostas, mas ainda não é tão comum avaliá-las na perspectiva de APR, em especial com uma ferramenta disponível no mercado para o uso facilitado por parte das empresas que usam SR. Percebe-se que há vários estudos recentes que ainda continuam apenas empregando questionários padronizados.

Como parte da investigação realizada e para a avaliação experimental apresentada no Capítulo 5, foram considerados algoritmos de algumas pesquisas comentadas neste Capítulo 3. No próximo capítulo, são apresentadas as abordagens elaboradas, referentes às questões de pesquisa desta tese, incluindo-se as estratégias propostas considerando os aspectos destacados na última linha do Quadro 3.1.

---

<sup>1</sup>Pela revisão da literatura realizada, foi visto que as pesquisas referentes à recomendação baseada em personalidade, em geral, enfocam domínios específicos (especialmente *personality matching* com músicas e filmes), ou enfocam em melhorar aspectos diferentes da acurácia (como os estudos mais recentes explorando a diversidade das recomendações), ou enfocam em melhorar a acurácia em cenários de *cold-start* (tópico sempre destacado pelos estudos que revisam a literatura sobre SR baseados em personalidade). Isso, por exemplo, é reforçado por Lex et al. (2021, p. 34): “*The most common motivations for considering personality in the recommendation process include to alleviate cold-start situations (in particular for new users) and to improve the level of personalization (e.g., to increase recommendation list diversity)*”.

<sup>2</sup>Apesar desta tese não focar no aspecto da diversidade das recomendações, convém comentar, de maneira geral, como características da personalidade podem se relacionar à diversidade dos itens. No artigo intitulado *How Personality Influences Users' Needs for Recommendation Diversity?*, Chen, Wu e He (2013) indicaram que, quanto à diversidade global, pessoas com baixos índices de Conscienciosidade preferiram alto nível de diversidade; além disso, tais pesquisadores indicaram que alguns traços de personalidade têm impacto significativo na preferência de diversidade dos usuários sobre atributos específicos dos itens (filmes) — por exemplo: em relação a ator/atriz, sua diversidade é preferida por usuários imaginativos/criativos (traço Abertura), e pessoas mais reativas e nervosas (traço Neuroticismo) estão mais inclinadas a escolher diretores diversos.

## Capítulo 4

# Abordagens Elaboradas

Com base no referencial teórico, nos estudos relacionados e nas questões de pesquisa, apresentados nos capítulos anteriores, neste capítulo são comentadas as abordagens elaboradas concernentes à recomendação via FC baseada em personalidade.

Para todas as abordagens variantes baseadas em vizinhança, foi utilizada, em geral, a medida de Similaridade dos Cossenos, empregada em estudos como Srivastava, Bala e Kumar (2017) e Yang e Huang (2019), e aplicou-se  $K = 5$ , assim como definido por Hu e Pu (2011).

De forma genérica, nas estratégias propostas, cada usuário possui um perfil associado representado por um par ordenado, formalizado na Equação 4.1.

$$u = (R, P). \quad (4.1)$$

Para a Equação 4.1, tem-se que:

- $u$  representa um usuário;
- $u \in U$ ;
- $U$  representa o conjunto de todos os usuários do sistema;
- $R$  representa as avaliações (*ratings*) realizadas por determinado usuário para itens do sistema;
- $R = \{(i, r) | i \in I \wedge r \in \mathbb{N} \wedge 1 \leq r \leq 5\}$ ;
- $I$  representa o conjunto de todos os itens do sistema;
- $\forall((i_A, r_1), (i_B, r_2)), (i_A, r_1) \in R \wedge (i_B, r_2) \in R, i_A \neq i_B$ ;
- $|R| \leq |I|$ ;

- $P$  representa o perfil do usuário relativo às características de personalidade;
- $P = (T, F, V, N)$ ;
- $T$  representa o perfil do usuário associado aos cinco traços do modelo *Big Five*;
- $T = (t_1, t_2, t_3, t_4, t_5)$ ;
- $0 \leq t_x \leq 1 \wedge t_x \in \mathbb{R} \wedge x \in \{1, 2, 3, 4, 5\}$ ;

Sendo:  $t_1$  referente ao traço Abertura,  $t_2$  referente ao traço Conscienciosidade,  $t_3$  referente ao traço Extroversão,  $t_4$  referente ao traço Amabilidade e  $t_5$  referente ao traço Neuroticismo;

- $F$  representa o perfil do usuário associado às seis facetas de cada traço do modelo *Big Five*;
- $F = (f_1, f_2, f_3, \dots, f_{30})$ ;  $|F| = 30$ ;
- $0 \leq f_y \leq 1 \wedge f_y \in \mathbb{R} \wedge y \in \{1, 2, 3, \dots, 29, 30\}$ ;

Sendo:  $f_1$  referente à faceta Aventura,  $f_2$  referente à faceta Interesses artísticos,  $f_3$  referente à faceta Emotividade,  $f_4$  referente à faceta Imaginação,  $f_5$  referente à faceta Intelecto,  $f_6$  referente à faceta Liberalismo;  $f_7$  referente à faceta Esforço por realizações,  $f_8$  referente à faceta Cautela,  $f_9$  referente à faceta Senso de dever,  $f_{10}$  referente à faceta Ordem,  $f_{11}$  referente à faceta Autodisciplina,  $f_{12}$  referente à faceta Competência;  $f_{13}$  referente à faceta Atividade,  $f_{14}$  referente à faceta Assertividade,  $f_{15}$  referente à faceta Emoções positivas,  $f_{16}$  referente à faceta Busca por entusiasmo,  $f_{17}$  referente à faceta Simpatia,  $f_{18}$  referente à faceta Gregarismo;  $f_{19}$  referente à faceta Altruísmo,  $f_{20}$  referente à faceta Cooperação,  $f_{21}$  referente à faceta Modéstia,  $f_{22}$  referente à faceta Franqueza,  $f_{23}$  referente à faceta Complacência,  $f_{24}$  referente à faceta Confiança;  $f_{25}$  referente à faceta Hostilidade,  $f_{26}$  referente à faceta Ansiedade,  $f_{27}$  referente à faceta Depressão,  $f_{28}$  referente à faceta Impulsividade,  $f_{29}$  referente à faceta Autoconsciência e  $f_{30}$  referente à faceta Vulnerabilidade ao estresse;

- $V$  representa o perfil do usuário associado aos cinco componentes do modelo *Values*;
- $V = (v_1, v_2, v_3, v_4, v_5)$ ;
- $0 \leq v_x \leq 1 \wedge v_x \in \mathbb{R} \wedge x \in \{1, 2, 3, 4, 5\}$ ;

Sendo:  $v_1$  referente ao componente Conservação,  $v_2$  referente ao componente Entusiasmo,  $v_3$  referente ao componente Hedonismo,  $v_4$  referente ao componente Auto-aprimoramento e  $v_5$  referente ao componente Autotranscendência;

- $N$  representa o perfil do usuário associado aos doze componentes do modelo *Needs*;
- $N = (n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}, n_{11}, n_{12})$ ;
- $0 \leq n_z \leq 1 \wedge n_z \in \mathbb{R} \wedge z \in \{1, 2, 3, \dots, 11, 12\}$ ;

Sendo:  $n_1$  referente ao componente Desafio,  $n_2$  referente ao componente Proximidade,  $n_3$  referente ao componente Curiosidade,  $n_4$  referente ao componente Entusiasmo,  $n_5$  referente ao componente Harmonia,  $n_6$  referente ao componente Ideal,  $n_7$  referente ao componente Liberdade,  $n_8$  referente ao componente Amor,  $n_9$  referente ao componente Praticidade,  $n_{10}$  referente ao componente Autoexpressão,  $n_{11}$  referente ao componente Estabilidade e  $n_{12}$  referente ao componente Estrutura;

O perfil de cada usuário  $u$  contempla as avaliações deste para os itens (*ratings* em uma escala de 1 a 5), representadas pelo conjunto  $R$ . O conjunto  $R$  é formado por pares ordenados, que representam o item avaliado e a avaliação correspondente.

Considerando-se os modelos *Big Five* (incluindo facetas), *Needs* e *Values*, para a identificação das características de personalidade dos usuários,  $P$  representa uma quádrupla ordenada formada por tuplas ( $T$ ,  $F$ ,  $V$  e  $N$ ). Tais tuplas possuem elementos representativos do valor percentual (valor do tipo real entre 0 e 1) referente a cada um dos componentes dos modelos, para determinado usuário.

## 4.1 Abordagens referentes à QP1

Com base na revisão da literatura, foram empregados os estudos de Tkalčič et al. (2009) e Hu e Pu (2011) para a análise da questão de pesquisa QP1, devido à relevância desses pesquisadores para a área.

Considerando que a personalidade do usuário não muda significativamente ao longo do tempo e, portanto, os vizinhos mais próximos do usuário podem ser calculados com antecedência, Tkalčič et al. (2009) propuseram uma abordagem de recomendação baseada em personalidade (seguindo o modelo *Big Five*) utilizando FC. Em sua abordagem, eles consideraram também a média geral dos *ratings* (de todos os usuários), apresentando duas variações: “public” (que considera um peso maior para a opinião de todos os outros usuários) e “neighbours” (que considera um peso maior para a opinião dos “vizinhos”).

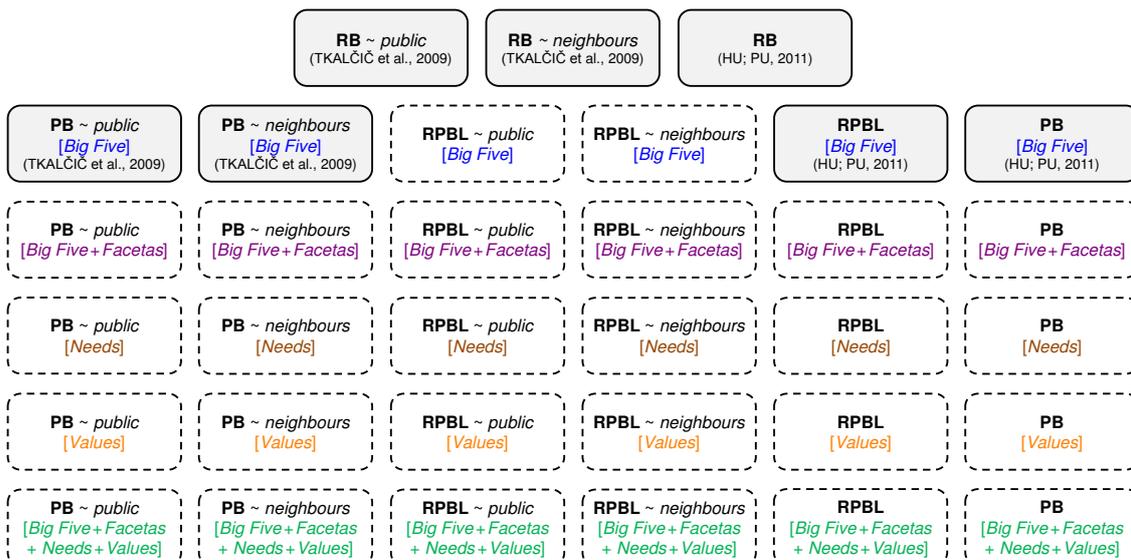
Hu e Pu (2011), além da abordagem tradicional *rating-based* (denominada RB) e da abor-

dagem *personality-based* “pura” (denominada PB), também compararam algumas outras estratégias híbridas elaboradas: RPBL (estratégia linear, combinando personalidade e *ratings* para calcular a similaridade entre usuários) e RPBC (estratégia cascata, usando o PB para estimar inicialmente os *ratings* para completar a matriz usuários-itens e, em seguida, aplicar o método RB na matriz densificada). Mesmo indicando melhor resultado com RPBC, as próprias autoras comentam que essa estratégia é custosa computacionalmente (demandando 2,4 vezes mais tempo que RB), devido à necessidade, inicialmente, de tornar densa a matriz usuários-itens (que, na prática, é bastante esparsa).

Embora RPBL tenha sido elaborada no estudo de Hu e Pu (2011), para a questão de pesquisa QP1, foram elaboradas abordagens variantes, empregando RPBL para as variações “neighbours” e “public” de Tkalcíč et al. (2009). Além disso, considerando o uso do *Watson Personality Insights* para o reconhecimento automático de personalidade, foram elaboradas novas abordagens quanto às características de personalidade utilizadas no cálculo da similaridade (abrangendo as facetas do *Big Five*, além dos modelos *Values* e *Needs*).

Considerando-se os comentários dos parágrafos anteriores, na Figura 4.1, em síntese, são mostradas as abordagens existentes consideradas na investigação referente à QP1 (com realce na cor cinza e bordas em linha contínua) e, baseadas nelas, as abordagens elaboradas (com linha tracejada nas bordas).

Figura 4.1: Abordagens de recomendação consideradas na investigação referente à QP1.



Fonte: autoria própria.

## 4.2 Abordagens referentes à QP2

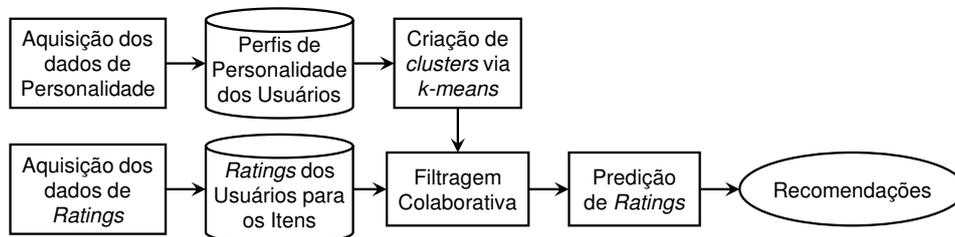
Tendo-se em vista a essência dos algoritmos de FC (que realizam recomendações/predições baseadas nas preferências anteriores dos usuários sobre os itens disponíveis), um elemento motivador desta tese, como comentado na Seção 1.1, é o entendimento de que a personalidade dos usuários é um fator que influencia todos os algoritmos de FC implicitamente (visto que utilizam os *likes* e/ou *ratings* atribuídos pelos usuários, cuja atribuição foi tacitamente influenciada pelos seus traços de personalidade).

A partir dessa ideia, como comentado na Seção 1.1, uma hipótese levantada nesta pesquisa é que a qualidade da recomendação/predição não diminuiria (e poderia aumentar) se, na execução de determinado algoritmo de FC (na fase de “treinamento”), forem sobretudo considerados os “dados de opinião” para os itens (por exemplo: *likes* ou *ratings* de 1 a 5) de usuários mais similares quanto à personalidade.

Portanto, visando-se a responder à questão de pesquisa QP2 (Seção 1.2), referente ao contexto do parágrafo anterior, foi elaborada uma abordagem com o uso de dados explícitos da personalidade dos usuários em algoritmos de FC (independentemente da categoria de FC), enfocando a “colaboração” (dados de opinião — *ratings*) de apenas usuários de um mesmo grupo quanto a seus traços de personalidade.

Assim, foi utilizado um processo de agrupamento (*clustering*) para desconsiderar os *ratings* de usuários não tão similares quanto à personalidade de um usuário-alvo de recomendação. Na Figura 4.2, é ilustrada a abordagem.

Figura 4.2: Representação geral da abordagem elaborada referente à QP2 (“tps”).



Fonte: autoria própria.

Como cada fator do modelo *Big Five* pode ser caracterizado por adjetivos em dois polos — polo do rótulo ou polo oposto, como apresentado em Jaques e Nunes (2021) —, foi definida a criação de dois grupos (*clusters*) por traço. Isto é, para executar determinado al-

goritmo de FC em relação a determinado usuário-alvo, foram previamente desconsideradas as opiniões (*ratings*) dos usuários menos similares em relação a cada traço; e, por fim, utilizada a média das execuções referentes aos cinco *clusters* dos quais determinado usuário-alvo participa.

Uma forma de definir os dois *clusters*, por traço, seria simplesmente agrupar os usuários com base em valores maiores e menores que 50%. Entretanto, as pessoas não são uniformemente distribuídas quanto aos traços de personalidade, podendo, inclusive, em uma certa população, estarem todas caracterizadas por valores acima de 50% em relação a algum(ns) traço(s), mas ainda se diferindo em relação ao nível percentual (mais próximo ou não de 100%). Assim, decidiu-se usar o algoritmo *k-means* (HARTIGAN; WONG, 1979) para definir os dois *clusters* por traço.

Considerando-se a abordagem descrita nos parágrafos anteriores, e tornando-a genérica (para usar modelos além do *Big Five*), o cálculo referente à predição de um *rating* de determinado usuário  $u$  para determinado item  $i$  ( $\tilde{r}_{u,i}$ ) foi formalmente definido na Equação 4.2.

$$\tilde{r}_{u,i} = \frac{1}{|M|} \sum_{x=1}^{|M|} \widetilde{r_{FC\_G_x u, i}}. \quad (4.2)$$

Na Equação 4.2,  $M \in \{T, V, N\}$ , isto é,  $M$  representa um dos perfis do usuário formalizados na tupla  $P$  (Equação 4.1), sendo  $|M| = 5$ , se  $M = T$  (modelo *Big Five*) ou  $M = V$  (modelo *Values*), ou  $|M| = 12$  se  $M = N$  (modelo *Needs*).  $FC\_G_x$  representa a execução de um algoritmo de FC (por exemplo, o SVD++) considerando apenas os dados oriundos dos usuários pertencentes ao *cluster*  $G_x$ , grupo ao qual o usuário  $u$  pertence. O usuário  $u$  pertence a  $|M|$  *clusters* (dentre os  $2 \times |M|$  *clusters* criados via *k-means*) —  $G_1$  representa o *cluster* que o usuário  $u$  pertence em relação ao primeiro componente do modelo em uso,  $G_2$  representa o *cluster* que  $u$  pertence em relação ao segundo componente, e assim sucessivamente. Dessa forma,  $\widetilde{r_{FC\_G_x u, i}}$  representa o *rating* estimado por determinado algoritmo de FC com base no *cluster*  $G_x$ , sendo  $u \in G_x$ .

### 4.3 Abordagens referentes à QP3

Para analisar o impacto dos traços de personalidade nas recomendações, cada perfil de usuário consistiu em um vetor numérico de tamanho até 5 (cinco), com os valores percentuais para cada traço do modelo *Big Five*. Foram variados os tamanhos dos vetores representativos dos usuários, retirando valores para analisar se algum traço (ou conjunto de traços) se destaca em relação a considerar todos os traços do *Big Five*. Nesse contexto, empregou-se a Distância Euclidiana para o cálculo da similaridade entre os perfis dos usuários, visto que tal medida funciona bem com vetores de tamanho 1 (um), diferentemente da medida de Similaridade dos Cossenos, por exemplo.

Assim, foram geradas primeiramente 31 variações do algoritmo de recomendação baseado em traços de personalidade: uma variação representando a abordagem tradicional (incluindo todos os traços), variações com os cinco traços individualmente (no caso, cinco variações) e as combinações de 2 a 4 traços. Depois, foram geradas variações com ponderações distintas. Devido às inúmeras ponderações que poderiam ser analisadas, foi planejada a estratégia de variar pesos para os traços de 20% em 20% (120 variações).

Essas 151 variações foram nomeadas com o padrão `***-***-***-***-***`, em que `***` representa um valor numérico inteiro entre 0 e 100, indicando o percentual de influência de um traço (na ordem em que aparece). Os valores são respectivos ao acrônimo OCEAN (como também é conhecido o *Big Five*): primeiro valor para o traço O (*Openness*), segundo valor para o traço C (*Conscientiousness*), terceiro valor para o traço E (*Extraversion*), quarto valor para o traço A (*Agreeableness*) e quinto valor para o traço N (*Neuroticism*). Os valores representados pelos asteriscos somam 100, representando 100% (com exceção de algumas variações somando 99, para representar os casos com apenas três traços de mesmo peso).

Para facilitar o entendimento, seguem alguns exemplos: a variação nomeada `020-060-000-020-000` consiste na variação que desconsidera os traços E e N (0% para cada) e considera 20% para o traço O, 60% para o traço C e 20% para o traço A; a variação nomeada `000-000-100-000-000` consiste na variação que considera unicamente o traço E; a variação nomeada `033-033-033-000-000` consiste na variação que desconsidera os traços A e N, dando o mesmo peso aos traços O, C e E; e a variação nomeada `020-020-020-020-020` representa o que é tradicionalmente empregado, com o mesmo peso para cada traço.

Além dessa análise inicial, foram analisadas também outras variações com ponderações, mas empregando sempre todos os traços. Para isso, foram variados os pesos em níveis de 10% (126 variações). Por exemplo, a variação nomeada 010-010-040-020-020 consiste em uma variação com menor influência para os traços O e C (10% para cada), sendo o peso maior para o traço E (40%), e os traços A e N com peso 20%, cada.

Além disso, foi elaborada uma estratégia para analisar variações considerando além dos cinco traços do *Big Five* — incluindo os componentes dos modelos *Needs* e *Values*. Atualmente, a maioria dos pesquisadores que estuda os traços de personalidade concorda com o número cinco (para os traços dominantes) em relação à questão “quantos traços ou disposições pessoais o indivíduo possui?” (FEIST; FEIST; ROBERTS, 2015). Assim, foram analisadas combinações de 5 componentes. Inicialmente, isso totalizou 26.334 combinações, dado que  $(22!)/(5! \times (22 - 5)!) = 26.334$ , sendo 22 a soma dos componentes do *Big Five* (5), *Needs* (12) e *Values* (5). Adicionando as 30 facetas do *Big Five* (6 facetas por traço), tem-se  $(52!)/(5! \times (52 - 5)!) = 2.598.960$  combinações, sendo 52 a soma dos componentes do *Big Five* (5+30), *Needs* (12) e *Values* (5).

Dado que analisar todas essas combinações seria custoso, então foi aplicado um Algoritmo Genético (GA — *Genetic Algorithm*) para analisar melhorias em relação à combinação clássica do *Big Five*. O GA empregado foi fundamentado na obra de Linden (2012). Os indivíduos foram representados como cadeias de 52 bits, sendo cada bit referente a uma característica de personalidade. A função de *fitness* consistiu no cálculo da métrica RMSE. A taxa de *crossover* foi de 80% e a taxa de mutação foi de aproximadamente 2% (1/52). Foi empregada uma seleção elitista, mantendo-se os 5 melhores indivíduos da geração anterior e gerando-se 200 novos. O critério de parada consistiu em alcançar o limite de 2.000 gerações ou passar 10 gerações seguidas sem encontrar uma solução melhor.

## 4.4 Abordagens referentes à QP4

Nesta seção, é descrita a elaboração de estratégias para a geração de recomendações personalizadas baseadas nas características de personalidade dos usuários e suas preferências anteriores. Mais especificamente, objetivou-se incorporar estratégias aos algoritmos do estado-da-arte de FC (utilizando explicitamente dados de personalidade, identificados automaticamente

via APR baseado em texto) para obter recomendações mais relevantes, com base nos resultados dos estudos anteriores (referentes às questões de pesquisa QP1, QP2 e QP3).

Considerando-se as poucas pesquisas no estado-da-arte referentes à “personalidade dos itens”, foi proposta uma estratégia similar à abordagem elaborada na Seção 4.2, mas realizando o pré-agrupamento com base nessa “personalidade dos itens” (Subseção 4.4.1).

Pelos resultados experimentais referentes à QP2, houve melhoria referente aos algoritmos baseados em vizinhança. Assim, a abordagem descrita na Seção 4.2 também se relaciona à QP4. Todavia, ainda em relação aos resultados experimentais referentes à QP2, houve casos de piora na acurácia. Então, em relação à QP4, foram planejadas estratégias híbridas (*ensemble design*) para analisar a melhoria da acurácia de abordagens baseadas em modelos aliadas a abordagens baseadas em vizinhos com personalidade similar: uma estratégia *weighted* (Subseção 4.4.2) e uma estratégia *cascade re-ranking* (Subseção 4.4.3).

Inicialmente, planejava-se propor uma estratégia de recomendação ponderando as características de personalidade (empregando, por exemplo, um GA, como similarmente comentado na Seção 4.3), mas a proposta de tal estratégia não foi formalizada devido aos resultados obtidos na investigação experimental referente à QP3 (Subseção 5.2.3).

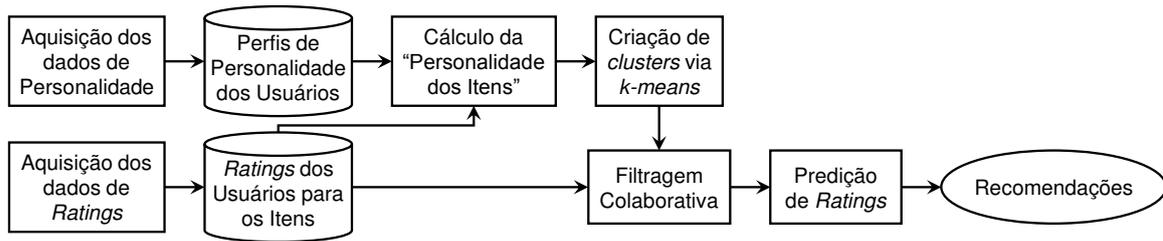
Convém também expor que, em relação à QP4, foram elaboradas outras estratégias, especificamente para o domínio educacional, incluindo-se a teoria dos Estilos de Aprendizagem, considerada em outras pesquisas na Educação. Tais estratégias foram publicadas em conferências, com resultados indiciando a obtenção de recomendações mais acuradas em comparação a outras abordagens (AGUIAR; ARAÚJO; COSTA, 2018a) (AGUIAR; ARAÚJO; COSTA, 2018b).

Nas subseções seguintes, estão os detalhes das estratégias propostas para a geração de recomendações sem vinculação a domínios específicos.

#### 4.4.1 Estratégia 1

Considerando-se a abordagem elaborada referente à QP2 (Seção 4.2), além de considerar que, em relação à FC baseada em vizinhança, calcular a vizinhança baseada no item é uma estratégia relevante em estudos sobre SR, foi elaborada a Estratégia 1. Tal estratégia, ilustrada na Figura 4.3, foi denominada “itps” para lembrar o foco em **it**ens com **tra**ços de **p**ersonalidade similares.

Figura 4.3: Representação geral da estratégia “itps”.



Fonte: autoria própria.

Embora os itens a serem recomendados não sejam geralmente indivíduos (como na recomendação de pessoas em sites de relacionamento), a ideia criada é calcular um “perfil de personalidade” para cada item, com base nos usuários que o avaliaram anteriormente.

Nesse sentido, a “personalidade do item”  $i$  foi definida como a média ponderada dos valores referentes aos traços dos usuários que avaliaram  $i$ , cujos pesos são as notas que os usuários atribuíram ao avaliar tal item. Por exemplo: considere o item  $i_A$ , que recebeu as notas 4, 2 e 5 dos usuários  $u_1$ ,  $u_2$  e  $u_3$ , respectivamente (ou seja:  $(i_A, 4) \in R_{u_1}$ ,  $(i_A, 2) \in R_{u_2}$  e  $(i_A, 5) \in R_{u_3}$ ); e considere os seguintes perfis referentes ao *Big Five* de tais usuários:

- $T_{u_1} = (0, 12; 0, 35; 0, 34; 0, 75; 0, 89)$
- $T_{u_2} = (0, 85; 0, 91; 0, 56; 0, 58; 0, 12)$
- $T_{u_3} = (0, 25; 0, 38; 0, 45; 0, 63; 0, 85)$

Com isso, finalizando o exemplo, a “personalidade do item”  $i_A$  é estimada como  $(0, 31; 0, 47; 0, 43; 0, 66; 0, 73)$ , sendo:

- $0, 31 \approx (4 \times 0, 12 + 2 \times 0, 85 + 5 \times 0, 25)/(4 + 2 + 5)$
- $0, 47 \approx (4 \times 0, 35 + 2 \times 0, 91 + 5 \times 0, 38)/(4 + 2 + 5)$
- $0, 43 \approx (4 \times 0, 34 + 2 \times 0, 56 + 5 \times 0, 45)/(4 + 2 + 5)$
- $0, 66 \approx (4 \times 0, 75 + 2 \times 0, 58 + 5 \times 0, 63)/(4 + 2 + 5)$
- $0, 73 \approx (4 \times 0, 89 + 2 \times 0, 12 + 5 \times 0, 85)/(4 + 2 + 5)$

A partir disso, similarmente ao apresentado na Seção 4.2, é utilizado um processo de *clustering* para desconsiderar os *ratings* de itens vizinhos não tão similares quanto à “personalidade do item” cogitado para a recomendação.

Assim como na Seção 4.2, foi usado o algoritmo *k-means* para definir dois *clusters* por traço do *Big Five*. Isto é, para executar determinado algoritmo de FC em relação a deter-

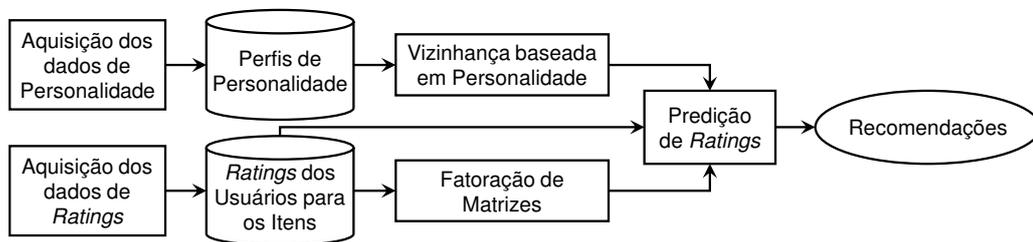
minado usuário-alvo, foram previamente desconsideradas as opiniões (*ratings*) atribuídas a itens menos similares em relação a cada traço; e, por fim, utilizada a média das execuções referentes aos cinco *clusters* dos quais cada item participa.

Após a definição da “personalidade dos itens” e do processo de *clustering*, é empregada uma estratégia de FC (por exemplo, UserKNN) para os *clusters*. Com isso, são calculadas as notas estimadas (*rating prediction*) para os itens candidatos ao usuário-alvo da recomendação (de modo similar à Equação 4.2, na estratégia “tps”), sendo recomendados aqueles itens com maiores notas.

#### 4.4.2 Estratégia 2

Nesta seção, é apresentada a Estratégia 2, ilustrada na Figura 4.4, uma estratégia híbrida (*weighted*) combinando duas abordagens: FC baseada em modelo via fatoração de matriz e FC baseada em vizinhança de usuários definida via personalidade.

Figura 4.4: Representação geral da estratégia *weighted*.



Fonte: autoria própria.

Para definir a vizinhança, são considerados os dados do Perfil de Personalidade de cada usuário (representado pela tupla  $P$  da Equação 4.1). Assim, o Perfil de Personalidade influenciará a recomendação dos itens a partir das preferências dos usuários com personalidade similar à personalidade do usuário-alvo da recomendação.

Para definir os itens mais relevantes, é estimada a nota (*rating*)  $\tilde{r}_{u,i}$  que um usuário-alvo  $u$  daria a um item  $i$  ainda não avaliado/adquirido por  $u$ . Nesta estratégia, a nota estimada é calculada como indicado na Equação 4.3.

$$\tilde{r}_{u,i} = \alpha \times \widetilde{r}_{MF_{u,i}} + (1 - \alpha) \times \widetilde{r}_{PB_{u,i}}. \quad (4.3)$$

Na Equação 4.3,  $\alpha$  é um parâmetro numérico, com valor real entre 0 (zero) e 1 (um), que controla o percentual de contribuição da predição via fatoração de matriz  $\widetilde{r}_{MF_{u,i}}$  na predição final. O valor  $\widetilde{r}_{PB_{u,i}}$  é calculado com base na Equação 4.4.

$$\widetilde{r}_{PB_{u,i}} = \frac{w_1 \times \widetilde{r}_{PB\_Tu,i} + w_2 \times \widetilde{r}_{PB\_Vu,i} + w_3 \times \widetilde{r}_{PB\_Nu,i}}{w_1 + w_2 + w_3}. \quad (4.4)$$

Na Equação 4.4, o valor  $\widetilde{r}_{PB\_Tu,i}$  refere-se à predição usando a abordagem nomeada PB por Hu e Pu (2011), que considera os Traços do modelo *Big Five*, representados pelo vetor  $T$  da tupla  $P$ . Os cálculos para  $\widetilde{r}_{PB\_Vu,i}$  e  $\widetilde{r}_{PB\_Nu,i}$  são análogos a  $\widetilde{r}_{PB\_Tu,i}$ , mas, em vez de usar o modelo *Big Five*, usam, respectivamente, os modelos *Values* — similar ao proposto por Srivastava, Bala e Kumar (2017) — e *Needs*, representados pelos vetores  $V$  e  $N$  da tupla  $P$ . Os valores  $w_1$ ,  $w_2$  e  $w_3$  são parâmetros numéricos representativos do peso de contribuição da predição oriunda de cada um desses modelos de personalidade.

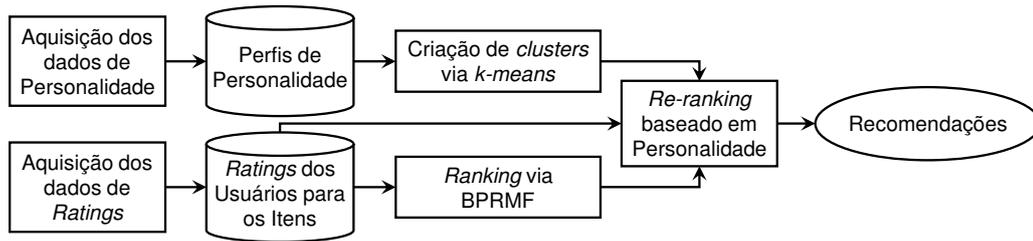
Dado que o modelo *Big Five* é o mais comumente empregado, pode-se definir  $w_1 = 1$  e  $w_2 = w_3 = 0$ , e usar a estratégia sem considerar os modelos *Values* e *Needs*.

A partir do cálculo de  $\widetilde{r}_{u,i}$  para os itens ainda não acessados/avaliados por determinado usuário, é realizada a ordenação final dos itens, de maneira decrescente (os itens com as maiores notas estimadas são recomendados ao usuário-alvo).

Destaca-se que algumas ações podem ser realizadas uma vez, após o cadastro dos usuários, e reutilizadas depois, a exemplo do cálculo da vizinhança baseada em personalidade (os vizinhos só mudariam considerando-se entradas de usuários).

### 4.4.3 Estratégia 3

Nesta subseção, assim como na Subseção 4.4.2, é apresentada uma estratégia utilizando conceitos referentes à personalidade, na construção do perfil dos usuários, e uso de FC baseada em vizinhança e também baseada em fatoração de matriz. Todavia, elaborou-se uma estratégia híbrida com foco em *re-ranking*. Na Figura 4.5, é ilustrada a Estratégia 3.

Figura 4.5: Representação geral da estratégia *re-ranking*.

Fonte: autoria própria.

Inicialmente, é considerado o algoritmo BPRMF (*Bayesian Personalized Ranking Matrix Factorization*) (RENDLE et al., 2009), sendo pré-selecionados, para o usuário-alvo, os primeiros  $n$  itens ranqueados definidos por este algoritmo. Em seguida, é calculado para cada item um índice de re-ranqueamento baseado nas opiniões dos vizinhos mais similares quanto à personalidade, em relação a cada traço do modelo *Big Five*. Baseado nesse cálculo (detalhado a seguir), os itens a serem recomendados são reordenados e apresentados ao usuário.

Para cada item  $i$  do *ranking* (lista  $L_u$ ) com os itens sugeridos para o usuário  $u$  (resultante do algoritmo BPRMF), é calculado um coeficiente de recomendação  $C_{i,u}$ , com valor real entre 0 (zero) e 1 (um), considerando o tamanho da lista ( $|L_u|$ ) e a posição que o item ocupa na lista do usuário ( $pos(i, L_u)$ ). O cálculo é realizado conforme apresentado na Equação 4.5, considerando-se  $0 < pos(i, L_u) \leq |L_u| \leq n$ .

$$C_{i,u} = 1 / \left( 1 + \left( \frac{|L_u|}{|L_u| - pos(i, L_u) + 1} - \frac{1}{pos(i, L_u)} \right) \right). \quad (4.5)$$

Como exemplo, considere que, para determinado usuário  $u_1$ , o algoritmo BPRMF resultou  $L_{u_1} = i_A, i_B, i_C, i_D, i_E, i_F, i_G, i_H$  (isto é:  $|L_{u_1}| = 8$ ,  $pos(i_A, L_{u_1}) = 1$ ,  $pos(i_B, L_{u_1}) = 2$ ,  $pos(i_C, L_{u_1}) = 3$ , e assim sucessivamente). Logo:

- $C_{i_A, u_1} = 1, 0$
- $C_{i_B, u_1} = 0, 609$
- $C_{i_C, u_1} = 0, 5$
- $C_{i_D, u_1} = 0, 426$
- $C_{i_E, u_1} = 0, 357$
- $C_{i_F, u_1} = 0, 286$
- $C_{i_G, u_1} = 0, 206$
- $C_{i_H, u_1} = 0, 113$

Se, para um usuário  $u_2$ , a lista fosse  $L_{u_2} = i_A, i_B, i_C, i_D, i_E$ , ter-se-ia:

- $C_{i_A, u_2} = 1, 0$
- $C_{i_B, u_2} = 0, 571$
- $C_{i_C, u_2} = 0, 429$
- $C_{i_D, u_2} = 0, 308$
- $C_{i_E, u_2} = 0, 172$

Ou seja, o coeficiente de recomendação para o item  $i_D$  é maior sendo 4º lugar dentre 8 indicações (caso do  $u_1$ ) do que sendo 4º lugar dentre 5 indicações (caso do  $u_2$ ).

Com base nisso, é calculado o índice de re-ranqueamento  $Rank_{i,u}$ , como indicado na Equação 4.6.

$$Rank_{i,u} = \alpha \times Rank_{BPRMF_{i,u}} + (1 - \alpha) \times Rank_{TPSi,u}. \quad (4.6)$$

Na Equação 4.6,  $\alpha$  é um parâmetro numérico, com valor real entre 0 (zero) e 1 (um), que controla o percentual de contribuição do *ranking* original ( $Rank_{BPRMF_{i,u}}$ ) no *re-ranking*. O valor de  $Rank_{BPRMF_{i,u}}$  consiste no valor de  $C_{i,u}$ . O valor de  $Rank_{TPSi,u}$  é calculado com base na Equação 4.7.

$$Rank_{TPSi,u} = \frac{\sum_{x=1}^{|T|} Rank_{TP}(t_x)}{|T|}. \quad (4.7)$$

Na Equação 4.7, é mostrado que, para cada traço  $t_x$  do modelo *Big Five* ( $|T| = 5$ ), calcula-se  $Rank_{TP}(t_x)$ , que consiste na média dos coeficientes dos vizinhos para o item  $i$  — para cada vizinho  $v$  (usuário do mesmo *cluster* referente ao traço  $t_x$ ) do usuário  $u$ , é calculado  $C_{i,v}$  (caso  $i \in L_v$ ).

Por fim, a lista final recomendada para o usuário-alvo  $u$  consiste no *re-ranking* contemplando os  $n$  itens ordenados decrescentemente pelos valores referentes aos índices de re-ranqueamento ( $Rank_{i,u}$ ).

## 4.5 Considerações Finais do Capítulo

Nas seções deste capítulo, foram apresentadas as abordagens elaboradas relacionadas aos objetivos desta tese. Em geral, a elaboração dessas abordagens visa a contribuir com novas estratégias para a área de FC baseada em personalidade.

Ressalta-se que, para aplicar as estratégias propostas, não necessariamente se deve usar o *Watson Personality Insights* como instrumento de mensuração das características de personalidade dos usuários. Entretanto, o(s) instrumento(s) utilizado(s) — APR ou questionários — devem ser baseados nos modelos *Big Five*, *Needs* e *Values*.

Ressalta-se ainda que, para a Estratégia 3 (Subseção 4.4.3), é possível substituir o algoritmo BPRMF por outro algoritmo de recomendação, além de ser possível adaptar a estratégia para empregar características de personalidade diferentes do modelo *Big Five*.

No próximo capítulo, são abordados os procedimentos metodológicos e os resultados obtidos referentes à investigação experimental realizada concernente às abordagens elaboradas.

# Capítulo 5

## Procedimentos Metodológicos e Resultados

Com base nas explicações dos conceitos e estudos relacionados, apresentados nos capítulos anteriores, neste capítulo são apresentados os aspectos metodológicos adotados, além dos resultados obtidos, relacionados aos objetivos específicos estabelecidos na Seção 1.2.

Em SR, o processo de validação é frequentemente realizado empregando técnicas de validação cruzada, como *K-fold Cross-validation* e *Random Subsampling* (BOBADILLA et al., 2013). O método *Random Subsampling* (validação cruzada com subamostragem aleatória) consiste na separação aleatória dos dados em dois subconjuntos mutuamente exclusivos (um para treinar o modelo e outro para testá-lo), sendo tal processo repetido por  $k$  vezes e sendo comum atribuir o percentual de dois terços (HAN; KAMBER; PEI, 2012).

Nesta pesquisa, para as análises experimentais, foi empregado o método *Random Subsampling*, sendo divididos, aleatoriamente, por 50 vezes, os dados de cada conjunto de dados em dois subconjuntos: um para treinamento (2/3 dos dados) e outro para teste (1/3 dos dados). Os conjuntos de dados utilizados estão descritos na Subseção 5.1.

Nos experimentos realizados, do tipo comparativo, a variável independente foi simplificada nomeada ‘Algoritmo’ (representação de uma estratégia de recomendação que tem como entrada os dados do conjunto de treinamento) e seus níveis consistiram nos algoritmos elaborados para as estratégias e variações propostas, além de algoritmos comumente utilizados em SR (a exemplo do UserKNN e do SVD++).

Considerando-se o foco da pesquisa definido no Capítulo 1, as variáveis de resposta

dos experimentos (variáveis dependentes) consistiram em valores numéricos, decimais, de natureza quantitativa: a métrica *F-Measure* (com limiar 3,5) e a métrica RMSE, comentadas na Seção 2.2, relacionadas à relevância (acurácia) da recomendação.

Para análise dos valores das variáveis dependentes, foram utilizados estes testes estatísticos: ANOVA e T (paramétricos) e Kruskal-Wallis e Mann–Whitney U (não paramétricos) (BOSLAUGH, 2012). De modo genérico, as seguintes hipóteses foram consideradas:

- Hipótese Nula ( $H_0$ ): O valor da variável dependente em análise é similar para os algoritmos em estudo.
- Hipótese Alternativa ( $H_1$ ): O valor da variável dependente em análise difere para os algoritmos em estudo.

Uma vez que  $H_0$  não seja confirmada, considera-se haver diferença entre os algoritmos e, com base nos *boxplots* e/ou intervalos de confiança para a média — fundamentados pelo Teorema Central do Limite (BOSLAUGH, 2012) —, analisa-se a variável dependente para responder à questão referente ao experimento. Para os intervalos de confiança gerados e para os testes estatísticos realizados, foi considerado o nível de significância de 0,05 (isto é:  $\alpha = 5\%$ ). A execução dos experimentos realizados envolveu as seguintes fases:

- Separação aleatória dos dados em conjuntos de treinamento e teste;
- Execução dos algoritmos; e
- Análise estatística dos resultados.

Para a execução de alguns algoritmos (UserKNN, ItemKNN, SMF, BMF, SVD++ e BPRMF), foi utilizada a biblioteca MyMediaLite (GANTNER et al., 2011). Para a execução de outros algoritmos — além da organização e separação aleatória dos dados — foram utilizadas as linguagens de programação Java, Python e R. A análise estatística dos resultados foi realizada por meio do ambiente R (R Core Team, 2021) (RStudio Team, 2021), para sumarizar os resultados obtidos, executar testes e elaborar *boxplots* e intervalos de confiança.

## 5.1 Conjuntos de Dados

Dentre os estudos referentes à recomendação baseada em personalidade, como apresentado no Capítulo 3, muitos destacam a influência da personalidade nos domínios de filme e música, inclusive associando determinados traços a determinados gêneros/categorias desses domínios (*personality matching*). Dado que, neste estudo, desejou-se explorar a importância da personalidade na recomendação via FC (bastante útil em domínios cujos itens possuem poucas informações como gêneros/categorias), foi inicialmente planejado explorar domínios menos explorados tradicionalmente (não enfocando em conjuntos de dados referentes a filmes e músicas).

Nesse sentido, para as avaliações experimentais, foi inicialmente definido o uso do conjunto de dados TripAdvisor e, além desse, foi utilizado um conjunto de dados da Amazon.

O conjunto de dados TripAdvisor, utilizado e disponibilizado por Alexandra Roshchina (ROSHCHINA; CARDIFF; ROSSO, 2015), contém dados referentes à personalidade dos usuários. Para inferir tais dados, os autores usaram a abordagem de APR proposta por Mairesse et al. (2007), um dos estudos pioneiros para inferência dos traços de personalidade a partir de textos, e uma das referências da IBM no projeto do *Watson Personality Insights*. Um dos principais motivos para escolher este conjunto de dados foi devido ao estudo de Roshchina, Cardiff e Rosso (2015) ser um dos poucos mais relacionados quanto à questão de pesquisa QP3.

Além de usar dados do TripAdvisor (domínio turístico), foi encontrado e utilizado um conjunto de dados da Amazon (*Fine Foods*), sendo motivada esta escolha pela importância de SR no comércio eletrônico, além da possibilidade de analisar o domínio alimentício, pouco analisado se comparado a livros, filmes e músicas<sup>1</sup>.

O conjunto de dados TripAdvisor<sup>2</sup> contém 32.580 *reviews* oriundos de 1.098 usuários para 26.282 itens (hotéis, restaurantes e atrações) do TripAdvisor, uma plataforma popular

---

<sup>1</sup>Notadamente, há mais estudos em alguns domínios, percebendo-se uma relação mais direta da influência da personalidade em uma tomada de decisão — a exemplo do domínio turístico, em que é facilmente possível associar alguns destinos a “comportamentos exploratórios” de um indivíduo, que se relacionam ao traço Abertura, como visto no Quadro 2.1. Por outro lado, no domínio alimentício, é mais difícil ilustrar uma associação simples, como esse exemplo anterior, mas há estudos relacionando traços de personalidade a preferências/hábitos alimentares (SPINELLI et al., 2018) (UFER; LIN; ORTEGA, 2019) (MACHADO-OLIVEIRA et al., 2020) (ESPOSITO; CERESA; BUOLI, 2021).

<sup>2</sup>Disponível em: [https://twin-persona.org/datasets/2016/dataset\\_description.txt](https://twin-persona.org/datasets/2016/dataset_description.txt). Acesso em: 30 de out. de 2021.

de viagens. O conjunto de dados *Fine Foods*<sup>3</sup> contém 568.454 *reviews* oriundos de 256.059 usuários para 74.258 itens (produtos alimentícios — “*gourmet foods*”) da Amazon, uma empresa popular de comércio eletrônico.

Com a possibilidade atual de empregar ferramentas de APR, e considerando-se o objetivo geral da pesquisa, a obtenção do Perfil de Personalidade dos usuários foi realizada com o *Watson Personality Insights*. Com tal ferramenta, foi analisado o texto dos *reviews* dos usuários (linguagem natural em inglês), sendo configurado o parâmetro de consulta *raw\_scores = true*, para computar pontuações brutas — que podem ser interpretadas como equivalentes às pontuações obtidas pelo autor do texto analisado se ele fizesse um teste de personalidade (IBM, 2021). No Apêndice C, é apresentado um exemplo.

Visto que se deve fornecer ao *Watson Personality Insights* pelo menos 600 palavras, por usuário, para o serviço produzir resultados aceitáveis<sup>4</sup>, foram desconsiderados os usuários que não atendiam a essa restrição. Além disso, foi também planejado restringir os itens, desconsiderando aqueles avaliados por menos de 5 usuários. Contemplando isso, as variações dos conjuntos de dados foram nomeadas *TripAdvisor* e *AmazonFoods*.

Com tais restrições, especialmente no conjunto *TripAdvisor*, a quantidade de *ratings* não ficou tão expressiva (com média de apenas aproximadamente dois *ratings* por usuário). Então, decidiu-se também usar variações sem a restrição dos itens serem avaliados por pelo menos 5 usuários (variações nomeadas *TripAdvisorCompleto* e *AmazonFoodsCompleto*). Nesse caso, embora a média de *ratings* por usuário tenha aumentado, foram contemplados bastantes itens pouco avaliados (no conjunto de dados *TripAdvisor*, a média ficou aproximadamente um *rating* por item)<sup>5</sup>. Na Tabela 5.1, são sintetizadas as informações sobre as variações dos conjuntos de dados utilizados.

<sup>3</sup>Disponível em: <https://snap.stanford.edu/data/web-FineFoods.html>. Acesso em: 30 de out. de 2021.

<sup>4</sup>Apesar de recomendar o uso com pelo menos 1.200 palavras, a IBM afirma que fornecer pelo menos 600 palavras produz resultados aceitáveis. Com menos de 600 palavras, o serviço analisa a entrada, mas gera um aviso. Ademais, a IBM informa que menos de 100 palavras gera um erro, e que a precisão máxima do serviço é alcançada com 3.000 palavras. (IBM, 2021)

<sup>5</sup>Com base nessas informações, foi percebido que o uso desses dados oriundos do *TripAdvisor* não seria tão válido quanto o uso dos dados oriundos da Amazon para analisar, por exemplo, as estratégias referentes à QP4. Todavia, preferiu-se continuar a investigação experimental também com tais variações (conjuntos *TripAdvisor* e *TripAdvisorCompleto*), especialmente considerando-se o objetivo específico referente à QP1.

Tabela 5.1: Características referentes aos conjuntos de dados.

Conjunto de dados	Quantidade de itens	Quantidade de usuários	Quantidade de ratings	Média de ratings por item	Média de ratings por usuário
TripAdvisor	192	641	1.250	6,51	1,95
TripAdvisorCompleto	26.844	1.058	31.645	1,18	29,91
AmazonFoods	4.306	10.889	106.371	24,70	9,77
AmazonFoodsCompleto	29.962	11.765	145.660	4,86	12,38

Fonte: autoria própria.

## 5.2 Análise Experimental

Nesta seção, são retomados os objetivos da pesquisa, explicitando outros aspectos metodológicos — específicos para a análise experimental referente a cada questão de pesquisa —, além de serem apresentados os resultados obtidos.

### 5.2.1 Experimento referente à QP1

Nesta subseção, é descrita a análise experimental objetivando rever pesquisas iniciais de referência — que propuseram estratégias de FC baseada em personalidade, identificando-a via questionários — para verificar se, ao usar APR (baseado em texto) e conjuntos de dados com maior número de usuários e itens, os resultados de tais estratégias continuam sendo melhores em relação à estratégia tradicional baseada apenas em *ratings*.

Como expresso na Seção 4.1, foram consideradas variações fundamentadas nas pesquisas de Tkálčič et al. (2009) e Hu e Pu (2011), sendo implementados os algoritmos, para o estudo comparativo, baseando-se em seus artigos. Em síntese, para este experimento, foram definidas estas variáveis independentes (e seus correspondentes níveis):

- Variação de FC: Hu&Pu, TkálčičEtAl-public, TkálčičEtAl-neighbours;
- Tipo de Vizinhança: RB, PB, RPBL; e
- Modelo de Personalidade: B (*Big Five*, sem incluir facetas), Bf (*Big Five*, incluindo facetas), V (*Values*), N (*Needs*), V-N-Bf (dados de todos os modelos conjuntamente).

Objetivou-se, inicialmente, avaliar se as abordagens considerando personalidade (uso de PB e RPBL) se comportam melhor que RB (independentemente da variação: Hu&Pu, TkálčičEtAl-public ou TkálčičEtAl-neighbours) sem uso de questionários — considerando os dados de personalidade inferidos via *Watson Personality Insights*. Adicionalmente,

objetivou-se avaliar o impacto dos modelos de personalidade utilizados, analisando se há diferenças ao usar modelos diferentes do popular *Big Five*, e avaliar se há melhorias ao considerar mais dados — a exemplo de usar dados das facetas do *Big Five* (visto que a maioria dos estudos considera apenas os dados dos seus cinco grandes fatores) e de outros modelos (usando *Big Five* junto a *Needs e Values*).

Os dados foram analisados por conjunto de dados e por agrupamentos referentes à variação de FC (Hu&Pu; TkalčičEtAl-public; TkalčičEtAl-neighbours). Os valores resultantes dos testes estatísticos encontram-se no Apêndice D (Tabela D.1).

Em resumo, não houve diferença estatística significativa entre as estratégias baseando-se nos dados oriundos do TripAdvisor — isso também foi comentado em um artigo publicado (AGUIAR; ARAÚJO; COSTA, 2020b). Nas Figuras 5.1 e 5.2, são ilustrados os *box-plots* referentes aos dados oriundos da Amazon, indicando que as estratégias baseadas em personalidade (PB e RPBL) foram melhores que a abordagem tradicional (RB), destacada em vermelho.

Considerando o objetivo referente à QP1 (Seção 1.2), é possível afirmar que, ao rever os primeiros estudos relevantes que propuseram estratégias de FC baseadas em traços de personalidade, tais estratégias tiveram desempenho no mínimo similar à estratégia tradicional (baseada apenas em *ratings*) ao avaliar com mais usuários e itens<sup>6</sup>, e personalidade identificada via *Watson Personality Insights* em vez de questionários.

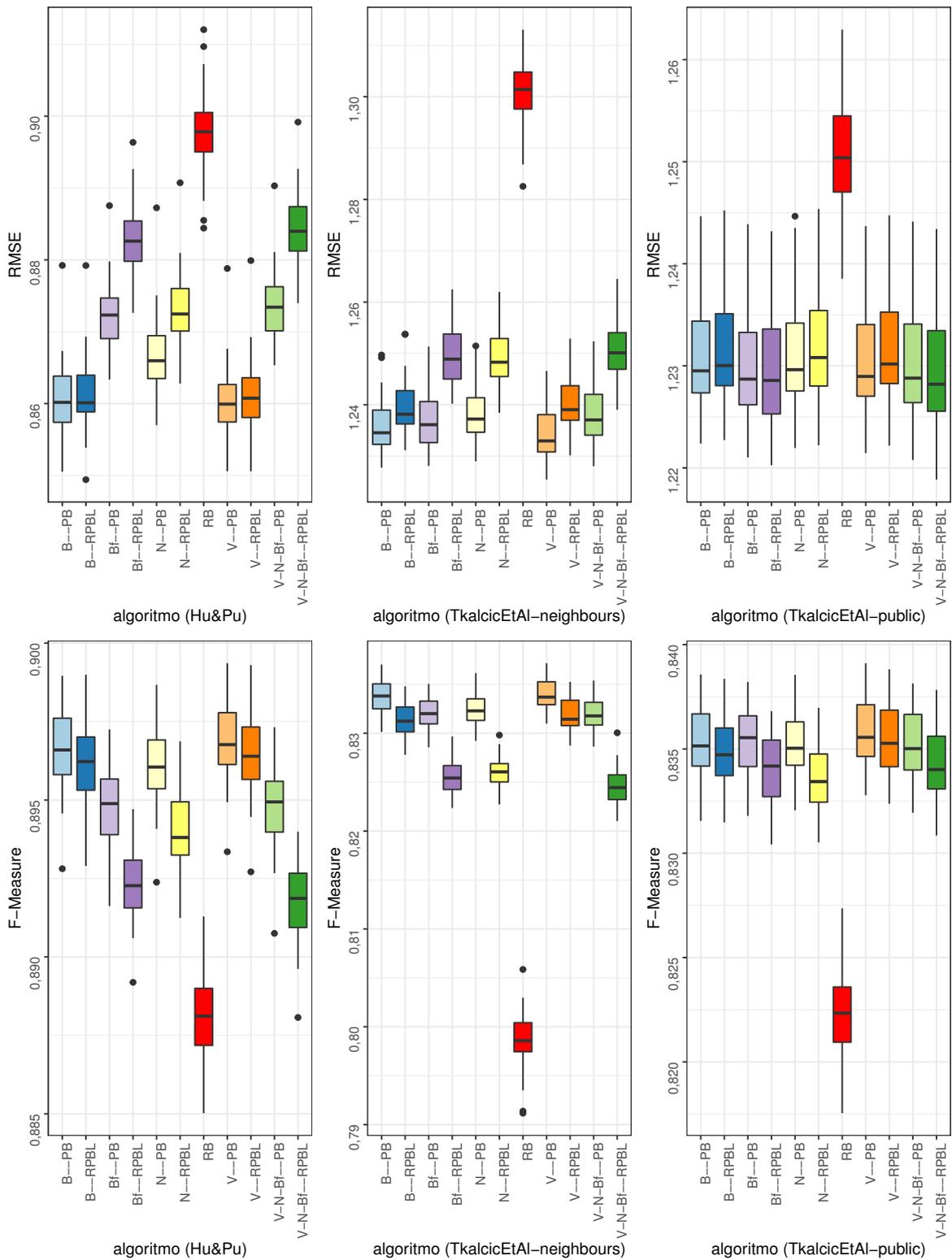
Tal resultado impacta na motivação de seguir nos estudos de estratégias baseadas em personalidade para, na prática, ser viável as empresas de comércio eletrônico, por exemplo, adotarem tais estratégias em seus SR, dado que indicá-las o uso de questionários não é motivador.

Em todos os cenários analisados com os dados oriundos da Amazon, o pior desempenho (maiores valores de RMSE e menores valores de *F-Measure*) foi proporcionado por RB, o que corrobora outros estudos, a exemplo de Hu e Pu (2011) e Srivastava, Bala e Kumar (2017). Por outro lado, as abordagens com determinado modelo de personalidade não resultaram sempre em melhor ou pior resultado quando comparadas às abordagens com os outros modelos.

---

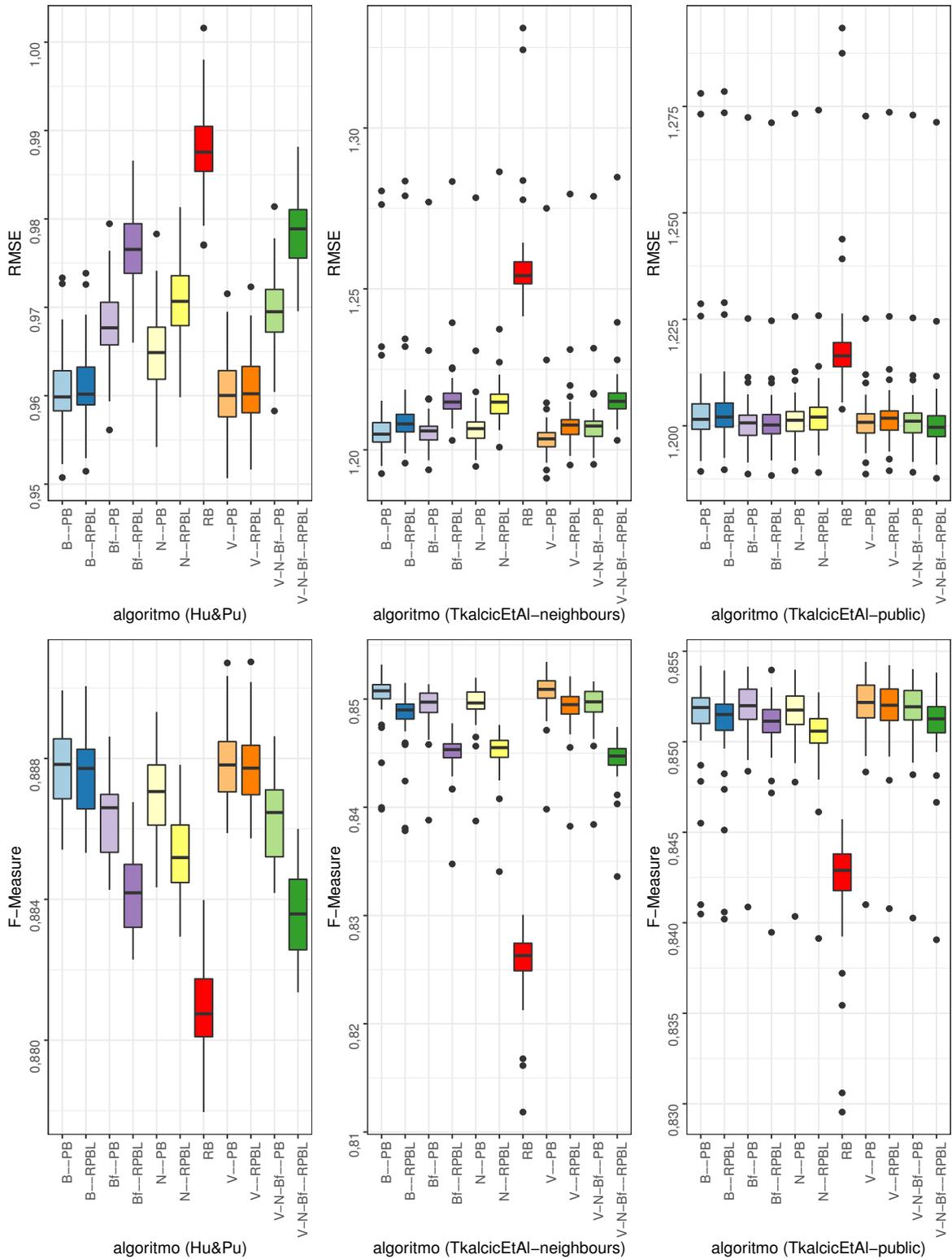
<sup>6</sup>Tkalčič et al. (2009) consideraram 52 usuários e 70 itens (imagens). Hu e Pu (2011) consideraram 230 usuários e 640 itens (músicas).

Figura 5.1: *Boxplots* referentes à QP1 com o conjunto de dados AmazonFoods.



Fonte: autoria própria.

Figura 5.2: *Boxplots* referentes à QP1 com o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Ademais, as abordagens aplicando RPBL não proporcionaram resultados melhores (em ambas as métricas) que as abordagens aplicando PB. Além do fato de RPBL ser mais custosa computacionalmente que PB, como indicado por Hu e Pu (2011), os resultados não indicaram ser tão útil, para a melhoria da acurácia, mesclar as duas formas de se definir a vizinhança (PB e RB), que é o foco de RPBL.

Ao analisar o impacto dos modelos nas abordagens, em relação à variação Hu&Pu (para ambas as métricas e tanto em PB quanto em RPBL), percebeu-se um desempenho melhor com B (modelo *Big Five*) e V (modelo *Values*), seguido de N (modelo *Needs*). Isso implica dizer que, com a variação Hu&Pu, as abordagens usando o modelo *Values* ou o modelo *Needs* proporcionaram melhores resultados que as abordagens incluindo as facetas do modelo *Big Five* (Bf), também não sendo útil compor o perfil de personalidade dos usuários com os três modelos de forma conjunta (V-N-Bf).

Em relação às variações TkalčičEtAl-neighbours e TkalčičEtAl-public, as abordagens com B e V continuaram tendo desempenhos semelhantes. Embora não seja possível, em todos os cenários, ressaltar veementemente a utilização de um modelo de personalidade em relação aos demais, ao analisar os valores obtidos com as variações Hu&Pu, TkalčičEtAl-neighbours e TkalčičEtAl-public, os melhores resultados (menores valores de RMSE e maiores valores de *F-Measure*) foram referentes à variação Hu&Pu. Assim, indicia-se a utilidade do modelo *Values* na recomendação baseada em personalidade, além do popular *Big Five*, mas sem incrementá-lo com os dados das suas facetas ou com os dados dos outros modelos analisados.

No âmbito de FC atualmente, é comum usar abordagens baseadas em modelo, como o uso de fatoração de matriz. Considerando-se que as abordagens analisadas são de FC baseada em memória, o resultado desta análise referente à QP1 motivou a QP2.

### **5.2.2 Experimento referente à QP2**

Nesta subseção, é descrita a investigação experimental objetivando analisar se a acurácia das predições de um algoritmo de FC é afetada (visando, especialmente, à melhoria) ao desconsiderar a “opinião” (*ratings*) de usuários não tão similares quanto à personalidade. Os comentários referentes à ocorrência de diferença estatística baseiam-se em execuções de testes T e testes Mann-Whitney U (Apêndice D — Tabelas D.2, D.3 e D.4).

Foi utilizada a estratégia descrita na Seção 4.1, variando-a com alguns algoritmos de FC. Considerando-se as duas categorias de FC (comentadas na Subseção 2.2.1), foram definidos os algoritmos BMF, SMF e SVD++ (da categoria *model-based*) e os algoritmos ItemKNN e UserKNN (da categoria *neighborhood-based*), disponíveis na biblioteca MyMediaLite. Para identificar simplificada-mente esses algoritmos, nas figuras do Apêndice E e no Quadro 5.1, foi utilizado o termo “base” (e a cor vermelha). Para identificar a variação dos algoritmos com a estratégia de agrupamento prévio de usuários com características de “traços de personalidade similares”, foi utilizado o termo “tps” (e a cor azul). Os dados foram analisados por conjunto de dados e por agrupamentos referentes ao modelo.

Os resultados obtidos neste experimento referente à QP2 estão detalhados no Apêndice E. No Quadro 5.1, tais resultados estão resumidos, indicando a variação (“base” ou “tps”) — para cada algoritmo de FC em análise — que estatisticamente proporcionou melhor resultado, em cada cenário analisado (considerando-se os diferentes conjuntos de dados, modelos de personalidade e métricas). O termo “base|tps” (com realce na cor cinza) indica que, em tal cenário, não houve diferença estatística de desempenho em relação às variações analisadas.

Quadro 5.1: Síntese dos resultados referentes à QP2.

Conjunto de dados	Algoritmo	Modelo <i>Big Five</i>		Modelo <i>Values</i>		Modelo <i>Needs</i>	
		RMSE	<i>F-Measure</i>	RMSE	<i>F-Measure</i>	RMSE	<i>F-Measure</i>
TripAdvisor	UserKNN	tps	base tps	tps	base tps	tps	base tps
	ItemKNN	tps	base tps	tps	base tps	tps	base tps
	SMF	base tps	base tps	base tps	base tps	base tps	base tps
	BMF	base tps	base tps	base tps	base tps	base tps	base tps
	SVD++	base tps	base tps	base tps	base tps	base tps	base tps
TripAdvisorCompleto	UserKNN	tps	base tps	tps	base tps	tps	base tps
	ItemKNN	tps	base tps	tps	base tps	tps	base tps
	SMF	tps	base tps	tps	base tps	tps	base tps
	BMF	base tps	base tps	base tps	base tps	base tps	base tps
	SVD++	base tps	base tps	base tps	base tps	base tps	base tps
AmazonFoods	UserKNN	tps	tps	tps	tps	tps	tps
	ItemKNN	tps	tps	tps	tps	tps	tps
	SMF	base	base	base	base	base	base
	BMF	base	base	base	base	base	base
	SVD++	tps	tps	base tps	tps	base	tps
AmazonFoodsCompleto	UserKNN	tps	tps	tps	tps	tps	tps
	ItemKNN	tps	base tps*	tps	tps	tps	tps
	SMF	base	base	base	base	base	base
	BMF	base	base	base	base	base	base
	SVD++	tps	base tps	base tps	base tps	base	base tps*

\* Diferença estatística não confirmada em ambos os testes analisados: o resultado do teste Mann-Whitney U não indicou diferença estatística (p-valor > 0,05), diferentemente do resultado do teste T (p-valor < 0,05); todavia, preferiu-se não rejeitar a hipótese nula, devido ao p-valor do teste T próximo a 0,05.

Fonte: autoria própria.

Respondendo à questão de pesquisa QP2, não é possível afirmar categoricamente que a acurácia das predições de um algoritmo de FC é afetada se o algoritmo for executado focando na opinião de usuários similares quanto à personalidade. Pelos resultados, em muitos casos não houve diferença estatística. Nos casos em que houve tal diferença, nem sempre a estratégia envolvendo personalidade proporcionou melhor desempenho.

Assim, não é possível afirmar que a acurácia das predições de um algoritmo de FC sempre se mantém (nem sempre tende a melhorar) ao desconsiderar os “dados de opinião” (*ratings*) de usuários menos similares quanto à personalidade (identificada automaticamente via *Watson Personality Insights*) de um usuário-alvo de recomendação.

Em geral, é possível perceber uma tendência de melhoria nos algoritmos de FC baseados em vizinhança (UserKNN e ItemKNN), especialmente em relação ao UserKNN. Dado que o UserKNN é focado na vizinhança dos usuários, então o pré-agrupamento de usuários baseados em personalidade se mostra útil ao empregar tal algoritmo de FC.

Para os algoritmos baseados em modelo, apesar de alguns casos tendendo à melhoria de desempenho (acurácia) aplicando personalidade, além dos casos em que o desempenho se manteve, por vezes o desempenho foi pior. Para esta categoria de algoritmos, não foi tão surpreendente ter casos de perda de desempenho com a redução de dados, pois foi “audacioso” esperar melhorias ao diminuir dados para a definição do modelo (outros aspectos importantes para definir o modelo de predição provavelmente não puderam ser captados ao dividir os dados). É conveniente destacar que, quando inicialmente analisado no contexto educacional (AGUIAR; ARAÚJO; COSTA, 2020c), não houve melhoria nos algoritmos baseados em modelo, mas também não houve piora significativa, provavelmente devido ao conjunto de dados não conter um grande volume como o da Amazon, em que os modelos internos puderam ser mais bem definidos.

Este resultado referente à QP2 alinha-se à motivação desta tese, quando comentado da premissa “pessoas similares quanto à personalidade tendem a preferir itens similares”. Reforça-se, então, a hipótese de que, especialmente nos algoritmos de FC baseados em vizinhança, se o “treinamento” para se definir um *rating* for realizado desconsiderando a “opinião” de usuários não tão similares quanto à personalidade, isso não tende a diminuir significativamente a acurácia da predição (possibilitando melhorias). Esse resultado se mostra instigante, pois, dado que o uso da personalidade vem sendo estudado para a melhoria da

qualidade das recomendações em aspectos além da acurácia (como a diversidade), o método genérico empregado em um algoritmo de FC possibilita utilizar características de personalidade em uma estratégia de recomendação, proporcionando possíveis outras vantagens (como melhoria relativa à diversidade) sem perder significativamente na acurácia e até possibilitando melhorar tal aspecto.

### 5.2.3 Experimento referente à QP3

Nesta subseção, é descrita a investigação experimental objetivando analisar a influência dos cinco traços do modelo de personalidade *Big Five* (identificados via APR baseado em texto). Inicialmente, conjecturou-se que, ao desconsiderar algum dos traços na construção do perfil de um usuário, o resultado do algoritmo de FC baseada em personalidade seria diferente. Assim, objetivou-se destacar se algum(ns) deles influencia(m) mais fortemente em melhores resultados, avaliando a possibilidade de elaborar uma estratégia com pesos diferenciados.

Os valores resultantes dos testes estatísticos (ANOVA e Kruskal-Wallis) referentes a esta questão de pesquisa encontram-se no Apêndice D (Tabelas D.5 e D.6). No Apêndice F, são apresentados os intervalos de confiança das variações de uso dos traços do *Big Five*, considerando alguns algoritmos base de FC. Os dados foram analisados por conjunto de dados e por algoritmo base.

Para o conjunto TripAdvisor, não foi possível considerar que houve diferença estatística entre variações, tanto referente ao RMSE quanto referente à *F-Measure*. Para o conjunto TripAdvisorCompleto, em relação a BMF e SVD++, também não houve diferença entre variações, para ambas as métricas em estudo. Para UserKNN (Figura F.1), ItemKNN (Figura F.2) e SMF (Figura F.3), houve diferença referente ao RMSE, mas não referente à *F-Measure* (e, ainda assim, nenhum com desempenho estatisticamente melhor que o da variação tradicional — as diferenças foram em relação à redução de desempenho).

Para o conjunto AmazonFoods (Figuras F.4, F.5, F.6, F.7 e F.8) e AmazonFoodsCompleto (Figuras F.9, F.10, F.11, F.12 e F.13), foi possível considerar que houve diferença estatística entre as variações, tanto referente ao RMSE quanto referente à *F-Measure* e em relação a todos os algoritmos.

Foram executados testes par a par (T e Mann-Whitney U) para identificar, estatisticamente, possíveis variações com melhor desempenho que a variação tradicional. Só houve

casos disso ao analisar o algoritmo SVD++ (como visualmente demonstrado pelos intervalos de confiança das Figuras F.8 e F.13) e algumas diferenças aconteceram em apenas uma das métricas (RMSE ou *F-Measure*). São listadas a seguir as variações que se destacaram em ambas as métricas.

- AmazonFoods

000-000-060-020-020  
*000-000-080-020-000*  
 000-020-060-020-000  
 000-020-080-000-000  
 000-040-040-020-000  
*000-040-060-000-000*  
 000-050-050-000-000

- AmazonFoodsCompleto

000-000-080-020-000  
 000-020-060-000-020  
 000-020-060-020-000  
 000-020-080-000-000  
 000-040-060-000-000

As que estão em itálico se destacaram nas duas variações do conjunto de dados. Além disso, está sublinhada a única variação que, em ambos os conjuntos de dados, permaneceu com diferença estatística ao considerar  $\alpha = 1\%$ . Nesse sentido, é possível notar, nesse resultado, a importância do terceiro traço do acrônimo OCEAN (E — Extroversão) aliando-o ao quarto traço (A — Amabilidade) e/ou ao segundo traço (C — Conscienciosidade). Entretanto, considerando que isso só foi percebido em relação ao SVD++, não é possível defender a possibilidade de omitir algum traço.

Em síntese, analisando os resultados anteriores, a variação considerando todos os traços, em geral, não tem desempenho significativamente pior a ponto de justificar a escolha de uma variação “desprezando” algum traço.

Nas Figuras F.14 a F.23, estão os intervalos de confiança mostrando variações sem omitir traços, mas variando o peso em níveis de 10%. Ou seja, considerando-se, pela análise anterior, que não foi justificável deixar de considerar algum traço, foi analisado se o aumento do peso de algum tenderia a melhorar a acurácia. Considerando-se os resultados anteriores, tal análise não foi realizada nos dados oriundos do TripAdvisor.

Nessa análise referente a pesos diferentes, não houve casos de variações com diferença estatística e resultado melhor, em ambas as métricas (RMSE e *F-Measure*), que a variação 020-020-020-020-020. Especificamente para a métrica RMSE, houve diferença estatística

em relação ao algoritmo SVD++ (como esperado pela análise anterior), e também algumas ocorrências referentes ao algoritmo UserKNN para AmazonFoodsCompleto), sendo pertinente destacar esta variação: 010-020-040-020-010 (novamente com maior importância do traço E, seguido dos traços C e A).

Entretanto, para os demais, apesar de haver variações com médias melhores, não se justificaria empregar alguma técnica (a exemplo de Algoritmos Genéticos) para internamente definir pesos, em vez de considerar os cinco traços com o mesmo peso. Esse resultado destaca a importância da atuação conjunta e igualmente ponderada desses cinco traços na caracterização da personalidade de uma pessoa.

Considerando o foco da questão de pesquisa QP3, foi usado, adicionalmente, o conjunto de dados Personality2018<sup>7</sup> (NGUYEN et al., 2018), que contém 1.028.751 *reviews* oriundos de 1.834 usuários do MovieLens, com identificação da personalidade via questionário TIPI. Na Tabela 5.2, estão as informações sobre a variação utilizada desse conjunto de dados. Foram desconsiderados alguns usuários, resultando em 1.819 usuários que avaliaram, no mínimo, 15 filmes. Além disso, foram mantidos apenas os filmes avaliados por pelo menos 5 usuários, visto que a quantidade de *ratings* ainda ficou expressiva (96% do conjunto total disponibilizado).

Tabela 5.2: Características referentes ao conjunto de dados Personality2018.

Quantidade de itens	Quantidade de usuários	Quantidade de <i>ratings</i>	Média de <i>ratings</i> por item	Média de <i>ratings</i> por usuário
14.868	1.819	984.499	66,22	541,23

Fonte: autoria própria.

A análise no conjunto de dados Personality2018 foi realizada com intento de verificar se, em caso de não empregar APR, também se destaca a influência dos cinco traços de maneira conjunta e igualmente ponderada. Observou-se exatamente isso, como mostrado nas Figuras F.24 e F.25. Para esta análise complementar, foram executados apenas os algoritmos base SMF e BMF.

Apesar do destaque do *Big Five* na Computação de Personalidade, julgou-se conveniente realizar experimentações referentes aos componentes de outros modelos, dado que outras características referentes à personalidade do indivíduo (como *Values* e *Needs*) também podem

<sup>7</sup>Disponível em: <https://grouplens.org/datasets/personality-2018/>. Acesso em: 30 de out. de 2021.

ser identificadas via APR.

Como comentado na Seção 4.3, foram primeiramente analisadas as 26.334 combinações (combinações de 5 componentes considerando o total dos 22 componentes oriundos dos modelos *Big Five*, *Needs* e *Values*). Posteriormente, foram consideradas as 52 características (adicionando as 30 facetas do *Big Five* aos 22 componentes), mas, considerando-se o vasto número de combinações neste caso (2.598.960), fez-se uso de um GA para avaliar se alguma combinação é significativamente melhor que apenas o modelo *Big Five*.

Considerando-se primeiramente todas as 26.334 combinações, foram realizadas análises para os 5 algoritmos (SMF, BMF, SVD++, UserKNN e ItemKNN). Em síntese, em relação a SMF e BMF, a melhor combinação (menor mediana referente ao RMSE) não foi melhor que o algoritmo baseline; todavia, a melhor combinação encontrada foi estatisticamente melhor que a combinação clássica do *Big Five* (os cinco traços). Em relação a SVD++, UserKNN e ItemKNN, a melhor combinação (menor mediana referente ao RMSE) foi melhor que a combinação clássica do *Big Five* e que o baseline.

Considerando-se o uso de GA, também foram realizadas análises para os cinco algoritmos (SMF, BMF, SVD++, UserKNN e ItemKNN). Em síntese, assim como visto ao analisar as combinações sem incluir as facetas, houve alguma melhoria (menor mediana referente ao RMSE) em relação à combinação clássica do *Big Five*. Em relação a SMF e BMF, não houve melhoria em relação ao baseline sem considerar personalidade. Em relação a SVD++, UserKNN e ItemKNN, o GA encontrou alguma combinação melhor que o baseline.

Adicionalmente, foi empregado o GA sem incluir as 30 características referentes às facetas. Apesar do GA considerando as 52 características atingir geralmente melhores resultados que o GA considerando as 22 características, o impacto na diminuição do RMSE não foi maior que 0,5%. Por fim, sobre a investigação de inclusão das facetas do *Big Five* e dos outros componentes (*Values* e *Needs*), realizada no conjunto de dados AmazonFoods, julgou-se que os resultados não justificariam empregar um GA para considerar adaptações em relação às características.

Respondendo à questão de pesquisa QP3, é possível dizer, em síntese, que os traços de personalidade influenciam de maneira diferente no resultado da recomendação, mas não há indícios de melhorias expressivas ao priorizar determinados traços.

### 5.2.4 Experimento referente à QP4

Nesta subsecção, é descrita a avaliação de estratégias para a geração de recomendações personalizadas baseadas nos traços de personalidade dos usuários e suas preferências anteriores.

O objetivo desta investigação consistiu em analisar algoritmos comumente utilizados em SR (citados no Capítulo 2) e os algoritmos implementados a partir das estratégias propostas (Capítulo 4), com a intenção de compará-los a respeito da relevância (acurácia) de suas recomendações referentes a itens, do ponto de vista de quem recebe tais recomendações, no contexto dos conjuntos de dados definidos na Seção 5.1.

#### Análise da Estratégia 1

Para a análise da Estratégia 1 (Subsecção 4.4.1), referente à incorporação da “personalidade dos itens” em algoritmos de FC, seguiu-se um método similar ao que foi empregado no experimento referente à questão de pesquisa QP2 (Subsecção 5.2.2). Para identificar a variação dos algoritmos analisados no experimento (com essa estratégia de agrupamento prévio de “itens com traços de personalidade similares”), foi utilizado o termo “itps” (e associou-se a cor amarela nas figuras). Para identificar os algoritmos UserKNN, ItemKNN, SMF, BMF, e SVD++, nas figuras, foi utilizado o termo “base”. Para a avaliação comparativa, também foi apresentada a estratégia “tps” (definida na Seção 4.2 e avaliada na Subsecção 5.2.2).

É pertinente comentar que não foi apresentada, nesta subsecção, a avaliação da estratégia “itps” para TripAdvisorCompleto, pois, para tal conjunto de dados, não convém calcular a “personalidade dos itens”, dado que a média de *ratings* por item (1,18) é notadamente baixa. Em outras palavras, há vários itens pouco avaliados, que poderiam aparecer apenas no conjunto de teste (cenário de novo item no sistema: sem avaliação de outros usuários, ficando, assim, sem “perfil de personalidade de item”), não tendo sido tal item agrupado em um *cluster* — nesses casos, para a estratégia adotada, usou-se a média global dos *ratings* do conjunto.

Ao executar a estratégia “itps” para TripAdvisorCompleto, o uso da média global dos *ratings* foi de 80,4%, ou seja, se fosse comparar, não seria analisado de fato o impacto da estratégia “itps”. Com o AmazonFoodsCompleto, houve também esse cenário, mas com menor impacto (14,7% de uso da média global), visto que a média de *ratings* por item é

de 4,86. Para TripAdvisor, o uso da média global foi de apenas 0,435%, e de 0,014% para AmazonFoods.

Considerando o conjunto de dados TripAdvisor, têm-se os *boxplots* da Figura 5.3. Com esse conjunto de dados, comparando em relação ao UserKNN, houve diferença estatística (par a par) para a métrica RMSE, mas não para *F-Measure*. Os resultados dos testes estatísticos estão no Apêndice D (Tabela D.7). Nas ocorrências de diferença estatística, o melhor desempenho (menor RMSE) foi das abordagens considerando personalidade, tendo “itps” melhor desempenho que “tps”. Comparando em relação ao ItemKNN, o resultado foi similar ao que houve com o UserKNN: houve diferença estatística para a métrica RMSE, mas não para *F-Measure*, com os menores valores de RMSE para as abordagens considerando personalidade (e “itps” melhor que “tps”). Em relação a SMF, BMF e SVD++, não houve diferença estatística para ambas as métricas.

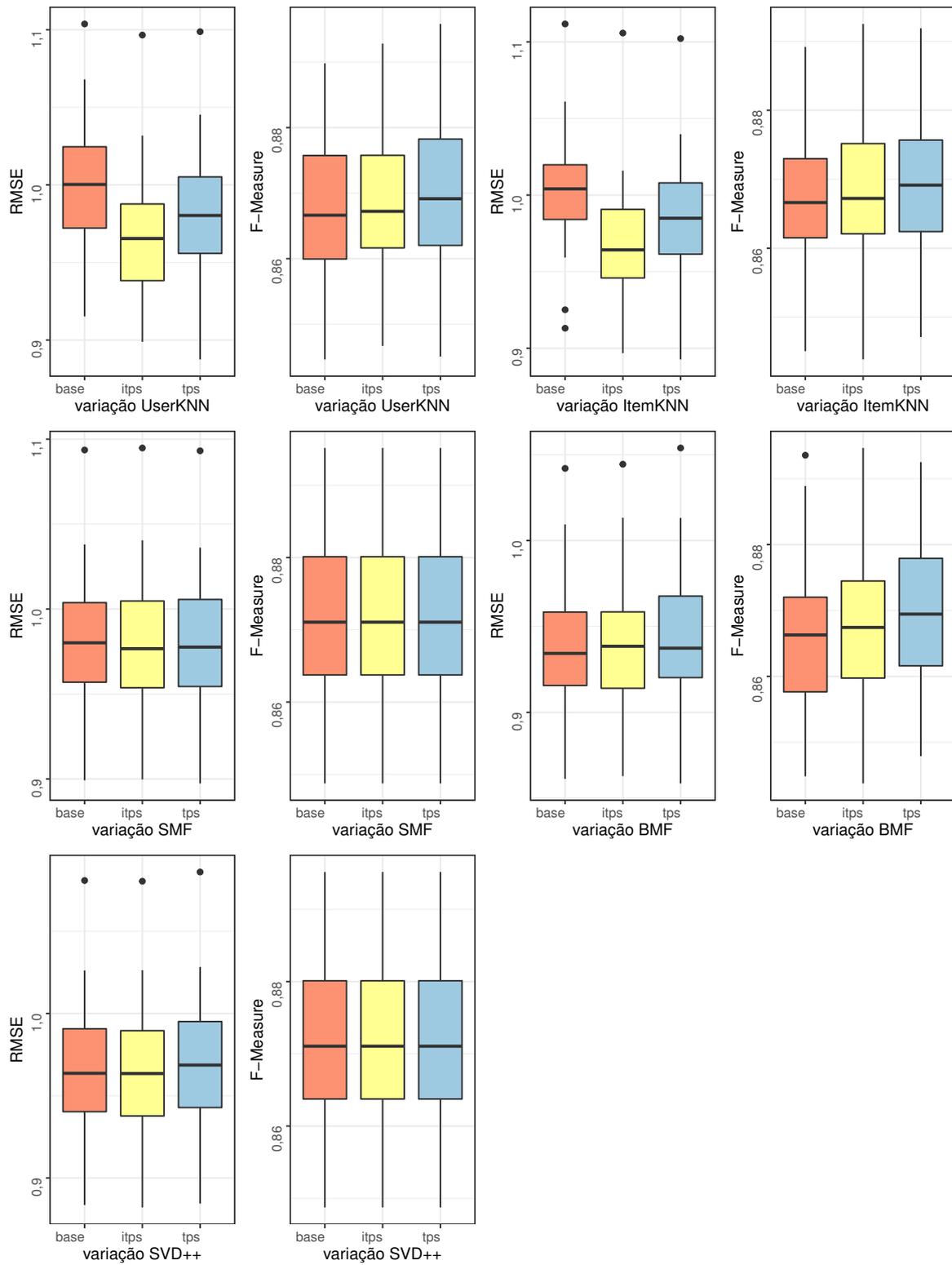
Considerando-se os conjuntos de dados AmazonFoods e AmazonFoodsCompleto, têm-se os *boxplots* das Figuras 5.4 e 5.5, respectivamente.

Comparando em relação ao UserKNN, tanto para o conjunto de dados reduzido quanto para o conjunto de dados completo, houve diferença estatística tanto para a métrica RMSE quanto para *F-Measure*. Percebe-se melhor desempenho (menor RMSE e maior *F-Measure*) das abordagens considerando personalidade, sendo “tps” melhor que “itps”.

Comparando em relação ao ItemKNN, em geral, também percebe-se melhor desempenho (menor RMSE e maior *F-Measure*) das abordagens considerando personalidade. Para o conjunto de dados reduzido e para o conjunto de dados completo, em geral, houve diferença estatística para a métrica RMSE (apenas não houve diferença entre “itps” e “tps” no AmazonFoodsCompleto). Também não houve diferença estatística entre “tps” e “itps” para o conjunto de dados reduzido em relação a *F-Measure*, mas houve diferença dessas abordagens em relação à abordagem sem personalidade. Para *F-Measure* no conjunto de dados completo, a estratégia “itps” apresentou o melhor desempenho.

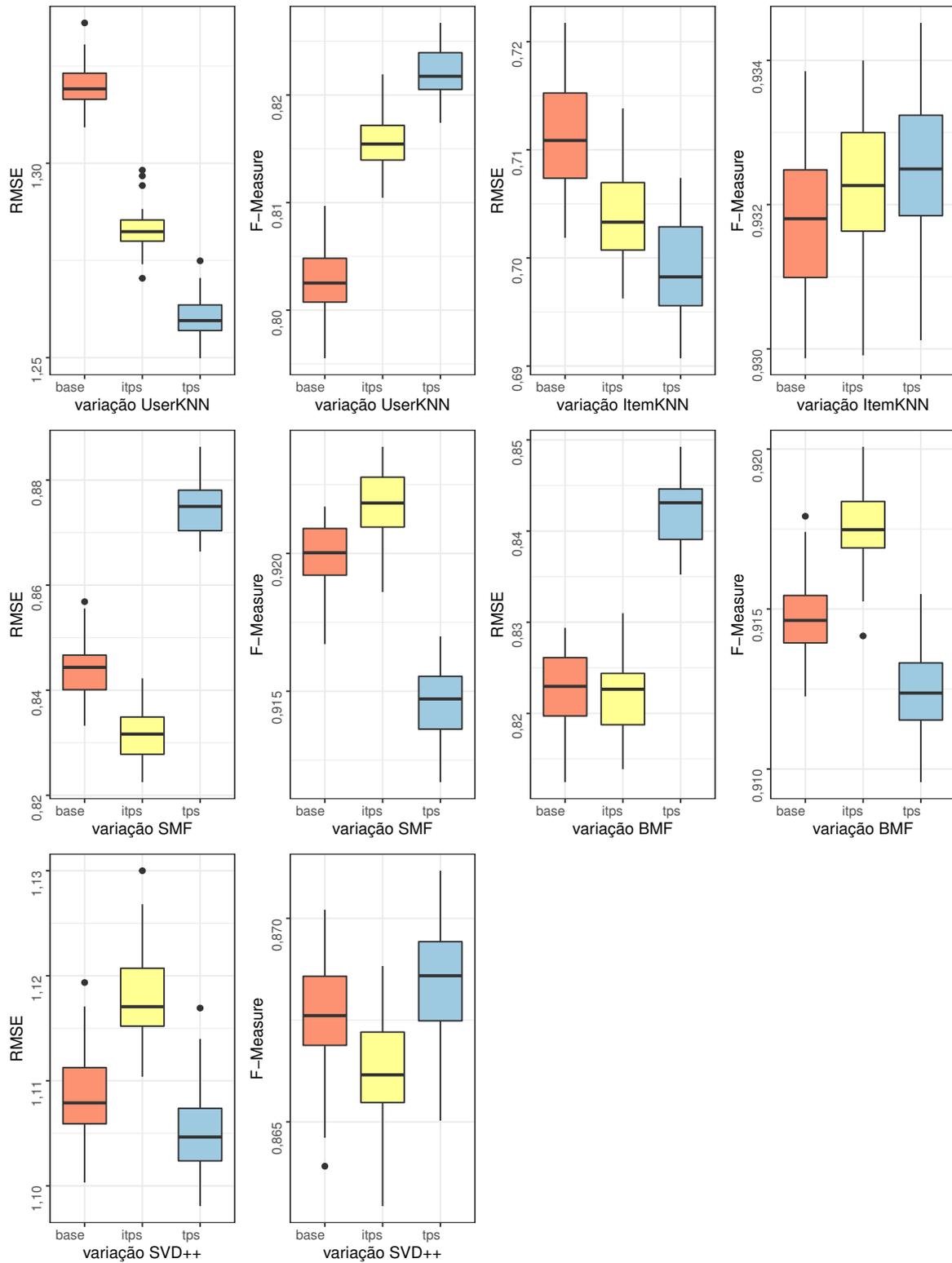
Comparando em relação ao SMF, tanto para a métrica RMSE quanto para *F-Measure*, e tanto para o conjunto de dados reduzido quanto para o conjunto de dados completo, houve diferença estatística. A estratégia “tps” considerando personalidade proporcionou pior desempenho (maior RMSE e menor *F-Measure*), mas a estratégia “itps” proporcionou o melhor desempenho.

Figura 5.3: *Boxplots* referentes à QP4 (estratégia “itps”) com o conjunto de dados TripAdvisor.



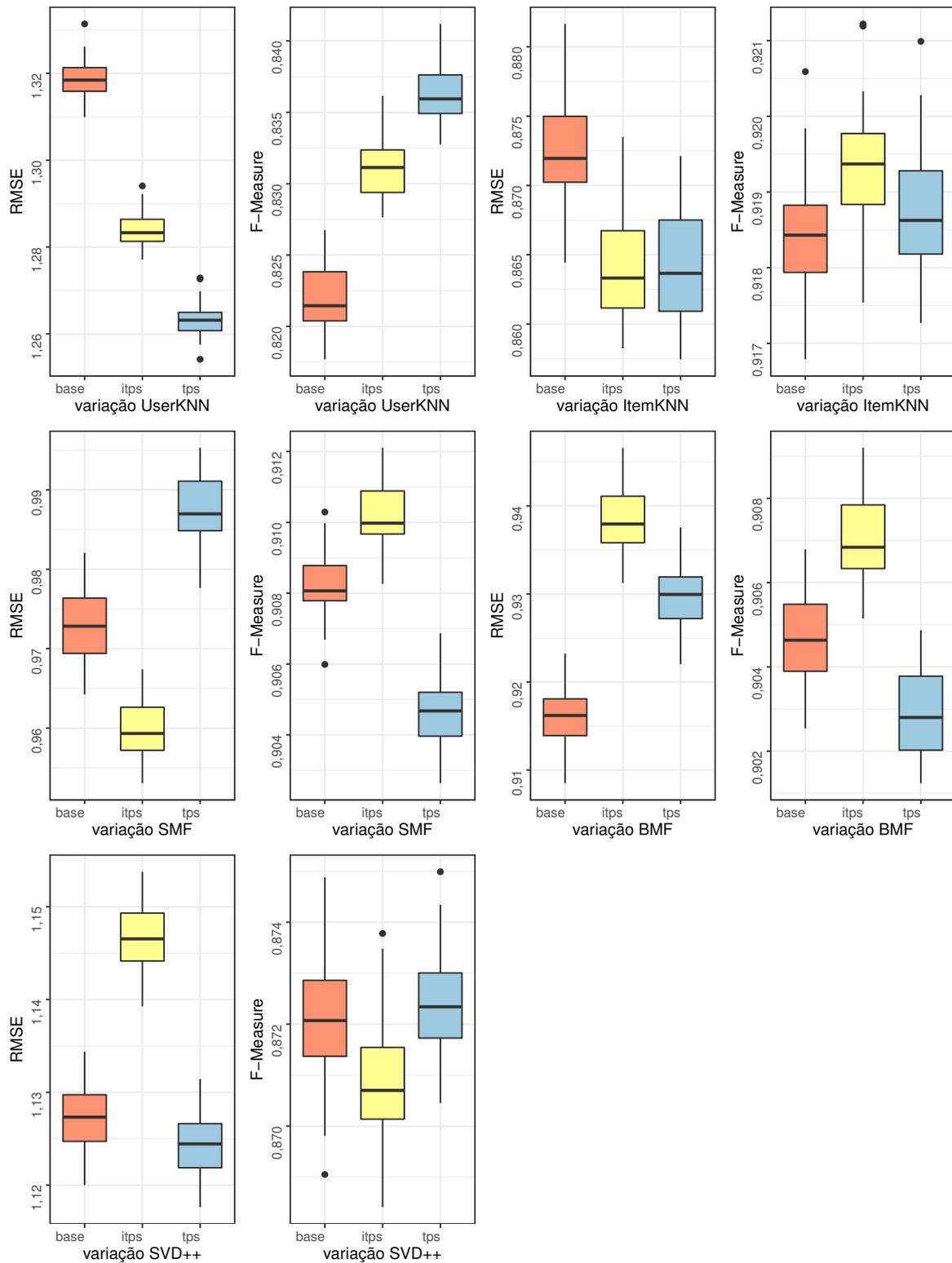
Fonte: autoria própria.

Figura 5.4: *Boxplots* referentes à QP4 (estratégia “itps”) com o conjunto de dados AmazonFoods.



Fonte: autoria própria.

Figura 5.5: *Boxplots* referentes à QP4 (estratégia “itps”) com o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Comparando em relação ao BMF, em relação à métrica *F-Measure*, o resultado foi similar ao que foi comentado com o SMF. Entretanto, em relação a RMSE, “itps” proporcionou pior desempenho no conjunto de dados AmazonFoodsCompleto, e proporcionou desempenho similar à abordagem sem personalidade no conjunto de dados AmazonFoods.

Comparando em relação ao SVD++, para a métrica *F-Measure* no conjunto de dados AmazonFoodsCompleto, não houve diferença estatística entre a estratégia “tps” e a abordagem sem personalidade. Porém, para os outros casos analisados, houve diferença estatística, com melhor desempenho (menor RMSE e maior *F-Measure*) para a estratégia “tps”, e pior desempenho (maior RMSE e menor *F-Measure*) para a estratégia “itps”.

Os resultados descritos anteriormente estão resumidos no Quadro 5.2, indicando a variação (“base” ou “tps” ou “itps”) — para cada algoritmo de FC em análise — que estatisticamente proporcionou melhor resultado, em cada cenário analisado (considerando-se os diferentes conjuntos de dados e métricas, com o modelo de personalidade *Big Five*). Os termos realçados na cor cinza indicam que, em tal cenário, não houve diferença estatística de desempenho em relação às variações citadas (caracterizando um empate de melhor desempenho com o baseline). O termo “tps|itps” (com realce na cor verde) indica que, em tal cenário, não houve diferença estatística de desempenho em relação às variações “tps” ou “itps”, mas proporcionaram melhor desempenho que o baseline.

Quadro 5.2: Síntese dos resultados referentes à QP4 (Estratégia 1).

Conjunto de dados	Algoritmo	RMSE	F-Measure
TripAdvisor	UserKNN	itps	base tps itps
	ItemKNN	itps	base tps itps
	SMF	base tps itps	base tps itps
	BMF	base tps itps	base tps itps
	SVD++	base tps itps	base tps itps
AmazonFoods	UserKNN	tps	tps
	ItemKNN	tps	tps itps
	SMF	itps	itps
	BMF	base itps	itps
	SVD++	tps	tps
AmazonFoodsCompleto	UserKNN	tps	tps
	ItemKNN	tps itps	itps
	SMF	itps	itps
	BMF	base	itps
	SVD++	tps	base tps

Fonte: autoria própria.

Em síntese, sobre a Estratégia 1 (“itps”), tem-se resultados variando em relação ao algoritmo base. Em relação a UserKNN, apesar de “itps” ter ganhos comparados ao baseline, seria mais pertinente aplicar a estratégia “tps”. Em relação a ItemKNN, “itps” e “tps” foram praticamente equivalentes, então, pela simplicidade da estratégia, também seria preferível aplicar “tps”. Em relação a SMF, “tps” tende a piorar o baseline, mas a estratégia “itps” demonstrou possibilidade de ganhos. Tal aspecto em relação a SMF não fica tão evidente em relação a BMF, mas, com *F-Measure* neste baseline, percebe-se comportamento similar ao comentado anteriormente (“tps” tendendo a piorar, mas “itps” tendendo a melhorar o baseline). Em relação a SVD++, há o comportamento inverso ao comentado sobre SMF (o “tps” tende a melhorar o baseline, mas o “itps” piora). Portanto, a opção de aplicar “itps” dependeria do algoritmo base a ser empregado.

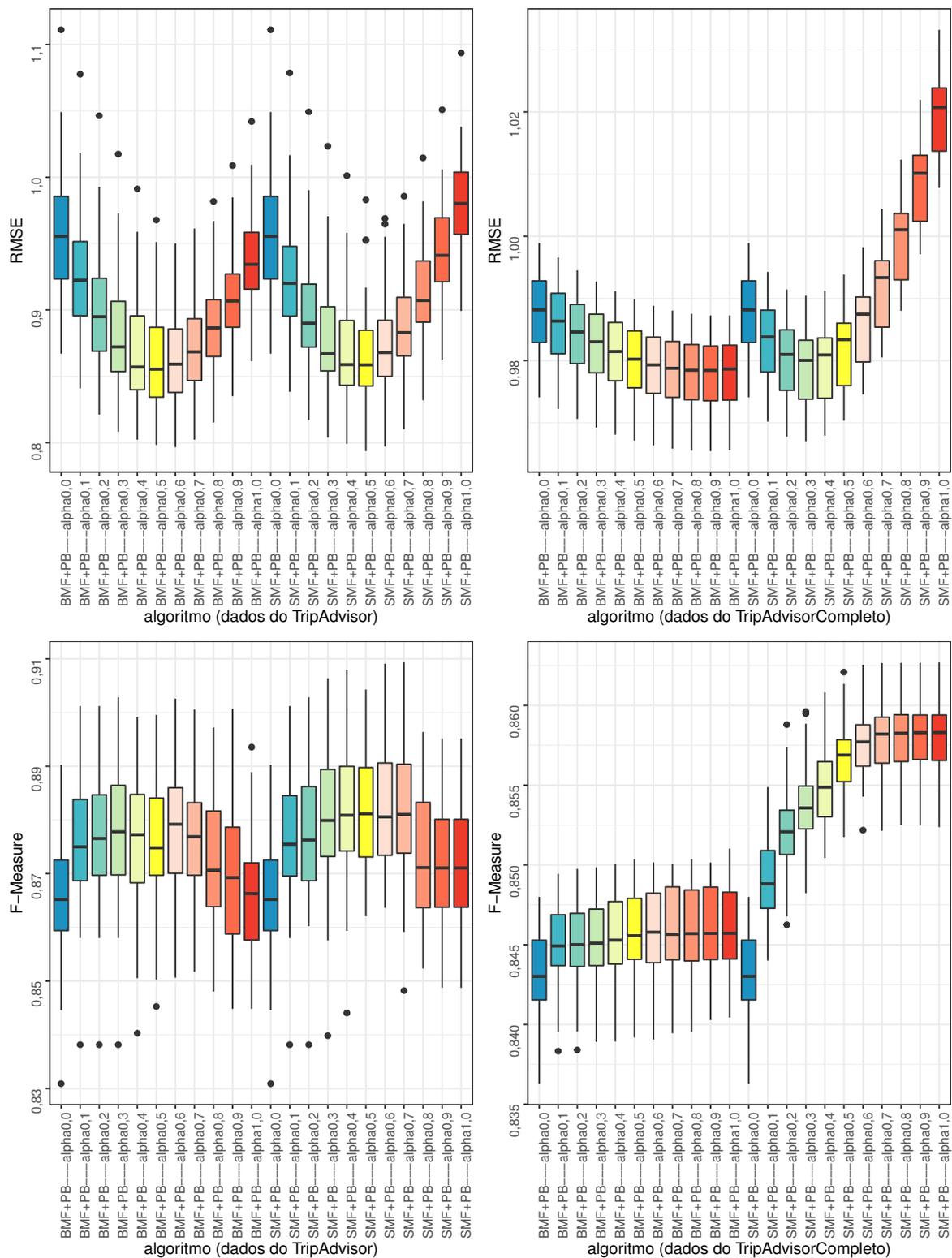
### Análise da Estratégia 2

Para a análise da Estratégia 2 (Subseção 4.4.2), considerando-se que a estratégia engloba um algoritmo de fatoração de matriz e uma abordagem baseada em personalidade, foram executadas variações referentes a SMF e BMF (algoritmos de fatoração de matriz que, nos experimentos referentes à QP2 e à QP3, foram melhores “puros” do que com personalidade). Foi considerado essencialmente o modelo *Big Five* (ou seja:  $w_1 = 1$  e  $w_2 = w_3 = 0$ ), e o valor de  $\alpha$  variou de 0,0 a 1,0, com incrementos de 0,1.

Nos *boxplots* das Figuras 5.6 e 5.7 — considerando-se o significado do valor de  $\alpha$  comentado na Subseção 4.4.2 —, as variações BMF+PB---alpha0,0 e SMF+PB---alpha0,0 (em azul) representam, em outras palavras, apenas o algoritmo PB. Por outro lado, as variações BMF+PB---alpha1,0 e SMF+PB---alpha1,0 (em vermelho) consistem, respectivamente, nos algoritmos BMF e SMF (sem influência de personalidade). A variação ponderando igualmente os algoritmos da estratégia híbrida ( $\alpha = 0,5$ ) está destacada em amarelo, para facilitar a visualização.

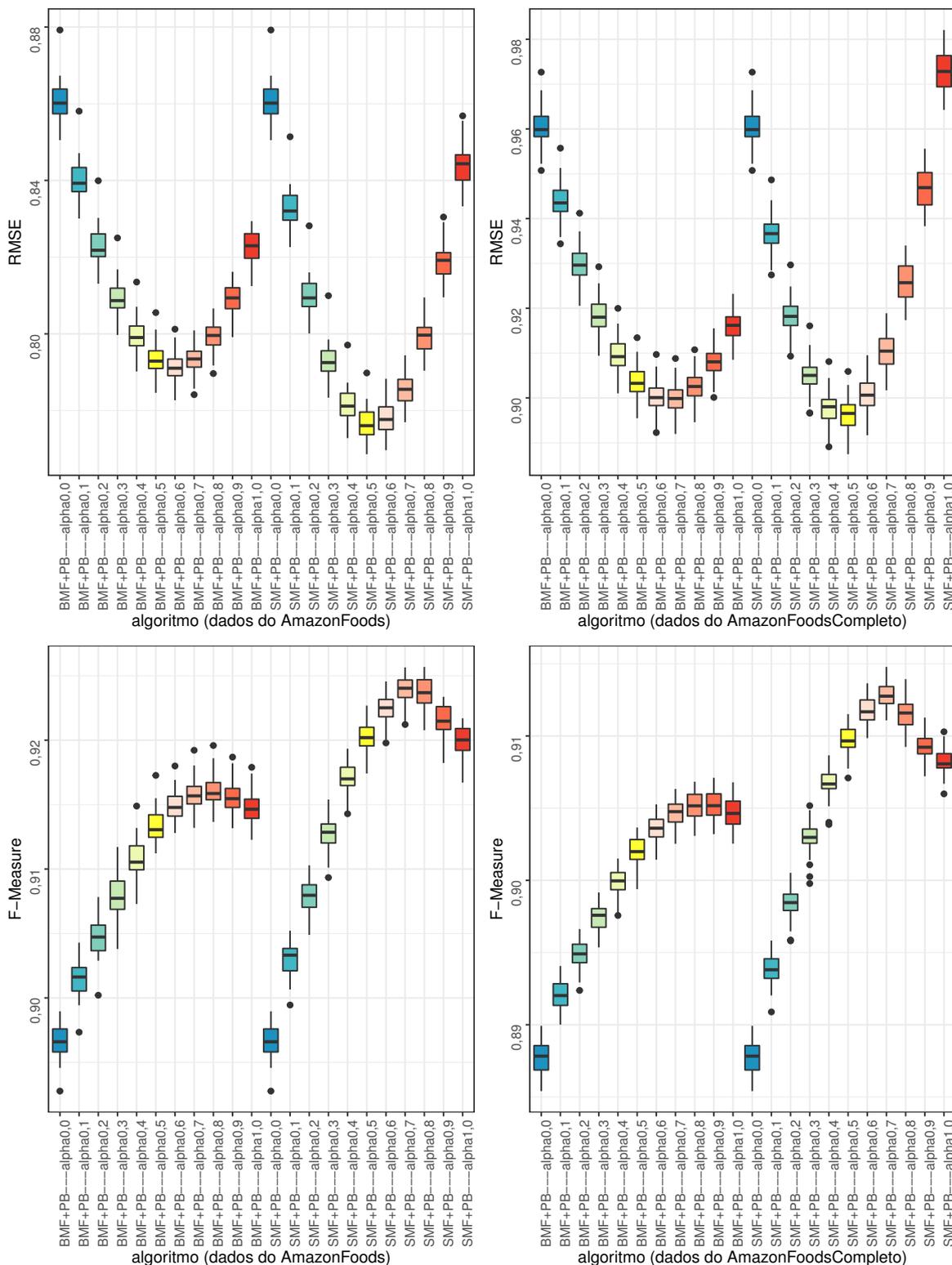
De acordo com os dados ilustrados nas Figuras 5.6 e 5.7, em resumo, a estratégia híbrida (*weighted ensemble*) apresentou geralmente alguma melhoria, especialmente em relação aos conjuntos de dados AmazonFoods e AmazonFoodsCompleto, e em relação à métrica RMSE. Os valores resultantes dos testes estatísticos encontram-se no Apêndice D (Tabela D.8).

Figura 5.6: *Boxplots* referentes à QP4 (estratégia *weighted*) com os conjuntos de dados TripAdvisor e TripAdvisorCompleto.



Fonte: autoria própria.

Figura 5.7: *Boxplots* referentes à QP4 (estratégia *weighted*) com os conjuntos de dados AmazonFoods e AmazonFoodsCompleto.



Fonte: autoria própria.

No conjunto de dados TripAdvisorCompleto, a estratégia híbrida não melhorou BMF, mas houve melhoria em relação ao RMSE ao aplicar a estratégia com SMF; por outro lado, no conjunto de dados TripAdvisor, tanto para BMF quanto SMF, há melhorias com a estratégia híbrida (especialmente com valores de  $\alpha$  de 0,4 a 0,6). Nos conjuntos de dados AmazonFoodsCompleto e AmazonFoods, com SMF, destaca-se o híbrido com valores de  $\alpha$  de 0,5, a 0,9; com BMF, destaca-se o híbrido com valores de  $\alpha$  de 0,7 a 0,9.

O foco no modelo *Big Five* neste experimento foi motivado pelos resultados apresentados na Seção 5.2.1 (referente à QP1). Todavia, foram realizadas também outras análises referentes a esta estratégia híbrida. No artigo publicado no WebMedia (AGUIAR; ARAÚJO; COSTA, 2020a), foram analisadas variações empregando além do modelo *Big Five* (por exemplo:  $w_1 = w_2 = w_3 = 1$ ;  $w_1 = w_2 = 1, w_3 = 0$ ;  $w_1 = 1, w_2 = 0, w_3 = 2$ ), com foco no conjunto de dados AmazonFoods. Em síntese, não houve diferença estatística entre considerar apenas o modelo *Big Five* ou considerá-lo junto aos modelos *Needs* e/ou *Values*.

### Análise da Estratégia 3

Para a análise experimental da Estratégia 3 (Subseção 4.4.3), foram avaliadas as listas finais de recomendação (foco da proposta ao empregar *re-ranking*). Nesse sentido, foram empregadas estas métricas como variáveis dependentes: *F-Measure* (para o *Top 5* das listas), MAP, MRR e NDCG<sup>8</sup> — métricas relacionadas à relevância (acurácia) da lista de itens recomendados. Além disso, considerou-se a métrica MSI (para o *Top 1*, o *Top 5* e o *Top 10* das listas), comentada na Seção 2.2, para avaliar o aspecto da novidade nessas listas. Para tais métricas, quanto maior o valor, melhor desempenho.

Para avaliar a estratégia (com métricas considerando, por exemplo, o *Top 5* da lista final de recomendações), foram considerados os primeiros 30 itens ranqueados pelo algoritmo BPRMF ( $n = 30$ ) e, assim como na Subseção 5.2.4, o valor de  $\alpha$  variou de 0,0 a 1,0, com incrementos de 0,1.

No conjunto de dados TripAdvisor, não houve mudanças significativas estatisticamente ao variar o  $\alpha$  no tocante às métricas de acurácia (Apêndice D — Tabela D.9), mas os menores

---

<sup>8</sup>MAP (*Mean Average Precision*), MRR (*Mean Reciprocal Rank*) e NDCG (*Normalized Discounted Cumulative Gain*) são métricas úteis para avaliar conjuntos de itens ordenados. A métrica MAP consiste na média das precisões médias (*average precision*) referentes aos itens da lista. A métrica MRR avalia se o primeiro elemento recuperado de uma lista é um elemento relevante. A métrica NDCG avalia a relevância dos itens considerando sua posição na lista ordenada (AGGARWAL et al., 2016)

valores de  $\alpha$  (maior impacto da personalidade no processo) resultaram em maior MSI (Figura 5.8). No conjunto de dados TripAdvisorCompleto, ao variar o valor de  $\alpha$ , há algumas mudanças referentes à acurácia (alguns valores de  $\alpha$  proporcionando diminuição no desempenho), mas há valores de  $\alpha$  mantendo a acurácia e impactando positivamente na métrica MSI (Figura 5.9).

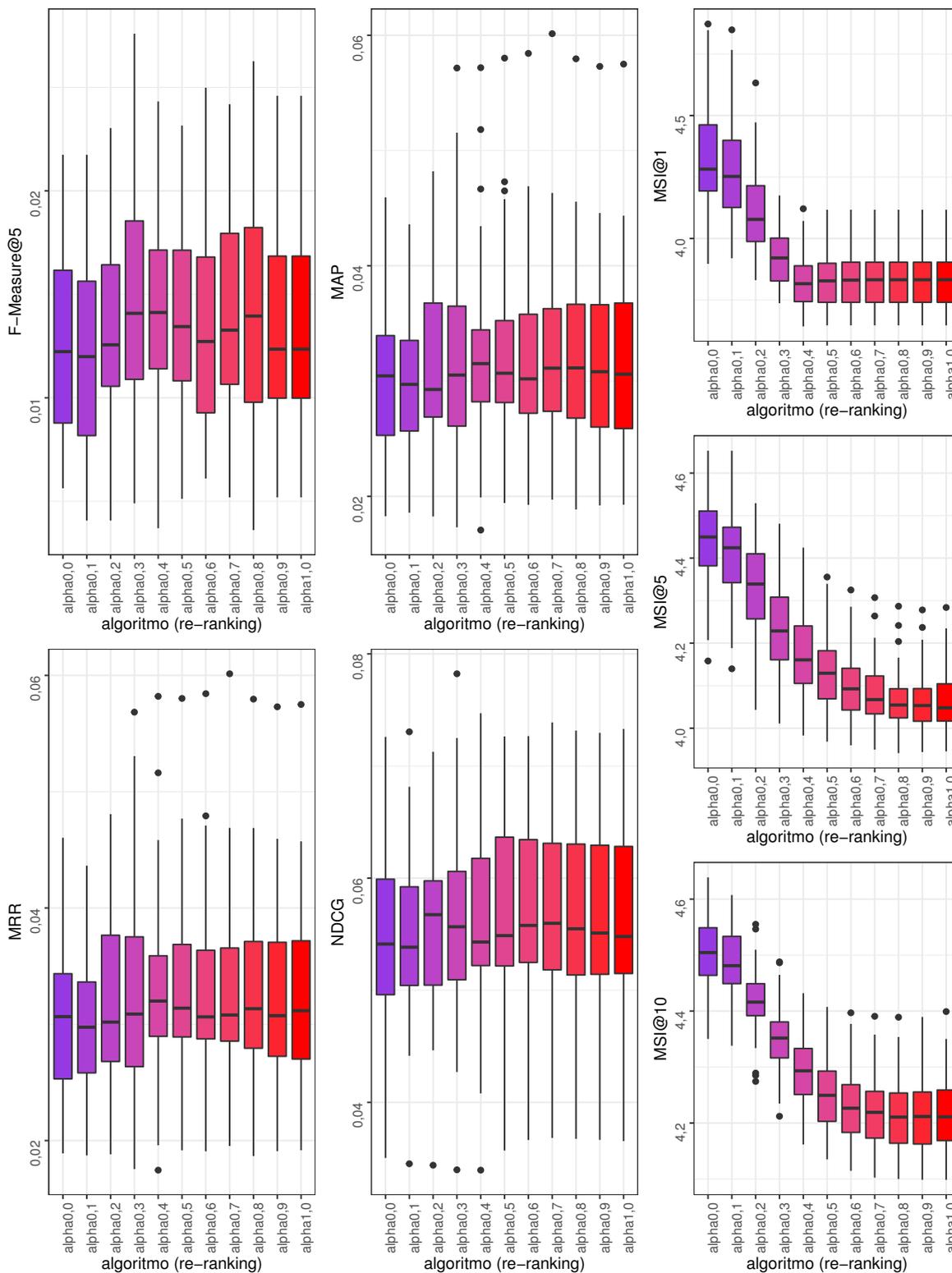
No conjunto de dados AmazonFoods, o *re-ranking* baseado em personalidade resultou em melhoria da acurácia (métricas *F-Measure*, MAP, MRR e NDCG) especialmente com  $\alpha = 0,2$  e  $\alpha = 0,3$ . Além disso, com esses valores de  $\alpha$ , há ganho na novidade (MSI), além de ter ainda mais ganho com  $\alpha$  tendendo a zero (Figura 5.10). No conjunto de dados AmazonFoodsCompleto, com  $\alpha$  tendendo a zero, há uma tendência de decréscimo nas métricas referentes à acurácia enquanto há ganho em novidade (Figura 5.11).

Apesar de as análises nesta pesquisa considerarem cenários gerais, foi adicionalmente analisado o re-ranqueamento por grupos de usuários (*clusters* referentes aos traços de personalidade), com o objetivo de identificar diferenças por *cluster*, especialmente nos casos em que se observou decréscimo, nas métricas referentes à acurácia, ao diminuir o valor de  $\alpha$ .

Em geral, as “curvas” foram similares às referentes aos cenários gerais anteriormente comentados (Figuras 5.8 a 5.11), especialmente referente à métrica MSI e aos dados oriundos do TripAdvisor. Contudo, houve alguns casos com esta situação: “curvas” diferentes para os *clusters*, não havendo decréscimo na acurácia para determinado *cluster* (sendo o decréscimo no outro *cluster* o motivo do decréscimo no cenário geral).

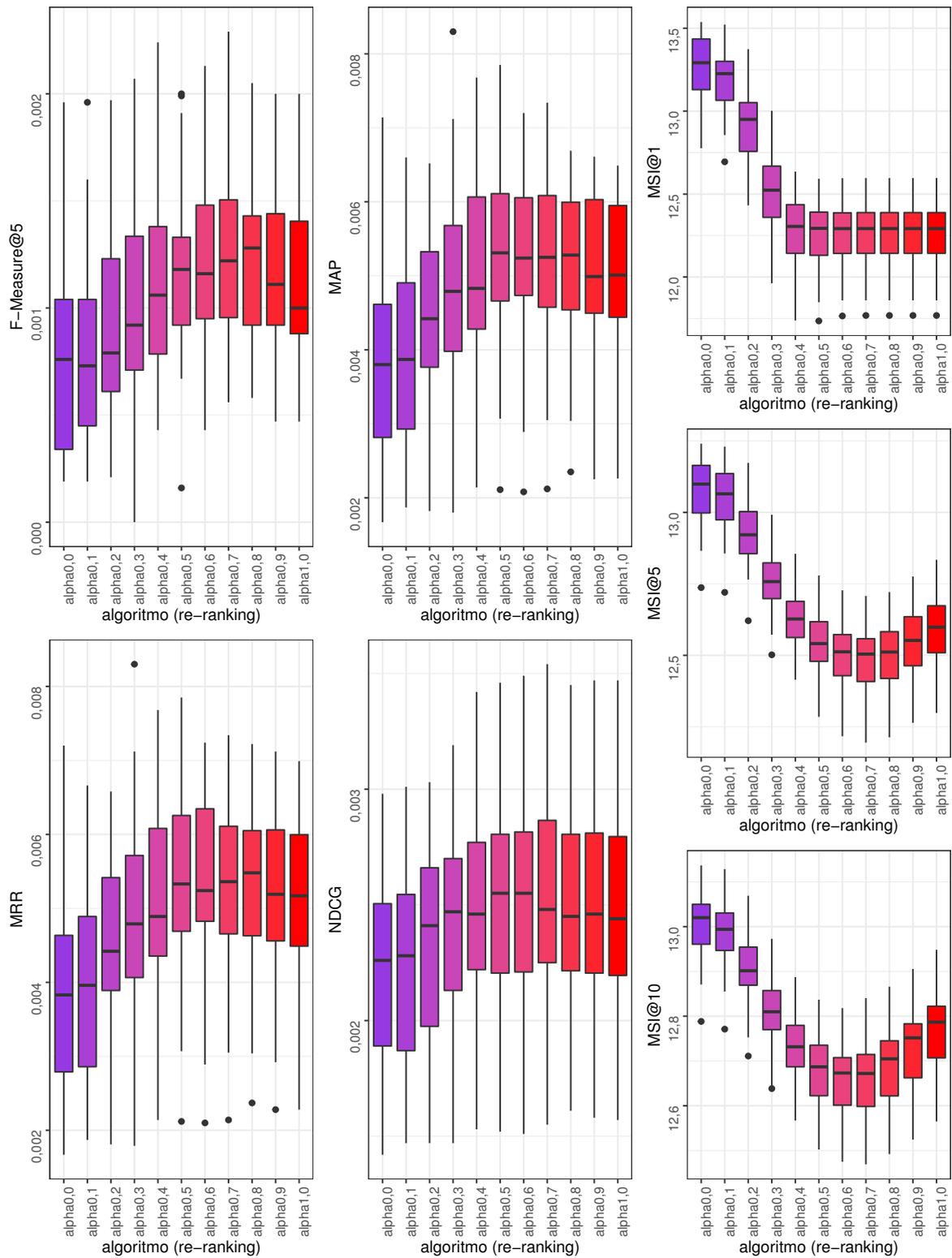
Tal situação foi observada nos conjuntos de dados AmazonFoods (Figura 5.12) e AmazonFoodsCompleto (Figura 5.13), especificamente em relação aos *clusters* referentes ao traço Amabilidade (quarto traço do *Big Five*). Os *re-rankings* para os usuários do *cluster* com maiores valores em Amabilidade (aproximadamente 42% dos usuários, representados na cor amarela nos *boxplots* elaborados) tenderam a manter — ou até a melhorar — a acurácia, o que difere em relação aos usuários do *cluster* com menores valores em Amabilidade (aproximadamente 58% dos usuários, representados na cor verde nos *boxplots* elaborados).

Figura 5.8: *Boxplots* referentes à QP4 (estratégia *re-ranking*) com o conjunto de dados TripAdvisor.



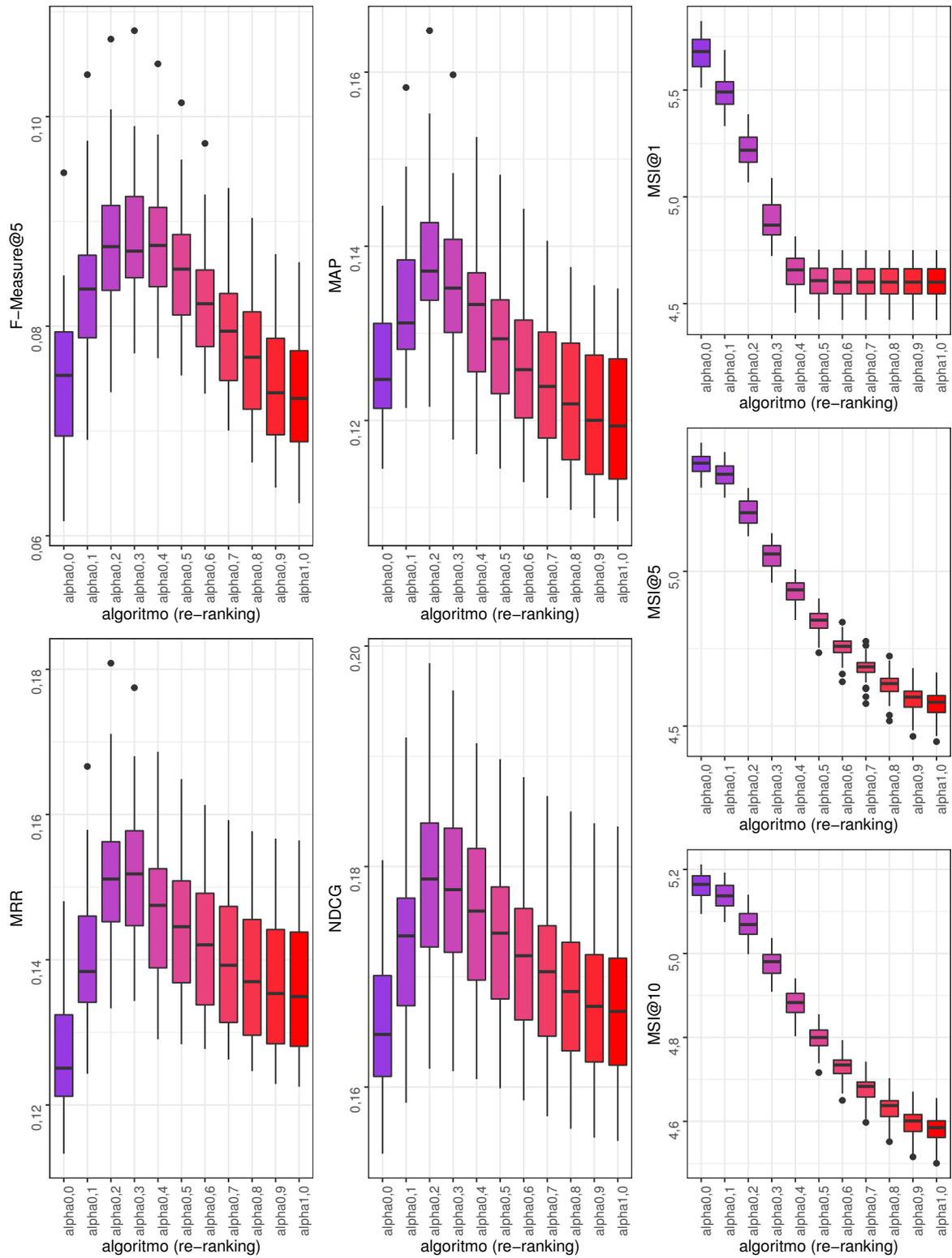
Fonte: autoria própria.

Figura 5.9: *Boxplots* referentes à QP4 (estratégia *re-ranking*) com o conjunto de dados TripAdvisorCompleto.



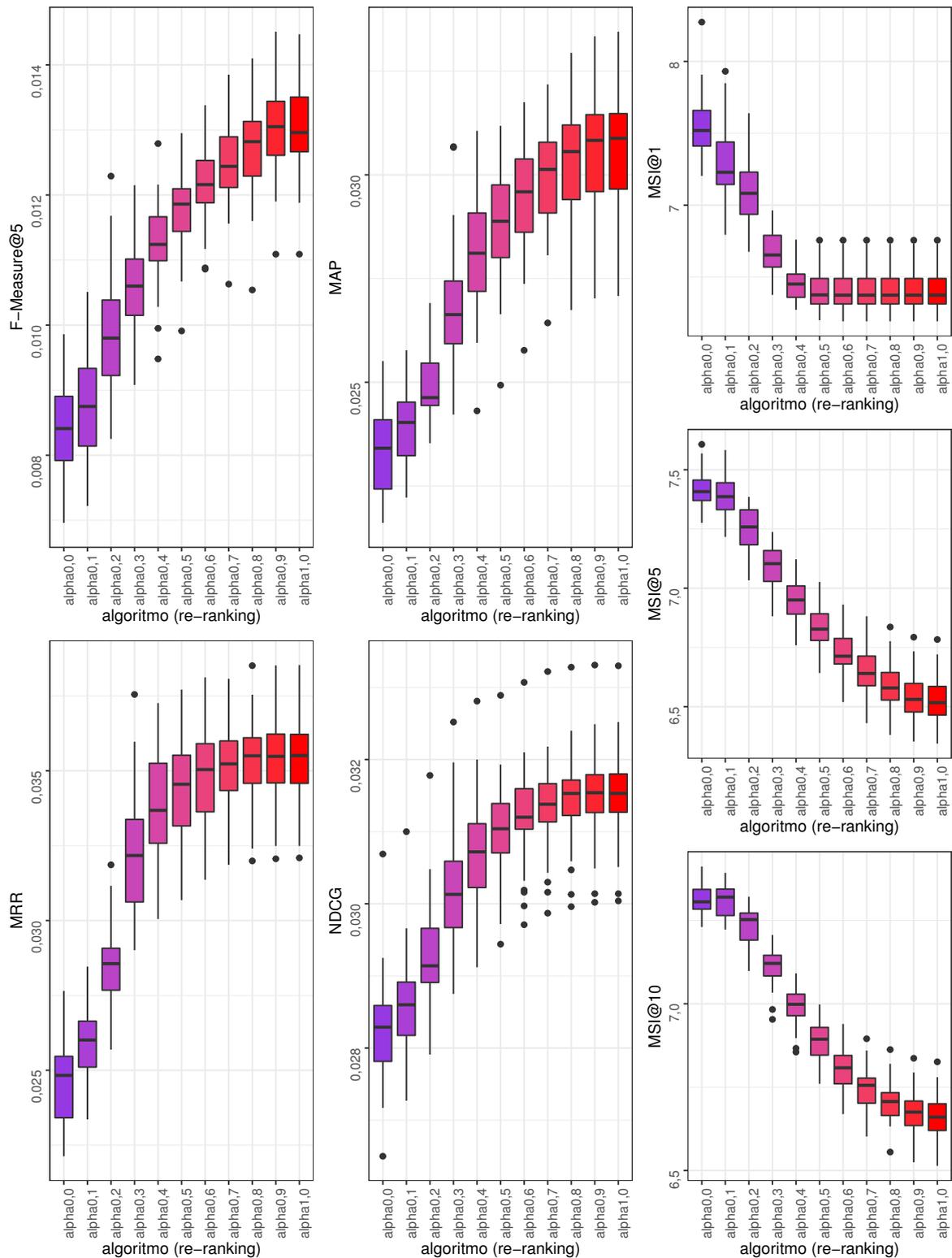
Fonte: autoria própria.

Figura 5.10: *Boxplots* referentes à QP4 (estratégia *re-ranking*) com o conjunto de dados AmazonFoods.



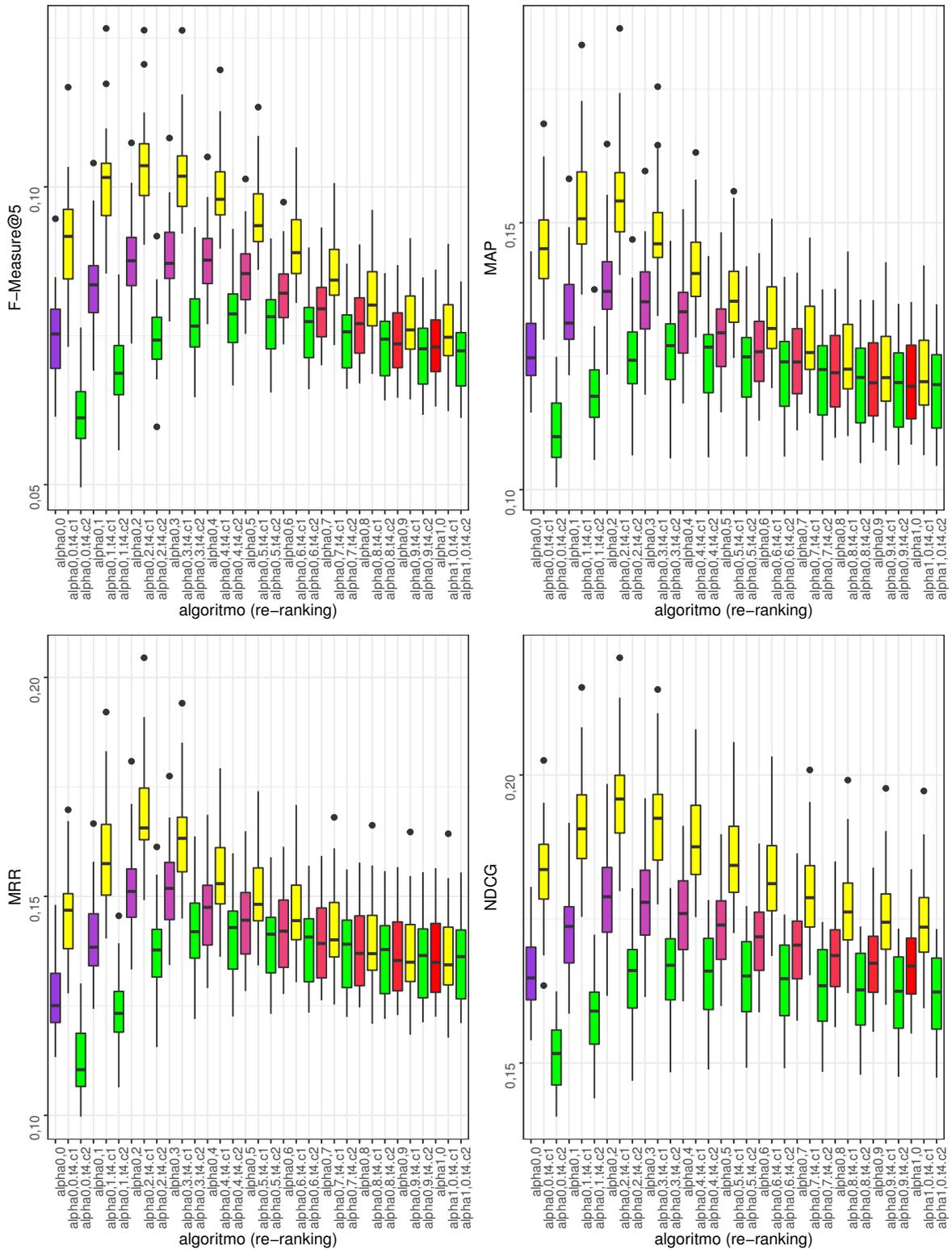
Fonte: autoria própria.

Figura 5.11: *Boxplots* referentes à QP4 (estratégia *re-ranking*) com o conjunto de dados AmazonFoodsCompleto.



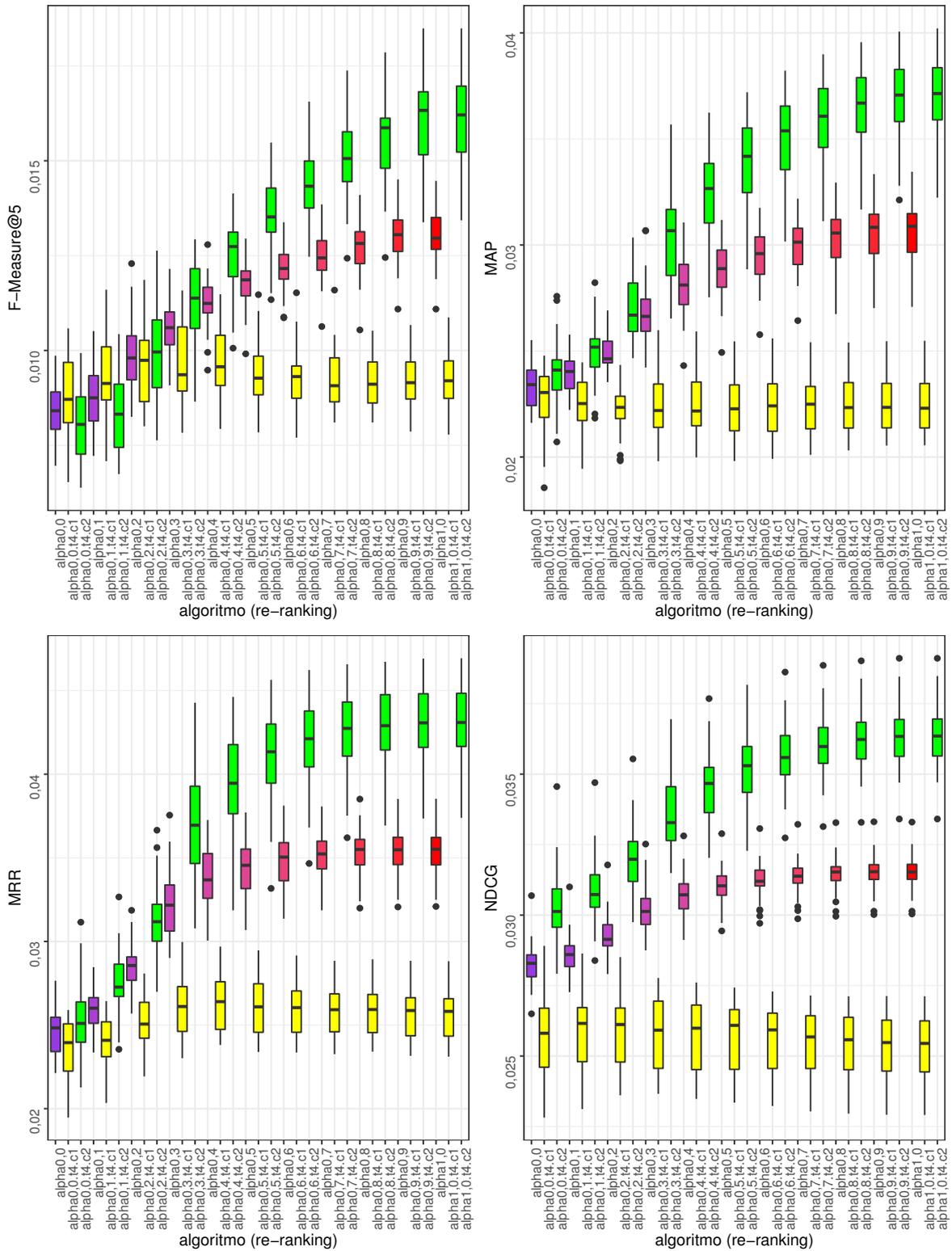
Fonte: autoria própria.

Figura 5.12: *Boxplots* referentes à QP4 (estratégia *re-ranking*) com o conjunto de dados AmazonFoods, destacando avaliação por *cluster*, com o traço Amabilidade (t4).



Fonte: autoria própria.

Figura 5.13: *Boxplots* referentes à QP4 (estratégia *re-ranking*) com o conjunto de dados AmazonFoodsCompleto, destacando avaliação por *cluster*, com o traço Amabilidade (t4).



Fonte: autoria própria.

Com os resultados desta Subseção 5.2.4, referentes à relevância das recomendações oriundas das estratégias propostas, é possível responder positivamente à questão de pesquisa QP4. Em geral, os resultados da avaliação experimental reforçam ter algum impacto ao empregar conceitos referentes às características da personalidade dos usuários em SR.

Os resultados obtidos na investigação realizada indiciam ser pertinente inserir aspectos de personalidade em abordagens convencionais, por não haver, em geral, consideráveis perdas de acurácia e, especialmente, por possibilitar ganhos em outros aspectos atrativos do ponto de vista do usuário, como a novidade. Em resumo, no Quadro 5.3, encontram-se as informações resultantes da investigação experimental realizada nesta pesquisa.

### 5.3 Considerações Finais do Capítulo

Visando-se a realizar análises experimentais referentes aos objetivos desta tese (incluindo a avaliação das abordagens apresentadas no Capítulo 4), foi realizada a investigação empírica descrita neste capítulo.

Os resultados da QP1 contribuem como incentivo para o uso de APR em estudos sobre SR. Particularmente, os resultados contribuíram para estimular a continuidade desta pesquisa, motivando a realização das investigações referentes às demais questões (QP2, QP3 e QP4).

Os resultados da QP2 são úteis para que futuras pesquisas referentes a SR, especialmente as que adotarem alguma abordagem de FC baseada em vizinhança, possam refletir a possibilidade de considerar agrupamentos de usuários no tocante à sua personalidade (estratégia “tps”).

A investigação referente à QP3 é útil para pesquisas futuras em FC baseada em componentes da personalidade humana. Os resultados não corroboram a ideia de criar um SR ponderando os componentes do *Big Five*; e também não corroboram a ideia de criar um SR baseado em personalidade autorregulador (substituindo algum traço tradicional por alguma faceta específica do *Big Five* ou componente dos modelos *Values* e *Needs*). Os resultados da QP3 reforçam a importância conjunta das dimensões do tradicional modelo *Big Five* para a representação da personalidade do indivíduo (que pode ser explorada em outras subáreas da computação além de SR).

Quadro 5.3: Síntese da investigação experimental realizada.

Objetivos	Principais aspectos metodológicos	Resumo dos principais resultados
<b>O1:</b> Rever pesquisas iniciais de referência para verificar se os resultados se mantêm ao obter dados de personalidade via APR (baseado em texto) e conjuntos de dados com maior número de usuários e itens.	Análise de abordagens elaboradas relacionadas às pesquisas de Tkalčič et al. (2009) e Hu e Pu (2011), mesclando estratégias e variando modelos de personalidade ( <i>Big Five</i> , <i>Needs</i> e <i>Values</i> ).	Foram mantidos alguns resultados de pesquisas iniciais relevantes concernentes a FC baseada em personalidade — as abordagens <i>personality-based</i> proporcionaram, no mínimo, acurácia similar às abordagens <i>rating-based</i> .
<b>O2:</b> Analisar se a acurácia das predições de um algoritmo de FC se mantém (ou tende a melhorar) ao desconsiderar os “dados de opinião” ( <i>ratings</i> ) de usuários menos similares quanto à personalidade (identificada automaticamente via textos) de um usuário-alvo de recomendação.	Análise de uma proposta genérica de agrupamento com <i>k-means</i> (nomeada “tps”), aplicada em diferentes algoritmos de FC (UserKNN, ItemKNN, SMF, BMF e SVD++), para realizar predições baseadas nos <i>ratings</i> de usuários mais similares quanto às características de personalidade (componentes dos modelos <i>Big Five</i> , <i>Needs</i> e <i>Values</i> ).	É possível haver efeito na acurácia das predições de um algoritmo de FC quando este é executado desconsiderando a opinião de usuários menos similares a um usuário-alvo quanto à personalidade (identificada via APR baseado em texto). Em geral, com o agrupamento proposto, foi proporcionado efeito positivo na acurácia dos algoritmos baseados em vizinhança. Em alguns cenários com algoritmos baseados em modelo, foi proporcionado efeito negativo na acurácia.
<b>O3:</b> Analisar a influência dos traços de personalidade (identificados via APR baseado em texto) visando a destacar se algum(ns) deles influencia(m) mais fortemente na melhoria dos resultados dos algoritmos de FC baseada em personalidade.	Análise, em diferentes algoritmos base (UserKNN, ItemKNN, SMF, BMF e SVD++), de variações elaboradas diversificando a influência — individual, agrupada e ponderada — dos traços do modelo <i>Big Five</i> , na FC baseada em personalidade. Análise complementar (incluindo uso de GA) abrangendo as facetas do <i>Big Five</i> e as dimensões dos modelos <i>Needs</i> e <i>Values</i> .	Os traços do <i>Big Five</i> podem influenciar de maneira diferente na acurácia dos algoritmos de FC baseada em personalidade, mas, de modo geral, as variações analisadas (ignorando traços ou ponderando-os) proporcionaram redução de desempenho ao comparar com a variação tradicional (atuação conjunta e igualmente ponderada dos traços). Na análise complementar (abrangendo características além dos cinco traços do <i>Big Five</i> ), foi possível encontrar, em alguns cenários, alguma variação com desempenho estatisticamente melhor que a variação tradicional.
<b>O4:</b> Propor e avaliar estratégias para a geração de recomendações personalizadas baseadas nas características de personalidade dos usuários e suas preferências anteriores, indicando se algoritmos de FC geram resultados mais relevantes ao considerar explicitamente a personalidade dos usuários, identificada automaticamente via textos.	Proposta de estratégias híbridas (motivadas pela revisão de literatura e pelos resultados das análises referentes a O1, O2 e O3): a Estratégia 1 (“itps”), referente à “personalidade de itens”; a Estratégia 2 ( <i>weighted</i> ), considerando fatoração de matriz (via SMF ou BMF) e vizinhança baseada em personalidade; e a Estratégia 3 ( <i>re-ranking</i> ), considerando fatoração de matriz (via BPRMF) aliada à personalidade dos usuários. Avaliação da Estratégia 1 aplicada em diferentes algoritmos de FC (UserKNN, ItemKNN, SMF, BMF e SVD++), comparando-a também a “tps”. Avaliação das Estratégias 2 e 3 variando o parâmetro $\alpha$ que controla o percentual de contribuição do algoritmo de fatoração de matriz (SMF, BMF ou BPRMF).	É possível obter efeitos positivos na acurácia de algoritmos de FC ao aplicar as estratégias propostas (com personalidade identificada via APR baseado em texto). Sobre “itps”, os resultados variaram em relação aos cenários analisados. Nos algoritmos baseados em vizinhança, “itps” proporcionou resultados geralmente melhores (no mínimo similares) em comparação ao algoritmo base, mas por vezes não superou “tps”. Nos algoritmos baseados em modelo, por vezes foram observados casos de efeito positivo de “itps” em relação a SMF e BMF (diferentemente de “tps”, que proporcionou efeito negativo) e foram observados casos de efeito negativo de “itps” em relação a SVD++ (enquanto “tps” proporcionou resultados no mínimo similares ao algoritmo base). Sobre <i>weighted</i> , por vezes houve efeito positivo na acurácia ao atribuir para $\alpha$ um valor diferente de 0,0 e de 1,0 (ou seja, resultados melhores ao considerar ambos os algoritmos que compõem o híbrido em comparação ao resultado desses algoritmos individualmente). Sobre <i>re-ranking</i> , houve efeito positivo na acurácia da lista final de recomendação em poucos casos, mas, em geral, os menores valores de $\alpha$ (isto é, maior influência da personalidade) proporcionaram mais novidade (métrica MSI) na lista final de recomendação. Ademais, houve indícios de que a estratégia <i>re-ranking</i> seria mais relevante se não direcionada aos usuários do <i>cluster</i> com menores valores em Amabilidade.

Fonte: autoria própria.

A QP4 foi definida de modo dependente aos resultados das questões anteriores. Visto que, pela QP3, não se identificou uma combinação específica de traços — inclusive ao empregar GA — para se ter uma mudança significativamente positiva no desempenho, não foi elaborada para a QP4 uma estratégia baseada em GA, como inicialmente idealizado. Todavia, foram elaboradas propostas com potencial de melhorar os SR ao serem incorporadas no processo de recomendação.

Sobre a Estratégia 1 referente à QP4, a estratégia “itps” não se mostrou a melhor opção para aplicar com os algoritmos baseados em vizinhança e o SVD++ (sendo melhor aplicar a estratégia “tps”), mas demonstrou ser pertinente aplicá-la com os outros algoritmos baseados em modelo (que tendiam a piorar com “tps”); assim, é conveniente que futuros estudos considerem aplicar agrupamentos no tocante à personalidade, avaliando a aplicação da estratégia “itps” caso a “tps” não seja satisfatória.

Ainda sobre a QP4, mostrou-se a possibilidade de melhorias no processo de recomendação com o *weighted ensemble* (Estratégia 2) e com o *re-ranking* (Estratégia 3). Especificamente sobre a estratégia *re-ranking*, é pertinente comentar que esta não é a estratégia pioneira envolvendo personalidade. Lex et al. (2021) e Tran, Felfernig e Tintarev (2021) destacam a estratégia de Lu e Tintarev (2018), mas esta é dependente das relações extraídas entre traços de personalidade e atributos dos itens (atributos musicais). Diferentemente disso, a Estratégia 3 independe de domínio.

Finalmente, com base na investigação experimental realizada, é possível afirmar que a principal hipótese formulada nesta tese (Seção 1.1) foi corroborada. Isto é: foi possível obter recomendações de itens mais acuradas para os usuários, em cenários gerais, ao incorporar informações referentes à personalidade, obtidas via APR baseado em textos (linguagem natural) desses usuários. Em síntese, foi percebido que, para a obtenção de efeitos positivos na acurácia das recomendações, é pertinente variar a estratégia baseada em personalidade a ser adotada, a depender do algoritmo de recomendação base considerado.

Dentre as abordagens elaboradas nesta tese, considerando-se a importância atual das abordagens de fatoração de matriz e os resultados analisados (sintetizados no Quadro 5.3), indica-se mais fortemente, para aplicação e novos estudos, a Estratégia 3 (Subseção 4.4.3), especialmente pela análise realizada, com esta estratégia *re-ranking* demonstrando ganhos em novidade aliados a efeitos positivos na acurácia com alguns valores do parâmetro  $\alpha$ .

# Capítulo 6

## Considerações Finais

Neste capítulo, são expostas as últimas considerações referentes a esta pesquisa, incluindo-se conclusões e limitações, além de direcionamentos para pesquisas futuras.

### 6.1 Conclusões

Sobre avançar o estado-da-arte, como comentado ao final da Seção 3.1, esta tese agrega valor à comunidade científica no sentido de possibilitar melhoria na acurácia dos SR — considerando características de personalidade sem uso de questionários e sem determinar um domínio específico (via FC) — em cenários gerais (diferentemente da maioria das pesquisas que possibilita melhoria na acurácia em cenários específicos, como situações de *cold-start*, ou melhoria na diversidade).

O uso de características de personalidade identificadas implicitamente é potencialmente útil para a personalização dos SR. Como expõem Jaques e Nunes (2021), os clientes habitualmente não estão interessados em questionários para as empresas melhorarem suas vendas, mas eles desejam, mesmo indiretamente, receber cada vez mais informações que satisfaçam suas expectativas.

Envolvendo a Computação da Personalidade (*Personality Computing*), esta tese contribui com a linha de pesquisa recentemente denominada *Psychology-informed Recommender Systems* (LEX et al., 2021). Quanto aos procedimentos técnicos realizados durante a investigação apresentada nesta tese, os Capítulos 2 e 3 associam-se à revisão bibliográfica, enquanto os Capítulos 4 e 5 associam-se à pesquisa experimental.

Considerando-se a revisão bibliográfica realizada, é possível concluir que a área envolvendo a personalidade humana é antiga, mas é recente a linha de pesquisa que envolve a personalidade em SR, com os primeiros estudos considerando APR datados por volta de 2015.

A revisão bibliográfica possibilitou o entendimento do estado-da-arte. Não foram identificadas muitas pesquisas apresentando ganhos expressivos na acurácia das recomendações dos sistemas baseados em personalidade, em âmbito geral. Comumente, as pesquisas nessa linha concentram-se em expor melhorias em âmbitos específicos: ao associar traços de personalidade a categorias de itens em determinados domínios; ao avaliar as recomendações do ponto de vista da diversidade; e ao analisar cenários de *cold-start*.

Nesse sentido, objetivou-se investigar a utilização de características da personalidade das pessoas (empregando-se APR baseado em texto e modelos além do tradicional *Big Five*) para melhorar, em cenários gerais, a acurácia das recomendações dos sistemas que aplicam FC no processo de recomendação. Para isso, foram elencadas questões de pesquisa (Seção 1.2) e, com a pesquisa experimental realizada, visou-se a respondê-las.

O uso de questionários pode dificultar a aplicabilidade de SR. Ao pedir ao usuário para responder a questionários sobre personalidade, perde-se a essência de recomendar algo a tal usuário sem precisar que este realize tarefas cansativas/entediadas. Embora o campo referente a APR ainda esteja avançando (por exemplo, há escassez de ferramentas validadas e facilmente disponíveis para identificar características de personalidade via textos em língua portuguesa), esta tese mostrou ser possível obter melhorias em FC ao incorporar dados de personalidade não oriundos dos tradicionais questionários.

É conveniente ressaltar que uma das principais motivações para a criação de SR baseados em personalidade consiste na premissa de que pessoas com personalidade similar têm preferências similares. Em outras palavras, pessoas semelhantes quanto à personalidade tenderiam a atribuir *ratings* similares ao avaliarem os mesmos itens. Isso motivou realizar a investigação, apresentada nesta tese, com foco na acurácia (com métricas como RMSE). Isso não significa que outras métricas (como as referentes à diversidade e à novidade) não são importantes para avaliar estratégias de recomendação; apenas não foram consideradas para o foco investigativo desta tese. Todavia, ainda foi realizada uma análise, com a métrica MSI, referente ao aspecto da novidade dos itens recomendados na estratégia de *re-ranking*.

A premissa expressa no início do parágrafo anterior também justifica a motivação de realizar a investigação considerando-se FC — e não, por exemplo, FBC (*personality matching*). Além disso, considerou-se FC almejando-se elaborar um método padronizado (genérico) para empregar a personalidade (dada sua importância na personalização) em recomendações para variados domínios (abrangendo aqueles com dificuldade em empregar FBC, por ausência de dados sobre conteúdo), inclusive domínios menos explorados, além daqueles que se tornem futuramente propícios para realizar recomendações.

Nesta tese, foi aludida a ideia que todos os algoritmos de FC implicitamente lidam com os traços de personalidade dos usuários e, ao dividi-los em grupos baseados explicitamente nesses traços de personalidade, conjecturou-se que o desempenho referente à acurácia dos algoritmos não seria prejudicado — e a aplicação desses agrupamentos poderia melhorar tal acurácia. Ao avaliar isso (QP2) experimentalmente, não se observou, em geral, melhoria nos algoritmos *model-based* (havendo casos de diminuição no desempenho), contudo, observou-se uma tendência de melhoria nos algoritmos *neighborhood-based*. Assim, indica-se que futuras pesquisas considerem a aplicação de agrupamentos prévios baseados em personalidade ao empregarem FC baseada em vizinhança (estratégia “tps”). Para os que empregarem FC baseada em modelo, indica-se analisar a aplicação de agrupamentos prévios baseados em “personalidade de itens” (conforme comentado ao avaliar a estratégia “itps” referente à QP4).

Neste documento, foram apresentadas investigações objetivando a elaboração de estratégias híbridas de SR, baseando-se em características de personalidade (empregando “personalidade de itens”, *weighted ensemble* e *re-ranking*). Os experimentos indicaram haver importância, para o processo de recomendação, de considerar características de personalidade (extraídas sem uso de questionário) para a criação do perfil dos usuários. Entretanto, para garantir maior relevância das recomendações, é preciso ir além dos métodos básicos baseados em personalidade.

Apesar de apresentar diferenças estatísticas, comentando-se termos como melhor/pior desempenho, ao longo do Capítulo 5, por vezes os ganhos e perdas não foram tão acentuados. Alguns estudos anteriores aplicando personalidade demonstraram ganhos mais expressivos avaliando cenários específicos (a exemplo de *cold-start*), mas, nesta tese, decidiu-se contribuir avaliando impactos da aplicação da personalidade em cenários mais gerais.

Nesse sentido, considerando a principal hipótese formulada nesta tese (Seção 1.1) e os resultados obtidos (Capítulo 5), é possível afirmar que usar dados explícitos de personalidade no processo de recomendação, inferidos via APR, não garante mudança substancial concernente à acurácia das categorias de FC. Por exemplo: se em determinado contexto, o SVD++ é mais acurado que o UserKNN, há pouca expectativa de que, ao empregar personalidade no UserKNN, este se torne mais acurado que o SVD++ sem empregar personalidade. Entretanto, aplicar a personalidade no processo de recomendação pode favorecer ganhos em outros aspectos (como incipientemente mostrado nos gráficos referentes à métrica MSI) concomitante a não ter perdas expressivas na acurácia geral das recomendações (podendo inclusive ter pequenos ganhos).

Recentemente, Moreno et al. (2021) concluíram que, apesar de a linguagem escrita analisada por métodos computacionais ser útil para identificar informações da personalidade, mais pesquisas ainda são necessárias para prever/explicar diferenças individuais, uma vez que os tamanhos de efeito da meta-análise realizada foram pequenos a moderados. Tais autores indicaram que se espera uma melhoria disso em um futuro próximo. Nesse sentido, os ganhos referentes à acurácia, nas estratégias apresentadas nesta tese, podem ser mais expressivos com essa futura melhoria em APR baseado em texto. Então, similar a como Moreno et al. (2021) concluíram seu estudo, conclui-se que as informações da investigação apresentada nesta tese são relevantes para o aprimoramento de propostas de pesquisas futuras.

É pertinente reforçar que alguns aspectos conclusivos, obtidos durante o desenvolvimento desta pesquisa, foram divulgados por meio de publicação de artigos em periódicos e conferências. Tais publicações, elencadas no Apêndice A, foram anteriormente citadas (Seção 1.3), relacionando-as às questões de pesquisa estabelecidas.

## **6.2 Limitações e Sugestões para Pesquisas Futuras**

Conforme comentado na Seção 1.3, a motivação mudou durante o desenvolvimento da pesquisa. Apesar dos variados materiais educacionais disponíveis, quando se especifica um assunto de uma determinada área, a quantidade de materiais filtrados não necessariamente permanece expressiva para motivar a proposição de um SR focado na adequação dos itens (após tal filtragem) às características psicológicas dos estudantes — tais como traços de

personalidade e estilos de aprendizagem. Por exemplo, no contexto brasileiro atual, acaba sendo utópico ter professores criando e disponibilizando diversos objetos de aprendizagem do mesmo conteúdo de uma aula do ensino fundamental, variando características referentes às preferências no modo de aprender dos estudantes. Isso impacta, por exemplo, em dificuldades referentes ao uso de dados não sintéticos para investigações empíricas. Nesse sentido, preferiu-se realizar a pesquisa menos dependente do domínio educacional<sup>1</sup>.

Sobre ameaças à validade, convém inicialmente comentar que os domínios analisados, específicos de determinados sites (TripAdvisor e Amazon), com uma determinada amostra da população de usuários desses sites, podem ter características próprias, dificultando a generalização dos resultados, o que configura uma ameaça à validade externa. Na Seção 5.1, ao apresentar as características dos conjuntos de dados utilizados, são comentados alguns aspectos limitadores (especialmente referentes aos dados oriundos do TripAdvisor). É pertinente, portanto, realizar esta investigação também com outras fontes de dados.

Nesse ponto, é pertinente comentar que, embora o foco investigativo desta tese tenha sido em relação a *personality neighborhood filtering* (e não a *personality matching*), é provável que haja ganhos mais expressivos ao incorporar a personalidade no processo de FC se, no domínio envolvido, houver itens com características mais fortemente relacionadas a características de personalidade. Como comentado na Seção 5.1, por exemplo, não é trivial estabelecer associações diretas entre o *Big Five* e itens no domínio alimentício.

Convém cogitar que informações sobre características da personalidade do usuário podem não impactar significativamente, em um processo de recomendação, em alguns domínios (e/ou cenários) específicos, por tais domínios/cenários contemplarem outros tipos de informação mais relevantes para o objetivo do SR. Aliado a isso, há as limitações próprias do processo de APR (por exemplo, um usuário, em determinados domínios/cenários, pode se expressar diferentemente do habitual, dificultando reconhecer as verdadeiras características de sua personalidade). Esses aspectos podem impactar em resultados não muito expressivos ou negativos quanto à qualidade das recomendações.

Ainda sobre os dados, outra ameaça à validade consiste na possibilidade de alguns *ratings*

---

<sup>1</sup>Destaca-se que não está sendo defendido, neste ponto, que propostas de SR educacionais não são importantes. Entretanto, alguns estudos, na temática discutida, acabarão sendo mais bem aproveitados quando se perceber a necessidade e houver diversos materiais produzidos considerando, por exemplo, os diferentes estilos de aprendizagem dos estudantes (AGUIAR, 2015).

não representarem de fato a opinião de determinado usuário, e alguns *reviews* podem conter palavras que não se associam realmente à personalidade de tal usuário, pois este pode ter disponibilizado sua conta (*login/senha*) para uso de outra pessoa (cônjuge, amigo, etc.).

A restrição inicial do número mínimo de *ratings* por item (pelo menos 5) limitou os dados das unidades experimentais (conjuntos de treinamento), mas, para minimizar essa ameaça à validade, também foram considerados os conjuntos de dados sem essa restrição (como comentado na Seção 5.1). Além disso, desconsiderar os usuários com menos de 600 palavras (para uso do *Watson Personality Insights*) foi uma restrição necessária que gera uma ameaça à validade.

Outra ameaça à validade seria ter considerado resultados dos testes estatísticos com suposições violadas (por exemplo: empregar testes paramétricos com dados não seguindo uma distribuição normal). Todavia, há testes estatísticos (como o teste T) razoavelmente robustos para dados sem uma distribuição normal. Além disso, também foram empregados testes não paramétricos para comparação dos resultados, obtendo-se geralmente resultados equivalentes. Houve apenas dois casos discordantes, para os quais não foi rejeitada a hipótese nula ( $p$ -valores do teste T menores que 0,05, mas maiores que 0,03), para evitar cometer o erro de considerar diferença estatística sem haver.

Considerando-se os objetivos desta tese, foram conduzidos especificamente experimentos off-line (utilizando o histórico de usuários reais), mas, posteriormente, é pertinente conduzir também experimentos on-line (AGGARWAL et al., 2016), com foco, por exemplo, na avaliação da estratégia *re-ranking*. Ainda em relação a aspectos experimentais, é pertinente estudar novos fatores e opções específicas dos algoritmos. Convém lembrar que, nesta pesquisa, se enfocou na FC. Estudos futuros podem, por exemplo, expandir a abordagem referente à questão de pesquisa QP2 para que, após o agrupamento (*clustering*), sejam empregados outros métodos, a exemplo de FBC.

Como comentado na Seção 4.4, planejava-se inicialmente propor uma estratégia de recomendação com GA, ponderando as características de personalidade, mas, devido aos resultados obtidos na investigação experimental referente à QP3, tal proposta não foi apresentada formalmente nesta tese. Todavia, sugere-se que a análise referente à QP3 seja realizada em variados domínios — e repetidamente em diferentes conjuntos de dados de um mesmo domínio — com propósito de descobrir se, para domínios específicos, há alguma

ponderação específica relevante. Assim, a estratégia com GA (abrangendo a diversidade de características de personalidade a considerar) poderia possibilitar, em alguns domínios, ter um grupo diferente de traços impactando mais fortemente na relevância (e/ou outros aspectos) das recomendações. Por exemplo: com itens musicais, determinados traços podem ser mais influentes do que com itens alimentícios.

Como visto na revisão bibliográfica, e destacado recentemente por Lex et al. (2021), modelar a “personalidade de itens” ainda é uma questão pouco pesquisada, sendo necessários métodos aprimorados para derivar traços de personalidade no nível do item. Nesta tese, também visou-se a contribuir com esta questão ao propor a estratégia “itps”. Embora tal estratégia não tenha sido significativamente vantajosa nas métricas avaliadas (foco na acurácia), é pertinente avaliá-la, em futuros estudos, em relação a outros aspectos, como explicabilidade (*explainability*). Além disso, identificar a “personalidade de itens” pode ser útil em outras abordagens além da FC. Por exemplo: um livro de autoajuda com alto valor calculado referente ao “perfil de Neuroticismo do item” talvez seja útil para ser sugerido na compra de outros livros também com valores altos no “perfil de Neuroticismo”, fortalecendo a FBC (que já se preocupa com a categoria do livro).

No tocante à explicabilidade, as estratégias elaboradas nesta tese podem contribuir com tal aspecto, ajudando no entendimento de uma recomendação. Por exemplo, no domínio educacional, explicar a um aluno/professor/gestor/pesquisador o motivo de uma recomendação fornecida a ele, gerada essencialmente via fatoração de matriz, provavelmente será menos fácil que se comentar que a recomendação foi definida baseando-se na opinião de pessoas com personalidade similar a dele.

Nesse contexto, Tran, Felfernig e Tintarev (2021) defendem ser essencial explorar características de personalidade para fins de explicação. Kouki et al. (2020) e Millicamp et al. (2020) usaram traços de personalidade para a explicação das recomendações, mas não para o processo de recomendação em si. Recentemente, Najafian et al. (2021), ao considerar a preocupação com a privacidade, indicaram que as explicações apresentadas a grupos de usuários precisam ser adaptadas à personalidade deles (em especial, referente aos traços Extroversão e Amabilidade<sup>2</sup>). É pertinente expandir as investigações nesse âmbito.

---

<sup>2</sup>Convém lembrar que, na análise experimental referente ao *re-ranking* (Subseção 5.2.4), o traço Amabilidade também se destacou em relação a ser potencialmente útil tratar diferentemente os grupos de usuário em relação a tal traço.

Embasados em uma revisão realizada, Lex et al. (2021) afirmam que, até o momento, não é bem compreendido até que ponto a personalidade influencia a qualidade da recomendação (nem a variabilidade disso entre os usuários). Segundo tais autores, pode ser valioso adequar recomendações aos traços de personalidade para alguns usuários e domínios, mas isso pode ser um sinal irrelevante para outros, podendo inclusive ser percebido por alguns usuários como algo invasivo em relação à privacidade.

Nesse sentido, Lex et al. (2021) destacam que incorporar a personalidade de uma forma consciente da privacidade é uma questão em aberto. Dhelim et al. (2021) comentam a privacidade dos dados de personalidade dos usuários, para evitar ações de indivíduos mal-intencionados, como um dos novos desafios (também envolvendo questões éticas) trazidos pelo uso da personalidade do usuário na recomendação. Nesse contexto, é conveniente citar o escândalo, destaque em 2018, envolvendo a empresa Cambridge Analytica: um aplicativo do Facebook criado pela Cambridge Analytica coletou dados de personalidade de mais de 87 milhões de usuários, com o objetivo de manipular a escolha de voto na eleição presidencial dos Estados Unidos em 2016 (BBC, 2018) (PRESSE, 2019).

Sobre questões éticas, além de aspectos sobre privacidade, há outras preocupações concernentes ao estudo de sistemas baseados em traços de personalidade. Um exemplo de preocupação relevante é o uso de FC envolvendo o traço neuroticismo. Como apresentado no Quadro 2.1, pessoas com ansiedade excessiva (além de outros aspectos não tão positivos referentes à saúde mental) tendem a serem representadas com índices altos de neuroticismo e podem, por exemplo, se interessar por conteúdos nocivos à sua saúde (aumentando sintomas de ansiedade, por exemplo). Nesse contexto, dependendo da construção do SR, é provável que usuários com índices altos de neuroticismo recebam recomendações desses conteúdos nocivos (“tóxicos”). Portanto, ao incorporar estratégias de recomendação baseadas na personalidade humana, é preciso refletir, do ponto de vista ético, sobre situações não pertinentes para considerar no processo de personalização.

Considerando-se a variedade de instrumentos existentes para identificar traços de personalidade, Lex et al. (2021) comentaram recentemente que a questão de quando usar cada instrumento desse — e que qualidade pode ser alcançada — ainda é objeto de investigação mais detalhada. Isso reforça a importância de pesquisas como a apresentada nesta tese, investigando impactos em SR com enfoque em uma forma de reconhecimento automático (sem

questionários).

Infortunadamente, nos últimos meses de realização desta pesquisa, foi visualizado na página Web da IBM um comunicado sobre o encerramento do serviço *Watson Personality Insights* no fim de 2021. Isso impacta na parte da motivação desse estudo referente ao uso de tal ferramenta — por parte, por exemplo, de SR comerciais já existentes — como meio facilitador para inserir características de personalidade em seus algoritmos de recomendação. Todavia, isso não invalida este estudo, pois, por exemplo, os resultados da QP1, referentes a não ter mudanças significativas ao inserir dados além do *Big Five*, acaba motivando o uso de ferramentas automáticas existentes que só identificam os cinco traços do *Big Five* — como a de Mairesse et al. (2007).

Além disso, embora sistemas futuros não possam usar o *Watson Personality Insights*, isso sugere a importância de a academia continuar avançando no desenvolvimento de ferramentas para identificar a personalidade humana, despertando-se, por exemplo, para incluir outros modelos de personalidade (mais focados em marketing, por exemplo, como o modelo *Needs*).

Ademais, com o término do serviço da IBM, a disponibilização do conjunto de dados utilizado neste estudo consiste em uma contribuição para a realização de estudos futuros em Computação da Personalidade (*Personality Computing*). Pretende-se publicar o conjunto de dados na próxima edição do SBBD DSW (*Dataset Showcase Workshop*), um workshop do Simpósio Brasileiro de Bancos de Dados (SBBD).

Estudos futuros sobre SR baseados em personalidade, especialmente no contexto brasileiro, serão mais produtivos quando as pesquisas avançarem mais no sentido de identificar a personalidade via textos brasileiros. É apropriado expor que Ferreira et al. (2020) exploraram o uso de processamento de linguagem natural para APR (modelo *Big Five*) a partir de textos em língua portuguesa (do Brasil), e ressaltam que a maioria dos estudos referentes a esse aspecto é para a língua inglesa. O *Watson Personality Insights*, por exemplo, não analisa textos em língua portuguesa.

Destaca-se, portanto, que as pesquisas em SR envolvendo traços de personalidade, em âmbito brasileiro, podem se beneficiar ainda mais com os avanços em APR, motivando ainda mais pesquisadores para essa área. Segundo Jaques e Nunes (2021), “é interessante focar em prover novas formas não intrusivas de inferência da personalidade em português brasileiro,

variando o espectro de teorias da personalidade usadas”. É pertinente, portanto, como sugestão de estudo futuro, focar nessa inferência no âmbito brasileiro, expandindo pesquisas existentes focadas no *Big Five* para, por exemplo, abranger modelos como *Values* e *Needs*.

Apesar de a necessidade de ter textos disponíveis dos usuários ser uma limitação para empregar APR em SR, é possível considerar os textos dos *reviews* em determinado domínio e, uma vez calculados os valores representativos da personalidade, usá-los em outros domínios (*cross-domain recommendation*). Sobre isso, Tran, Felfernig e Tintarev (2021) defendem que, embora existam algumas abordagens na literatura, mais pesquisas relacionadas são necessárias. Considerando-se tal possibilidade de reuso, é pertinente destacar que existe o PersonalityML (NUNES; BEZERRA; OLIVEIRA, 2012), uma linguagem de marcação, fundamentada em XML (*eXtensible Markup Language*), para padronizar a personalidade do usuário em SR.

Ainda sobre a limitação citada no início do parágrafo anterior, convém comentar que, embora um usuário possa inicialmente não ter informações textuais suficientes, o SR pode começar com um questionário breve (como o TIPI e o BFI-10) e depois atualizar os dados de personalidade, ou, a depender do domínio, iniciar com estratégias como a apresentada por Lima, Feijó e Furtado (2018) — o *Big Five Game Inventory* —, no universo de jogos.

Em relação aos resultados da análise complementar referente à estratégia de *re-ranking* proposta (avaliando os resultados referentes às métricas por grupos de usuário), pretende-se expandir tal análise em estudos futuros, com outros conjuntos de dados e domínios, podendo ser útil um recomendador híbrido alternado (*switching*) para que determinadas estratégias possam ser focadas para usuários com determinadas características de personalidade.

Outra sugestão de estudo futuro consiste em analisar como cada traço de personalidade influencia o *re-ranking* proposto, inclusive para métricas além da acurácia (como a MSI, empregada na análise experimental). As características de personalidade dos usuários poderiam ser úteis, por exemplo, para expandir a tese de Oliveira (2020) sobre diversificação em SR. Nesse ponto, convém destacar os seguintes aspectos comentados por Oliveira (2020):

Apesar da notável necessidade de considerar diversidade e novidade como métricas de utilidade, é arriscado um sistema de recomendação otimizar essas dimensões esquecendo a relevância, uma vez que o sistema pode não mais atender diretamente às preferências do usuário (RIBEIRO et al., 2012 [apud (OLIVEIRA, 2020)]). [...] Ao oferecer recomendações diversas, tem-se a ideia de ampliar a gama de opções

exibidas ao usuário, aumentando assim a chance de que alguma daquelas possibilidades adéque-se ao interesse momentâneo do usuário. Com a métrica novidade, tem-se a intenção de oferecer itens desconhecidos, porém próximos às preferências do usuário, enquanto serendipidade diz respeito à sensação de descoberta de um item valioso. Em todos os casos, as recomendações realizadas devem ser de interesse do usuário, ou seja, o sistema de recomendação deve buscar essas métricas com o menor prejuízo possível à métrica de precisão do sistema. (OLIVEIRA, 2020, p. 13 e 31).

Baseando-se em estudos recentes sobre recomendação baseada em personalidade focando em diversidade, julga-se pertinente, portanto, que estratégias de recomendação usem explicitamente os conceitos da personalidade dos indivíduos, mesmo com ganhos pequenos em acurácia. Em outras palavras, algumas estratégias podem não ter ganhos tão expressivos na acurácia das recomendações, mas podem ajudar na melhoria de outros aspectos (explicabilidade, diversidade, novidade e serendipidade).

É conveniente comentar que, além da personalidade, o tratamento de emoções é outro aspecto psicológico possível de ser considerado em SR (LEX et al., 2021). Entretanto, as emoções são mais inconstantes, sendo necessário identificá-las mais frequentemente. Perguntar diretamente ao usuário como ele está se sentindo, repetidamente, pode tornar o processo mais cansativo/entediante que usar os questionários de personalidade. Por outro lado, identificar automaticamente emoções com monitoramento constante de expressões faciais via *webcam*, por exemplo, não é um meio apazível para muitos usuários. No contexto educacional, isso pode ser inaceitável por determinados estudantes e/ou seus responsáveis, por possivelmente impactar negativamente no processo de aprendizagem. Assim, sugere-se, também, realizar estudos nessa temática, para aliar a aplicação da personalidade às emoções dos usuários em SR.

Sobre aspectos psicológicos além do *Big Five*, é pertinente comentar que, recentemente, Huang et al. (2020) indicaram que estilos de vida (*lifestyles*) on-line refletem e afetam as preferências dos consumidores de comércio eletrônico em vários cenários. Seus resultados experimentais mostraram os estilos de vida on-line e suas subdimensões melhorando o desempenho de predição de preferências e superando o uso do *Big Five*. Nesse sentido, é pertinente analisar a forma de identificação e a estratégia adotada no tocante à aplicação dos traços de personalidade desses consumidores para a realização da predição, além de investigar cenários além de comércio eletrônico.

Apesar dos SR anteciparem qual a necessidade/preferência dos usuários, estes podem, em alguns domínios, sentir-se atraídos pela ideia de controle da situação (como a experiência em “Black Mirror: Bandersnatch”, na Netflix<sup>3</sup>). Além de apresentar o resultado da recomendação “tradicional” do sistema, poderia também apresentar, por exemplo, outras poucas opções referentes a características associadas a determinados traços de personalidade — com a constante escolha em um tipo, tais traços seriam marcados com maior percentual em relação a outros, em um perfil dinâmico do usuário (ponderando um algoritmo baseado em traços de personalidade). Ademais, é incitante pensar também nesta opção para ser apresentada nos SR: “o que me seria recomendado se minha personalidade fosse diferente?” ou “o que me seria recomendado se minha personalidade fosse similar à personalidade da celebridade X?”.

Por fim, compreende-se que esta tese contribui para a área de SR ao apresentar uma investigação envolvendo a aplicação de teorias concernentes à personalidade, em algoritmos usuais para recomendação de recursos, visando a melhorar soluções referentes à personalização de ambientes computacionais. Reforça-se também que, além de considerar características de personalidade e FC em SR, é importante que o processo de recomendação englobe outras técnicas e aspectos, para atender particularidades de determinados domínios (como, por exemplo, analisar o conteúdo dos itens ao recomendar objetos de aprendizagem — domínio educacional).

---

<sup>3</sup>Disponível em: <https://www.netflix.com/br/title/80988062>. Acesso em: 30 de out. de 2021

# Referências Bibliográficas

ADAJI, I.; SHARMAINE, C.; DEBROWNEY, S.; OYIBO, K.; VASSILEVA, J. Personality Based Recipe Recommendation Using Recipe Network Graphs. In: MEISELWITZ, G. (Ed.). *Social Computing and Social Media. Technologies and Analytics*. Cham: Springer International Publishing, 2018. v. 10914, p. 161–170. ISBN 978-3-319-91484-8 978-3-319-91485-5. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-91485-5\\_12](https://doi.org/10.1007/978-3-319-91485-5_12). Acesso em: 29 de out. de 2021.

ADAMOPOULOS, P.; TODRI, V. Personality-Based Recommendations: Evidence from Amazon.com. In: CASTELLS, P. (Ed.). *Poster Proceedings of the 9th ACM Conference on Recommender Systems (RecSys)*. Vienna, Austria: CEUR-WS.org, 2015. (CEUR Workshop Proceedings, v. 1441). Disponível em: [http://ceur-ws.org/Vol-1441/recsys2015\\_poster16.pdf](http://ceur-ws.org/Vol-1441/recsys2015_poster16.pdf). Acesso em: 29 de out. de 2021.

ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, v. 17, n. 6, p. 734–749, jun. 2005. ISSN 1041-4347. Disponível em: <https://doi.org/10.1109/TKDE.2005.99>. Acesso em: 29 de out. de 2021.

AGGARWAL, C. C. et al. *Recommender Systems — The Textbook*. New York: Springer International Publishing, 2016. ISBN 978-3-319-29657-9. Disponível em: <https://doi.org/10.1007/978-3-319-29659-3>. Acesso em: 29 de out. de 2021.

AGUIAR, J.; BARBOSA, A.; FECHINE, J.; COSTA, E. Um Estudo sobre a Influência das Dimensões do Modelo Felder-Silverman na Recomendação de Recursos Educacionais baseada nos Estilos de Aprendizagem dos Alunos. In: *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*. Recife-PE: SBC (Sociedade Brasileira de Computação), 2017. v. 28, p. 1277–1286. Disponível em: <https://doi.org/10.5753/cbie.sbie.2017.1277>. Acesso em: 29 de out. de 2021.

AGUIAR, J.; FECHINE, J.; COSTA, E. Utilização da ferramenta Five Labs para Identificação de Traços de Personalidade dos Estudantes. In: *Anais do XXI Workshop de Informática na Escola (WIE 2015)*. Maceió, Alagoas, Brasil: SBC (Sociedade Brasileira de Computação), 2015. v. 21, p. 157–166. ISSN 2316-6541. Disponível em: <https://doi.org/10.5753/cbie.wie.2015.157>. Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B. *Recomendação de objetos de aprendizagem baseada em estilos de aprendizagem e traços de personalidade*. Dissertação (Mestrado) — Universidade Federal de Campina Grande (UFCG) — Programa de Pós-Graduação em Ciência da Computação, Campina Grande, Paraíba, Brasil, 2015. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/546>. Acesso em: 27 de dez. de 2021.

AGUIAR, J. J. B. Considerando Estilos de Aprendizagem, Emoções e Personalidade em Informática na Educação. *Informática na Educação: Teoria & Prática*, v. 20, n. 2 mai/ago, p. 85–102, set. 2017. ISSN 1982-1654, 1516-084X. Disponível em: <https://doi.org/10.22456/1982-1654.65333>. Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Estratégia baseada em Personalidade e Tendências para Recomendação de Objetos de Aprendizagem usando Algoritmo Genético. In: *I Workshop Latino-Americano de Trabalhos em Andamento em Computação (WLATAC) - Anais*. 1. ed. São Paulo–SP: SBC (Sociedade Brasileira de Computação), 2018. v. 1, p. 62–67. ISBN 978-85-7669-460-1. Disponível em: [http://portaldeconteudo.sbc.org.br/index.php/wlatac\\_clei/issue/view/282/wlatac\\_clei](http://portaldeconteudo.sbc.org.br/index.php/wlatac_clei/issue/view/282/wlatac_clei). Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Recomendação de Objetos de Aprendizagem utilizando Filtragem Colaborativa baseada em Tendências e em Estilos de Aprendizagem. In: *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*. Fortaleza–CE: SBC (Sociedade Brasileira de Computação), 2018. v. 29, p. 1423–1432. Disponível em: <https://doi.org/10.5753/cbie.sbie.2018.1423>. Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Experimentando a Influência dos Traços de Personalidade do Modelo Big Five na Recomendação de Recursos Educacionais. In: *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)*. Brasília–DF: SBC (Sociedade Brasileira de Computação), 2019. v. 30, p. 1711–1720. Disponível em: <https://doi.org/10.5753/cbie.sbie.2019.1711>. Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Collaborative Filtering Strategy for Product Recommendation Using Personality Characteristics of Customers. In: *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia 2020)*. São Luís–MA (Brazil): Association for Computing Machinery (ACM), 2020. p. 157–164. ISBN 978-1-4503-8196-3. Disponível em: <https://doi.org/10.1145/3428658.3430969>. Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Estudo Comparativo de Abordagens para Sistemas de Recomendação baseados em Personalidade com uso do serviço IBM Watson Personality Insights. *Revista Ibérica de Sistemas e Tecnologias de Informação (RISTI)*, Porto (Portugal), v. 40, p. 73–88, 2020. ISSN 1646-9895. Disponível em: <https://doi.org/10.17013/risti.40.73-88>. Acesso em: 29 de out. de 2021.

AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Estudo de Desempenho de Algoritmos de Filtragem Colaborativa para Sistemas de Recomendação Educacionais aplicando Agrupamento Prévio de Usuários com Traços de Personalidade Similares. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)*. Natal–RN: SBC (Sociedade Brasileira de Computação), 2020. v. 31, p. 1082–1091. Disponível em: <https://doi.org/10.5753/cbie.sbie.2020.1082>. Acesso em: 29 de out. de 2021.

ALHARTHI, H. *The Use of Items Personality Profiles in Recommender Systems*. Tese (Doutorado) — Université d’Ottawa/University of Ottawa, Ottawa, Canada, 2015.

Disponível em: [https://ruor.uottawa.ca/bitstream/10393/31922/1/Alharthi\\_Haifa\\_2015\\_the\\_sis.pdf](https://ruor.uottawa.ca/bitstream/10393/31922/1/Alharthi_Haifa_2015_the_sis.pdf). Acesso em: 29 de out. de 2021.

ALHARTHI, H.; TRAN, T. Item-based collaborative filtering using the big five personality traits. In: *Proceedings of the 1st Workshop on Recommendation Systems for Television and Online Video (RecSysTV) co-located with 8th ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2014.

ALLPORT, G. W. *Personality: A psychological interpretation*. Henry Holt, Oxford, England, p. xiv, 588, 1937.

ALVES, P.; SARAIVA, P.; CARNEIRO, J.; CAMPOS, P.; MARTINS, H.; NOVAIS, P.; MARREIROS, G. Modeling Tourists' Personality in Recommender Systems: How Does Personality Influence Preferences for Tourist Attractions? In: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. Genoa Italy: Association for Computing Machinery (ACM), 2020. p. 4–13. ISBN 978-1-4503-6861-2. Disponível em: <https://doi.org/10.1145/3340631.3394843>. Acesso em: 29 de out. de 2021.

ALVES, T.; NATÁLIO, J.; HENRIQUES-CALADO, J.; GAMA, S. Incorporating personality in user interface design: A review. *Personality and Individual Differences*, Elsevier, v. 155, p. 109709, 2020. ISSN 0191-8869. Disponível em: <https://doi.org/10.1016/j.paid.2019.109709>. Acesso em: 29 de out. de 2021.

ARNOUX, P.-H.; XU, A.; BOYETTE, N.; MAHMUD, J.; AKKIRAJU, R.; SINHA, V. 25 tweets to know you: A new model to predict personality with social media. In: . [s.n.], 2017. v. 11, n. 1, p. 472–475. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14963>. Acesso em: 29 de out. de 2021.

ASABERE, N. Y.; ACAKPOVI, A. *ROPPSA : TV Program Recommendation Based on Personality and Social Awareness. Mathematical Problems in Engineering*, v. 2020, p. 1–15, jun. 2020. ISSN 1024-123X, 1563-5147. Disponível em: <https://doi.org/10.1155/2020/1971286>. Acesso em: 29 de out. de 2021.

ASABERE, N. Y.; ACAKPOVI, A.; MICHAEL, M. B. Improving Socially-Aware Recommendation Accuracy Through Personality. *IEEE Transactions on Affective Computing*, v. 9, n. 3, p. 351–361, jul. 2018. ISSN 1949-3045, 2371-9850. Disponível em: <https://doi.org/10.1109/TAFFC.2017.2695605>. Acesso em: 29 de out. de 2021.

ATAS, M.; FELFERNIG, A.; POLAT-ERDENIZ, S.; POPESCU, A.; TRAN, T. N. T.; UTA, M. Towards psychology-aware preference construction in recommender systems: Overview and research issues. *Journal of Intelligent Information Systems*, set. 2021. ISSN 0925-9902, 1573-7675. Disponível em: <https://doi.org/10.1007/s10844-021-00674-5>. Acesso em: 29 de out. de 2021.

BACHRACH, Y.; KOSINSKI, M.; GRAEPEL, T.; KOHLI, P.; STILLWELL, D. Personality and patterns of facebook usage. In: *Proceedings of the 4th Annual ACM Web Science Conference (WebSci)*. New York, NY, USA: Association for Computing Machinery (ACM), 2012. p. 24–32. ISBN 9781450312288. Disponível em: <https://doi.org/10.1145/2380718.2380722>. Acesso em: 29 de out. de 2021.

BADENES, H.; BENGUALID, M. N.; CHEN, J.; GOU, L.; HABER, E.; MAHMUD, J.; NICHOLS, J. W.; PAL, A.; SCHOUDT, J.; SMITH, B. A.; XUAN, Y.; YANG, H.; ZHOU, M. X. System u: Automatically deriving personality traits from social media for people recommendation. In: *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2014. p. 373–374. ISBN 9781450326681. Disponível em: <https://doi.org/10.1145/2645710.2645719>. Acesso em: 29 de out. de 2021.

BALAKRISHNAN, V.; ARABI, H. HyPeRM: A hybrid personality-aware recommender for movie. *Malaysian Journal of Computer Science*, v. 31, n. 1, p. 48–62, jan. 2018. ISSN 0127-9084. Disponível em: <https://doi.org/10.22452/mjcs.vol31no1.4>. Acesso em: 29 de out. de 2021.

BARVINSKI, C.; RIBEIRO, A. C. R.; LONGHI, M.; BEHAR, P. A. Proposta de Modelo Socioafetivo de Aluno para a Recomendação de Estratégias Pedagógicas. In: . Recife, Pernambuco, Brasil: [s.n.], 2017. p. 1637. Disponível em: <https://doi.org/10.5753/cbie.sbie.2017.1637>. Acesso em: 29 de out. de 2021.

BBC. *Entenda o escândalo de uso político de dados que derrubou valor do Facebook e o colocou na mira de autoridades*. 2018. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/entenda-o-escandalo-de-uso-politico-de-dados-que-derrubou-valor-do-facebook-e-o-colocou-na-mira-de-autoridades.ghtml>. Acesso em: 29 de out. de 2021.

BEHESHTI, A.; YAKHCHI, S.; MOUSAEIRAD, S.; GHAFARI, S. M.; GOLUGURI, S. R.; EDRISI, M. A. Towards Cognitive Recommender Systems. *Algorithms*, v. 13, n. 8, p. 176, jul. 2020. ISSN 1999-4893. Disponível em: <https://doi.org/10.3390/a13080176>. Acesso em: 29 de out. de 2021.

BIAN, L.; HOLTZMAN, H.; HUYNH, T.; MONTPETIT, M.-J. MatchMaker: A friend recommendation system through TV character matching. In: *2012 IEEE Consumer Communications and Networking Conference (CCNC)*. Las Vegas, NV, USA: IEEE, 2012. p. 714–718. ISBN 978-1-4577-2071-0 978-1-4577-2070-3 978-1-4577-2069-7. Disponível em: <https://doi.org/10.1109/CCNC.2012.6180983>. Acesso em: 29 de out. de 2021.

BOBADILLA, J.; ORTEGA, F.; HERNANDO, A.; GUTIÉRREZ, A. Recommender systems survey. *Knowledge-Based Systems*, Elsevier, v. 46, p. 109–132, 2013. ISSN 0950-7051. Disponível em: <https://doi.org/10.1016/j.knosys.2013.03.012>. Acesso em: 29 de out. de 2021.

BOLOCK, A. E.; KADY, A. E.; HERBERT, C.; ABDENNADHER, S. Towards a Character-based Meta Recommender for Movies. In: ALFRED, R.; LIM, Y.; HAVILUDDIN, H.; ON, C. K. (Ed.). *Computational Science and Technology*. Singapore: Springer Singapore, 2020. v. 603, p. 627–638. ISBN 9789811500572 9789811500589. Series Title: Lecture Notes in Electrical Engineering. Disponível em: [https://doi.org/10.1007/978-981-15-0058-9\\_60](https://doi.org/10.1007/978-981-15-0058-9_60). Acesso em: 29 de out. de 2021.

BOLOGNA, C.; ROSA, A. C. D.; GAETA, M.; SANSONETTI, G.; VISERTA, V. Personality-Based Recommendation in E-Commerce. In: BERKOVSKY, S.; HERDER, E.; LOPS, P.; SANTOS, O. C. (Ed.). *Late-Breaking Results, Project Papers and Workshop*

*Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization (UMAP Workshops — EMPIRE 2013: Emotions and Personality in Personalized Services)*. Rome, Italy: CEUR-WS.org, 2013. (CEUR Workshop Proceedings, v. 997), p. 7–12. ISSN 1613-0073. Disponível em: [http://ceur-ws.org/Vol-997/empire2013\\_paper\\_1.pdf](http://ceur-ws.org/Vol-997/empire2013_paper_1.pdf). Acesso em: 29 de out. de 2021.

BOSLAUGH, S. *Statistics in a Nutshell*. 2nd ed. ed. Farnham, Surrey, England: O'Reilly, 2012. (In a Nutshell). OCLC: ocn820127295. ISBN 978-1-4493-1682-2.

BRAUNHOFER, M.; ELAHI, M.; GE, M.; RICCI, F. Context Dependent Preference Acquisition with Personality-Based Active Learning in Mobile Recommender Systems. In: HUTCHISON, D.; KANADE, T.; KITTLER, J.; KLEINBERG, J. M.; KOBSA, A.; MATTERN, F.; MITCHELL, J. C.; NAOR, M.; NIERSTRASZ, O.; RANGAN, C. P.; STEFFEN, B.; TERZOPOULOS, D.; TYGAR, D.; WEIKUM, G.; ZAPHIRIS, P.; IOANNOU, A. (Ed.). *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration*. Cham: Springer International Publishing, 2014. v. 8524, p. 105–116. ISBN 978-3-319-07484-9 978-3-319-07485-6. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-07485-6\\_11](https://doi.org/10.1007/978-3-319-07485-6_11). Acesso em: 29 de out. de 2021.

BRAUNHOFER, M.; ELAHI, M.; RICCI, F. Usability Assessment of a Context-Aware and Personality-Based Mobile Recommender System. In: HEPP, M.; HOFFNER, Y. (Ed.). *E-Commerce and Web Technologies*. Cham: Springer International Publishing, 2014. v. 188, p. 77–88. ISBN 978-3-319-10490-4 978-3-319-10491-1. Series Title: Lecture Notes in Business Information Processing. Disponível em: [https://doi.org/10.1007/978-3-319-10491-1\\_9](https://doi.org/10.1007/978-3-319-10491-1_9). Acesso em: 29 de out. de 2021.

BRAUNHOFER, M.; ELAHI, M.; RICCI, F. User Personality and the New User Problem in a Context-Aware Point of Interest Recommender System. In: TUSSYADIAH, I.; INVERSINI, A. (Ed.). *Information and Communication Technologies in Tourism 2015*. Cham: Springer International Publishing, 2015. p. 537–549. ISBN 978-3-319-14342-2 978-3-319-14343-9. Disponível em: [https://doi.org/10.1007/978-3-319-14343-9\\_39](https://doi.org/10.1007/978-3-319-14343-9_39). Acesso em: 29 de out. de 2021.

BUETTNER, R. Predicting user behavior in electronic markets based on personality-mining in large online social networks: A personality-based product recommender framework. *Electronic Markets*, v. 27, n. 3, p. 247–265, ago. 2017. ISSN 1019-6781, 1422-8890. Disponível em: <https://doi.org/10.1007/s12525-016-0228-z>. Acesso em: 29 de out. de 2021.

BURGER, J. M. *Personality*. 8. ed. Belmont, CA, USA: Wadsworth Publishing, 2010. ISBN 978-0495813965.

BURKE, R. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction (UMUAI)*, Springer, v. 12, n. 4, p. 331–370, 2002. ISSN 0924-1868. Disponível em: <https://doi.org/10.1023/A:1021240730564>. Acesso em: 29 de out. de 2021.

CANTADOR, I.; FERNÁNDEZ-TOBÍAS, I. On the Exploitation of User Personality in Recommender Systems. In: GE, M.; RICCI, F. (Ed.). *Proceedings of the First International Workshop on Decision Making and Recommender Systems (DMRS2014)*. Bolzano, Italy: CEUR-WS.org, 2014. (CEUR Workshop Proceedings, v. 1278), p. 42–45. Disponível em: <http://ceur-ws.org/Vol-1278/paper8.pdf>. Acesso em: 29 de out. de 2021.

CANTADOR, I.; FERNÁNDEZ-TOBÍAS, I.; BELLOGÍN, A. Relating personality types with user preferences in multiple entertainment domains. In: BERKOVSKY, S.; HERDER, E.; LOPS, P.; SANTOS, O. C. (Ed.). *Late-Breaking Results, Project Papers and Workshop Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization (UMAP Workshops — EMPIRE 2013: Emotions and Personality in Personalized Services)*. Rome, Italy: CEUR-WS.org, 2013. (CEUR Workshop Proceedings, v. 997), p. 13–28. ISSN 1613-0073. Disponível em: [http://ceur-ws.org/Vol-997/empire2013\\_paper\\_2.pdf](http://ceur-ws.org/Vol-997/empire2013_paper_2.pdf). Acesso em: 29 de out. de 2021.

CARVALHO, L. d. F.; PIANOWSKI, G.; GONÇALVES, A. P. Personality differences and covid-19: are extroversion and conscientiousness personality traits associated with engagement with containment measures? *Trends in Psychiatry and Psychotherapy*, SciELO, v. 42, n. 2, p. 179–184, 2020. ISSN 2237-6089. Disponível em: <https://doi.org/10.1590/2237-6089-2020-0029>. Acesso em: 29 de out. de 2021.

CHAKRABARTY, N.; CHOWDHURY, S.; KANNI, S. D.; MUKHERJEE, S. FAFinder: Friend Suggestion System for Social Networking. In: HEMANTH, D. J.; SHAKYA, S.; BAIG, Z. (Ed.). *Intelligent Data Communication Technologies and Internet of Things*. Cham: Springer International Publishing, 2020. v. 38, p. 51–58. ISBN 978-3-030-34079-7 978-3-030-34080-3. Series Title: Lecture Notes on Data Engineering and Communications Technologies. Disponível em: [https://doi.org/10.1007/978-3-030-34080-3\\_6](https://doi.org/10.1007/978-3-030-34080-3_6). Acesso em: 29 de out. de 2021.

CHAN, G.; ARYA, A.; WHITEHEAD, A. Keeping Players Engaged in Exergames: A Personality Matchmaking Approach. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: Association for Computing Machinery (ACM), 2018. p. 1–6. ISBN 978-1-4503-5621-3. Disponível em: <https://doi.org/10.1145/3170427.3188455>. Acesso em: 29 de out. de 2021.

CHEN, L.; WU, W.; HE, L. How personality influences users' needs for recommendation diversity? In: *Extended Abstracts on Human Factors in Computing Systems (CHI '13)*. New York, NY, USA: Association for Computing Machinery (ACM), 2013. p. 829–834. ISBN 9781450319522. Disponível em: <https://doi.org/10.1145/2468356.2468505>. Acesso em: 27 de dez. de 2021.

CHEN, L.; WU, W.; HE, L. Personality and recommendation diversity. In: TKALČIČ, M.; CAROLIS, B. D.; GEMMIS, M. de; ODIĆ, A.; KOŠIR, A. (Ed.). *Emotions and Personality in Personalized Services: Models, Evaluation and Applications*. Cham: Springer International Publishing, 2016, (Human–Computer Interaction Series). p. 201–225. ISBN 978-3-319-31413-6. Disponível em: [https://doi.org/10.1007/978-3-319-31413-6\\_11](https://doi.org/10.1007/978-3-319-31413-6_11). Acesso em: 29 de out. de 2021.

- CHENG, R.; TANG, B. A Music Recommendation System Based on Acoustic Features and User Personalities. In: CAO, H.; LI, J.; WANG, R. (Ed.). *Trends and Applications in Knowledge Discovery and Data Mining*. Cham: Springer International Publishing, 2016. v. 9794, p. 203–213. ISBN 978-3-319-42995-3 978-3-319-42996-0. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-42996-0\\_17](https://doi.org/10.1007/978-3-319-42996-0_17). Acesso em: 29 de out. de 2021.
- COSTA, E.; AGUIAR, J.; MAGALHÃES, J. Sistemas de recomendação de recursos educacionais: conceitos, técnicas e aplicações. In: MELO, A. M.; BORGES, M. A. F.; SILVA, C. G. da (Ed.). *Anais da II Jornada de Atualização em Informática na Educação (JAIE)*. Campinas–SP: SBC (Sociedade Brasileira de Computação), 2013. p. 57–78. ISBN 9788576692836. [Capítulo base de minicurso apresentado na II JAIE realizada no II Congresso Brasileiro de Informática na Educação (CBIE)]. Disponível em: <http://www.br-ie.org/pub/index.php/pie/article/view/2589>. Acesso em: 29 de out. de 2021.
- COSTA, P. T.; MCCRAE, R. R. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources, 1992.
- DACREMA, M. F.; CREMONESI, P.; JANNACH, D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In: *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2019. p. 101–109. ISBN 9781450362436. Disponível em: <https://doi.org/10.1145/3298689.3347058>. Acesso em: 29 de out. de 2021.
- DHELIM, S.; AUNG, N.; BOURAS, M. A.; NING, H.; CAMBRIA, E. A survey on personality-aware recommendation systems. *Artificial Intelligence Review*, set. 2021. ISSN 0269-2821, 1573-7462. Disponível em: <https://doi.org/10.1007/s10462-021-10063-7>. Acesso em: 29 de out. de 2021.
- DHELIM, S.; AUNG, N.; NING, H. Mining user interest based on personality-aware hybrid filtering in social networks. *Knowledge-Based Systems*, v. 206, p. 106227, out. 2020. ISSN 0950-7051. Disponível em: <https://doi.org/10.1016/j.knosys.2020.106227>. Acesso em: 29 de out. de 2021.
- DHELIM, S.; NING, H.; AUNG, N.; HUANG, R.; MA, J. Personality-Aware Product Recommendation System Based on User Interests Mining and Metapath Discovery. *IEEE Transactions on Computational Social Systems*, v. 8, n. 1, p. 86–98, fev. 2021. ISSN 2329-924X, 2373-7476. Disponível em: <https://doi.org/10.1109/TCSS.2020.3037040>. Acesso em: 29 de out. de 2021.
- ELAHI, M.; BRAUNHOFER, M.; RICCI, F.; TKALČIČ, M. Personality-Based Active Learning for Collaborative Filtering Recommender Systems. In: HUTCHISON, D.; KANADE, T.; KITTLER, J.; KLEINBERG, J. M.; MATTERN, F.; MITCHELL, J. C.; NAOR, M.; NIERSTRASZ, O.; RANGAN, C. P.; STEFFEN, B.; SUDAN, M.; TERZOPOULOS, D.; TYGAR, D.; VARDI, M. Y.; WEIKUM, G.; BALDONI, M.; BAROGLIO, C.; BOELLA, G.; MICALIZIO, R. (Ed.). *AI\*IA 2013: Advances in Artificial Intelligence*. Cham: Springer International Publishing, 2013. v. 8249, p. 360–371. ISBN

978-3-319-03523-9 978-3-319-03524-6. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-03524-6\\_31](https://doi.org/10.1007/978-3-319-03524-6_31). Acesso em: 29 de out. de 2021.

ESPOSITO, C. M.; CERESA, A.; BUOLI, M. The Association Between Personality Traits and Dietary Choices: A Systematic Review. *Advances in Nutrition*, v. 12, n. 4, p. 1149–1159, 01 2021. ISSN 2161-8313. Disponível em: <https://doi.org/10.1093/advances/nmaa166>. Acesso em: 27 de dez. de 2021.

FEIST, J.; FEIST, G. J.; ROBERTS, T.-A. *Teorias da personalidade*. 8. ed. Porto Alegre: AMGH (Artmed McGraw-Hill), 2015. (tradução: Sandra Maria Mallmann da Rosa ; revisão técnica: Maria Cecília de Vilhena Moraes, Odette de Godoy Pinheiro). ISBN 978-85-8055-460-1.

FENG, H.; QIAN, X. Recommendation via user's personality and social contextual. In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13*. San Francisco, California, USA: ACM Press, 2013. p. 1521–1524. ISBN 978-1-4503-2263-8. Disponível em: <https://doi.org/10.1145/2505515.2507834>. Acesso em: 29 de out. de 2021.

FERNÁNDEZ-TOBIÁS, I.; BRAUNHOFER, M.; ELAHI, M.; RICCI, F.; CANTADOR, I. Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction (UMUAI)*, v. 26, n. 2-3, p. 221–255, jun. 2016. ISSN 0924-1868, 1573-1391. Disponível em: <https://doi.org/10.1007/s11257-016-9172-z>. Acesso em: 29 de out. de 2021.

FERNÁNDEZ-TOBIÁS, I.; CANTADOR, I. Personality-Aware Collaborative Filtering: An Empirical Study in Multiple Domains with Facebook Data. In: HEPP, M.; HOFFNER, Y. (Ed.). *E-Commerce and Web Technologies*. Cham: Springer International Publishing, 2014. v. 188, p. 125–137. ISBN 978-3-319-10490-4 978-3-319-10491-1. Series Title: Lecture Notes in Business Information Processing. Disponível em: [https://doi.org/10.1007/978-3-319-10491-1\\_13](https://doi.org/10.1007/978-3-319-10491-1_13). Acesso em: 29 de out. de 2021.

FERNÁNDEZ-TOBIÁS, I.; CANTADOR, I. On the Use of Cross-Domain User Preferences and Personality Traits in Collaborative Filtering. In: RICCI, F.; BONTCHEVA, K.; CONLAN, O.; LAWLESS, S. (Ed.). *User Modeling, Adaptation and Personalization (UMAP)*. Cham: Springer International Publishing, 2015. v. 9146, p. 343–349. ISBN 978-3-319-20266-2 978-3-319-20267-9. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-20267-9\\_29](https://doi.org/10.1007/978-3-319-20267-9_29). Acesso em: 29 de out. de 2021.

FERREIRA, T. B.; BUIAR, J. A.; FERNANDES, M. A.; PIMENTEL, A. R.; OLIVEIRA, L. S. Regras para formação de grupos de colaboração utilizando detecção automática de traços de personalidade. *Brazilian Journal of Computers in Education*, v. 28, p. 273–296, 2020. ISSN 2317-6121. Disponível em: <https://doi.org/10.5753/rbie.2020.28.0.273>. Acesso em: 29 de out. de 2021.

FERWERDA, B.; GRAUS, M.; VALL, A.; TKALČIČ, M.; SCHEDL, M. The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation

Lists. In: TKALČIČ, M.; CAROLIS, B. D.; GEMMIS, M. de; KOŠIR, A. (Ed.). *Proceedings of the 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) co-located with 10th ACM Conference on Recommender Systems (RecSys)*. Boston, MA, USA: CEUR-WS.org, 2016. (CEUR Workshop Proceedings, v. 1680), p. 43–47. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-1680/paper6.pdf>. Acesso em: 29 de out. de 2021.

FERWERDA, B.; TKALČIČ, M. Predicting users' personality from instagram pictures: Using visual and/or content features? In: *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP)*. New York, NY, USA: Association for Computing Machinery (ACM), 2018. p. 157–161. ISBN 9781450355896. Disponível em: <https://doi.org/10.1145/3209219.3209248>. Acesso em: 29 de out. de 2021.

FERWERDA, B.; TKALČIČ, M.; SCHEDL, M. Personality Traits and Music Genre Preferences: How Music Taste Varies Over Age Groups. In: BIELIKOVA, M.; BOGINA, V.; KUFLIK, T.; SASSON, R. (Ed.). *Proceedings of the 1st Workshop on Temporal Reasoning in Recommender Systems (RecTemp) co-located with 11th ACM Conference on Recommender Systems (RecSys)*. Como, Italy: CEUR-WS.org, 2017. (CEUR Workshop Proceedings, v. 1922), p. 16–20. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-1922/paper4.pdf>. Acesso em: 29 de out. de 2021.

FERWERDA, B.; TKALČIČ, M.; SCHEDL, M. Personality Traits and Music Genres: What Do People Prefer to Listen To? In: *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP)*. Bratislava Slovakia: Association for Computing Machinery (ACM), 2017. p. 285–288. ISBN 978-1-4503-4635-1. Disponível em: <https://doi.org/10.1145/3079628.3079693>. Acesso em: 29 de out. de 2021.

FERWERDA, B.; YANG, E.; SCHEDL, M.; TKALČIČ, M. Personality and taxonomy preferences, and the influence of category choice on the user experience for music streaming services. *Multimedia Tools and Applications*, v. 78, n. 14, p. 20157–20190, jul. 2019. ISSN 1380-7501, 1573-7721. Disponível em: <https://doi.org/10.1007/s11042-019-7336-7>. Acesso em: 29 de out. de 2021.

FIVE. *Five Labs — See the personality behind your posts*. 2015. Product of Five.com. [The tool is no longer supported]. Disponível em: <http://labs.five.com/>. Acesso em: 24 de abr. de 2015.

FORD, K. *Brands Laid Bare: Using Market Research for Evidence-Based Brand Management*. 1. ed. England: John Wiley & Sons, 2005. ISBN 978-0-470-01283-3.

GANTNER, Z.; RENDLE, S.; FREUDENTHALER, C.; SCHMIDT-THIEME, L. Mymedialite: A free recommender system library. In: *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2011. p. 305–308. ISBN 9781450306836. Disponível em: <https://doi.org/10.1145/2043932.2043989>. Acesso em: 29 de out. de 2021.

GELLI, F.; HE, X.; CHEN, T.; CHUA, T.-S. How Personality Affects our Likes: Towards a Better Understanding of Actionable Images. In: *Proceedings of the 25th ACM international conference on Multimedia*. Mountain View California USA: Association for Computing

Machinery (ACM), 2017. p. 1828–1837. ISBN 978-1-4503-4906-2. Disponível em: [⟨https://doi.org/10.1145/3123266.3127909⟩](https://doi.org/10.1145/3123266.3127909). Acesso em: 29 de out. de 2021.

GIANOTTI, R. C.; CAZELLA, S. C.; BEHAR, P. A. A model for integrating personality traits into an educational recommender system. In: *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*. [s.n.], 2019. v. 2161-377X, p. 383–385. Disponível em: [⟨https://doi.org/10.1109/ICALT.2019.00119⟩](https://doi.org/10.1109/ICALT.2019.00119). Acesso em: 29 de out. de 2021.

GOLBECK, J.; NORRIS, E. Personality, movie preferences, and recommendations. In: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. New York, NY, USA: Association for Computing Machinery (ACM), 2013. p. 1414–1415. ISBN 9781450322409. Disponível em: [⟨https://doi.org/10.1145/2492517.2492572⟩](https://doi.org/10.1145/2492517.2492572). Acesso em: 29 de out. de 2021.

GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston: Addison-Wesley Publishing Company, 1989. ISBN 9780201157673.

GONZÁLEZ, G.; ROSA, J. L. D. L.; MONTANER, M.; DELFÍN, S. Embedding emotional context in recommender systems. In: IEEE. *2007 IEEE 23rd International Conference on Data Engineering Workshop*. 2007. p. 845–852. Disponível em: [⟨https://doi.org/10.1109/ICDEW.2007.4401075⟩](https://doi.org/10.1109/ICDEW.2007.4401075). Acesso em: 29 de out. de 2021.

GOSLING, S. D.; RENTFROW, P. J.; SWANN, W. B. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, Elsevier, v. 37, n. 6, p. 504–528, 2003. ISSN 0092-6566. Disponível em: [⟨https://doi.org/10.1016/S0092-6566\(03\)00046-1⟩](https://doi.org/10.1016/S0092-6566(03)00046-1). Acesso em: 29 de out. de 2021.

GUNTUKU, S. C.; ROY, S.; WEISI, L. Personality Modeling Based Image Recommendation. In: HE, X.; LUO, S.; TAO, D.; XU, C.; YANG, J.; HASAN, M. A. (Ed.). *MultiMedia Modeling*. Cham: Springer International Publishing, 2015. v. 8936, p. 171–182. ISBN 978-3-319-14441-2 978-3-319-14442-9. Series Title: Lecture Notes in Computer Science. Disponível em: [⟨https://doi.org/10.1007/978-3-319-14442-9\\_15⟩](https://doi.org/10.1007/978-3-319-14442-9_15). Acesso em: 29 de out. de 2021.

GUPTA, S.; GULATI, P.; BHATIA, S.; MADAN, R. An Automatic Approach to Music Recommendations Based on Individual Personality Traits. *SSRN Electronic Journal*, 2020. ISSN 1556-5068. Disponível em: [⟨https://doi.org/10.2139/ssrn.3565276⟩](https://doi.org/10.2139/ssrn.3565276). Acesso em: 29 de out. de 2021.

HAFSHEJANI, Z. Y.; KAEDI, M.; FATEMI, A. Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. *Electronic Commerce Research*, Kluwer Academic Publishers, USA, v. 18, n. 4, p. 813–836, December 2018. ISSN 1389-5753. Disponível em: [⟨https://doi.org/10.1007/s10660-018-9287-x⟩](https://doi.org/10.1007/s10660-018-9287-x). Acesso em: 29 de out. de 2021.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. 3. ed. Waltham, MA, USA: Morgan Kaufmann Publishers, 2012. (The Morgan Kaufmann Series in Data Management Systems). ISBN 978-0-12-381479-1. Disponível em: [⟨https://doi.org/10.1016/B978-0-12-381479-1⟩](https://doi.org/10.1016/B978-0-12-381479-1).

[//www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques](http://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques)). Acesso em: 29 de out. de 2021.

HARIADI, a. I.; NURJANAH, D. Hybrid attribute and personality based recommender system for book recommendation. In: *2017 International Conference on Data and Software Engineering (ICoDSE)*. Palembang: IEEE, 2017. p. 1–5. ISBN 978-1-5386-1449-5. Disponível em: <https://doi.org/10.1109/ICODSE.2017.8285874>). Acesso em: 29 de out. de 2021.

HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Wiley, Royal Statistical Society, v. 28, n. 1, p. 100–108, 1979. ISSN 00359254, 14679876. Disponível em: <https://doi.org/10.2307/2346830>). Acesso em: 29 de out. de 2021.

HERLOCKER, J. L.; KONSTAN, J. A.; TERVEEN, L. G.; RIEDL, J. T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, Association for Computing Machinery (ACM), New York, NY, USA, v. 22, n. 1, p. 5–53, jan. 2004. ISSN 1046-8188. Disponível em: <https://doi.org/10.1145/963770.963772>). Acesso em: 29 de out. de 2021.

HIRSH, J. B.; PETERSON, J. B. Personality and language use in self-narratives. *Journal of Research in Personality*, v. 43, n. 3, p. 524–527, 2009. ISSN 0092-6566. Disponível em: <https://doi.org/10.1016/j.jrp.2009.01.006>). Acesso em: 29 de out. de 2021.

HU, R. Design and user issues in personality-based recommender systems. In: *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys)*. Barcelona, Spain: ACM Press, 2010. p. 357. ISBN 978-1-60558-906-0. Disponível em: <https://doi.org/10.1145/1864708.1864790>). Acesso em: 29 de out. de 2021.

HU, R.; PU, P. Acceptance issues of personality-based recommender systems. In: *Proceedings of the Third ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2009. p. 221–224. ISBN 9781605584355. Disponível em: <https://doi.org/10.1145/1639714.1639753>). Acesso em: 29 de out. de 2021.

HU, R.; PU, P. A Study on User Perception of Personality-Based Recommender Systems. In: HUTCHISON, D.; KANADE, T.; KITTLER, J.; KLEINBERG, J. M.; MATTERN, F.; MITCHELL, J. C.; NAOR, M.; NIERSTRASZ, O.; RANGAN, C. P.; STEFFEN, B.; SUDAN, M.; TERZOPOULOS, D.; TYGAR, D.; VARDI, M. Y.; WEIKUM, G.; BRA, P. D.; KOBASA, A.; CHIN, D. (Ed.). *User Modeling, Adaptation, and Personalization (UMAP)*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. v. 6075, p. 291–302. ISBN 978-3-642-13469-2 978-3-642-13470-8. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-642-13470-8\\_27](https://doi.org/10.1007/978-3-642-13470-8_27)). Acesso em: 29 de out. de 2021.

HU, R.; PU, P. Using Personality Information in Collaborative Filtering for New Users. In: GEYER, W.; FREYNE, J.; MOBASHER, B.; ANAND, S. S.; DUGAN, C. (Ed.). *Proceedings of the 2nd Workshop on Recommender Systems and the Social Web (RSWeb) co-located with 4th ACM Conference on Recommender Systems (RecSys)*. Barcelona, Spain:

[s.n.], 2010. p. 17–24. Disponível em: <https://hci.epfl.ch/publications/2010/hupu-RSWE B2010.pdf>). Acesso em: 29 de out. de 2021.

HU, R.; PU, P. Enhancing collaborative filtering systems with personality information. In: *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys)*. Chicago, Illinois, USA: Association for Computing Machinery (ACM), 2011. p. 197–204. ISBN 978-1-4503-0683-6. Disponível em: <https://doi.org/10.1145/2043932.2043969>). Acesso em: 29 de out. de 2021.

HU, R.; PU, P. Exploring personality's effect on users' rating behavior. In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. Toronto Ontario Canada: Association for Computing Machinery (ACM), 2014. p. 2599–2604. ISBN 978-1-4503-2474-8. Disponível em: <https://doi.org/10.1145/2559206.2581317>). Acesso em: 29 de out. de 2021.

HUANG, Y.; LIU, H.; LI, W.; WANG, Z.; HU, X.; WANG, W. Lifestyles in Amazon: Evidence from online reviews enhanced recommender system. *International Journal of Market Research*, v. 62, n. 6, p. 689–706, nov. 2020. ISSN 1470-7853, 2515-2173. Disponível em: <https://doi.org/10.1177/1470785319844146>). Acesso em: 29 de out. de 2021.

IBM. *IBM Cloud Docs — Personality Insights*. 2021. Disponível em: <http://cloud.ibm.com/docs/personality-insights>). Acesso em: 29 de out. de 2021.

JAQUES, P. A.; NUNES, M. A. S. N. Computação afetiva aplicada à educação. In: SAMPAIO, F. F.; PIMENTEL, M.; SANTOS, E. O. (Ed.). *Informática na Educação: games, inteligência artificial, realidade virtual/aumentada e computação ubíqua*. Porto Alegre: SBC (Sociedade Brasileira de Computação), 2021, (Série Informática na Educação, v.7). Disponível em: <http://ieducacao.ceie-br.org/computacaoafetiva>). Acesso em: 29 de out. de 2021.

JEONG, C.-S.; LEE, J.-Y.; JUNG, K.-D. Adaptive Recommendation System for Tourism by Personality Type Using Deep Learning. *International Journal of Internet, Broadcasting and Communication*, v. 12, n. 1, p. 55–60, fev. 2020. Disponível em: <https://doi.org/10.7236/IJIBC.2020.12.1.55>). Acesso em: 29 de out. de 2021.

JOHN, O. P.; DONAHUE, E. M.; KENTLE, R. L. Big five inventory. *Journal of Personality and Social Psychology*, 1991. Disponível em: <https://doi.org/10.1037/t07550-000>). Acesso em: 29 de out. de 2021.

JOHN, O. P.; SRIVASTAVA, S. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In: *Handbook of personality: Theory and research, 2nd ed.* New York, NY, US: Guilford Press, 1999. p. 102–138. ISBN 1-57230-483-9 (Hardcover). Disponível em: <https://pages.uoregon.edu/sanjay/pubs/bigfive.pdf>). Acesso em: 29 de out. de 2021.

JOHNSON, J. A. Web-based personality assessment. In: *71st Annual Meeting of the Eastern Psychological Association*. Baltimore, MD: Pennsylvania State University, 2000. Disponível em: <http://www.personal.psu.edu/faculty/j/5/j5j/papers/ConferencePapers/2000EPA.pdf>). Acesso em: 29 de out. de 2021.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, American Association for the Advancement of Science, v. 349, n. 6245, p. 255–260, 2015. Disponível em: <https://doi.org/10.1126/science.aaa8415>. Acesso em: 27 de dez. de 2021.

KARUMUR, R. P.; KONSTAN, J. A. Relating Newcomer Personality to Survival and Activity in Recommender Systems. In: *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (UMAP)*. Halifax Nova Scotia Canada: Association for Computing Machinery (ACM), 2016. p. 195–205. ISBN 978-1-4503-4368-8. Disponível em: <https://doi.org/10.1145/2930238.2930246>. Acesso em: 29 de out. de 2021.

KARUMUR, R. P.; NGUYEN, T. T.; KONSTAN, J. A. Exploring the Value of Personality in Predicting Rating Behaviors: A Study of Category Preferences on MovieLens. In: *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)*. Boston Massachusetts USA: Association for Computing Machinery (ACM), 2016. p. 139–142. ISBN 978-1-4503-4035-9. Disponível em: <https://doi.org/10.1145/2959100.2959140>. Acesso em: 29 de out. de 2021.

KARUMUR, R. P.; NGUYEN, T. T.; KONSTAN, J. A. Personality, User Preferences and Behavior in Recommender systems. *Information Systems Frontiers*, v. 20, n. 6, p. 1241–1265, dez. 2018. ISSN 1387-3326, 1572-9419. Disponível em: <https://doi.org/10.1007/s10796-017-9800-0>. Acesso em: 29 de out. de 2021.

KHAN, E. M.; MUKTA, M. S. H.; ALI, M. E.; MAHMUD, J. Predicting Users' Movie Preference and Rating Behavior from Personality and Values. *ACM Transactions on Interactive Intelligent Systems*, v. 10, n. 3, p. 1–25, nov. 2020. ISSN 2160-6455, 2160-6463. Disponível em: <https://doi.org/10.1145/3338244>. Acesso em: 29 de out. de 2021.

KHODABANDEHLOU, S.; GOLPAYEGANI, S. A. H.; RAHMAN, M. Z. An effective recommender system based on personality traits, demographics and behavior of customers in time context. *Data Technologies and Applications*, v. 55, n. 1, p. 149–174, nov. 2020. ISSN 2514-9288. Disponível em: <https://doi.org/10.1108/DTA-04-2020-0094>. Acesso em: 29 de out. de 2021.

KHWAJA, M.; FERRER, M.; IGLESIAS, J. O.; FAISAL, A. A.; MATIC, A. Aligning daily activities with personality: Towards a recommender system for improving wellbeing. In: *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2019. p. 368–372. ISBN 9781450362436. Disponível em: <https://doi.org/10.1145/3298689.3347020>. Acesso em: 29 de out. de 2021.

KOUKI, P.; SCHAFFER, J.; PUJARA, J.; O'DONOVAN, J.; GETOOR, L. Generating and Understanding Personalized Explanations in Hybrid Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*, v. 10, n. 4, p. 1–40, dez. 2020. ISSN 2160-6455, 2160-6463. Disponível em: <https://doi.org/10.1145/3365843>. Acesso em: 29 de out. de 2021.

- LAMPRECHT, J.; AHMAD, R.; ROBRA-BISSANTZ, S. Nutzung von Persönlichkeitsprofilen zur Steigerung von Kooperation in virtuellen Teams. *HMD Praxis der Wirtschaftsinformatik*, Springer, v. 55, n. 3, p. 540–551, jun 2018. ISSN 2198-2775. Disponível em: <https://doi.org/10.1365/s40702-017-0377-3>. Acesso em: 29 de out. de 2021.
- LEX, E.; KOWALD, D.; SEITLINGER, P.; TRAN, T. N. T.; FELFERNIG, A.; SCHEDL, M. Psychology-informed Recommender Systems. *Foundations and Trends® in Information Retrieval*, v. 15, n. 2, p. 134–242, 2021. ISSN 1554-0669, 1554-0677. Disponível em: <https://doi.org/10.1561/15000000090>. Acesso em: 29 de out. de 2021.
- LI, L.; ZHU, H.; ZHAO, S.; DING, G.; JIANG, H.; TAN, A. Personality Driven Multi-task Learning for Image Aesthetic Assessment. In: *2019 IEEE International Conference on Multimedia and Expo (ICME)*. Shanghai, China: IEEE, 2019. p. 430–435. ISBN 978-1-5386-9552-4. Disponível em: <https://doi.org/10.1109/ICME.2019.00081>. Acesso em: 29 de out. de 2021.
- LI, L.; ZHU, H.; ZHAO, S.; DING, G.; LIN, W. Personality-Assisted Multi-Task Learning for Generic and Personalized Image Aesthetics Assessment. *IEEE Transactions on Image Processing*, v. 29, p. 3898–3910, 2020. ISSN 1057-7149, 1941-0042. Disponível em: <https://doi.org/10.1109/TIP.2020.2968285>. Acesso em: 29 de out. de 2021.
- LIMA, E. S. de; FEIJÓ, B.; FURTADO, A. L. Player behavior and personality modeling for interactive storytelling in games. *Entertainment Computing*, v. 28, p. 32–48, dez. 2018. ISSN 1875-9521. Disponível em: <https://doi.org/10.1016/j.entcom.2018.08.003>. Acesso em: 29 de out. de 2021.
- LINDEN, R. *Algoritmos Genéticos*. 3. ed. Rio de Janeiro: Ciência Moderna, 2012. ISBN 9788539901951.
- LIU, H.; HUANG, Y.; WANG, Z.; LIU, K.; HU, X.; WANG, W. Personality or Value: A Comparative Study of Psychographic Segmentation Based on an Online Review Enhanced Recommender System. *Applied Sciences*, v. 9, n. 10, p. 1992, maio 2019. ISSN 2076-3417. Disponível em: <https://doi.org/10.3390/app9101992>. Acesso em: 29 de out. de 2021.
- LIU, R.; HU, X. A Multimodal Music Recommendation System with Listeners' Personality and Physiological Signals. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. Virtual Event China: Association for Computing Machinery (ACM), 2020. p. 357–360. ISBN 978-1-4503-7585-6. Disponível em: <https://doi.org/10.1145/3383583.3398623>. Acesso em: 29 de out. de 2021.
- LIU, Z.; WANG, Y.; MAHMUD, J.; AKKIRAJU, R.; SCHOUDT, J.; XU, A.; DONOVAN, B. To buy or not to buy? Understanding the role of personality traits in predicting consumer behaviors. In: SPIRO, E.; AHN, Y.-Y. (Ed.). *Proceedings of the International Conference on Social Informatics (SocInfo)*. Cham: Springer International Publishing, 2016. p. 337–346. ISBN 978-3-319-47874-6. Disponível em: [https://doi.org/10.1007/978-3-319-47874-6\\_24](https://doi.org/10.1007/978-3-319-47874-6_24). Acesso em: 29 de out. de 2021.

LIU, Z.; XU, A.; WANG, Y.; SCHOUDT, J.; MAHMUD, J.; AKKIRAJU, R. Does personality matter? A study of personality and situational effects on consumer behavior. In: *Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT)*. New York, NY, USA: Association for Computing Machinery (ACM), 2017. p. 185–193. ISBN 9781450347082. Disponível em: <https://doi.org/10.1145/3078714.3078733>. Acesso em: 29 de out. de 2021.

LU, F.; TINTAREV, N. A diversity adjusting strategy with personality for music recommendation. In: BRUSILOVSKY, P.; GEMMIS, M. de; FELFERNIG, A.; LOPS, P.; O'DONOVAN, J.; SEMERARO, G.; WILLEMSSEN, M. C. (Ed.). *Proceedings of the 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS) co-located with 12th ACM Conference on Recommender Systems (RecSys)*. Vancouver, Canada: CEUR-WS.org, 2018. (CEUR Workshop Proceedings, v. 2225), p. 7–14. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-2225/paper2.pdf>. Acesso em: 29 de out. de 2021.

MACHADO-OLIVEIRA, M. C.; NEZLEK, J. B.; RODRIGUES, H.; SANT'ANA, A. S. Personality traits and food consumption: an overview of recent research. *Current Opinion in Food Science*, v. 33, p. 91–97, 2020. ISSN 2214-7993. Sensory Science & Consumer Perception • Food Physics and Material Science. Disponível em: <https://doi.org/10.1016/j.cofs.2020.02.005>. Acesso em: 27 de dez. de 2021.

MAIRESSE, F.; WALKER, M. A.; MEHL, M. R.; MOORE, R. K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, v. 30, p. 457–500, nov. 2007. ISSN 1076-9757. Disponível em: <https://doi.org/10.1613/jair.2349>. Acesso em: 29 de out. de 2021.

MARKOVIKJ, D.; GIEVSKA, S.; KOSINSKI, M.; STILLWELL, D. Mining facebook data for predictive personality modeling. *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, v. 7, p. 23–26, August 2021. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14466>. Acesso em: 29 de out. de 2021.

MASLOW, A. H. A theory of human motivation. *Psychological Review*, American Psychological Association, v. 50, n. 4, p. 370–396, 1943. ISSN 1939-1471, 0033-295X. Disponível em: <https://doi.org/10.1037/h0054346>. Acesso em: 29 de out. de 2021.

MCCRAE, R. R.; COSTA, P. T. Personality trait structure as a human universal. *The American Psychologist*, American Psychological Association, v. 52, n. 5, p. 509–516, May 1997. ISSN 0003-066X. Disponível em: <https://doi.org/10.1037//0003-066x.52.5.509>. Acesso em: 29 de out. de 2021.

MCCRAE, R. R.; COSTA, P. T. *Personality in Adulthood: A Five-Factor Theory Perspective*. 2. ed. New York, NY, USA: Guilford Press, 2003. ISBN 1-57230-827-3.

MEHTA, Y.; FATEHI, S.; KAZAMEINI, A.; STACHL, C.; CAMBRIA, E.; EETEMADI, S. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In: IEEE. *2020 IEEE International Conference on Data Mining (ICDM)*. 2020. p. 1184–1189. Disponível em: <https://doi.org/10.1109/ICDM50108.2020.00146>. Acesso em: 29 de out. de 2021.

MILLECAMP, M.; HTUN, N. N.; CONATI, C.; VERBERT, K. What's in a User? Towards Personalising Transparency for Music Recommender Interfaces. In: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. Genoa Italy: Association for Computing Machinery (ACM), 2020. p. 173–182. ISBN 978-1-4503-6861-2. Disponível em: <https://doi.org/10.1145/3340631.3394844>. Acesso em: 29 de out. de 2021.

MORENO, J. D.; MARTÍNEZ-HUERTAS, J. Á.; OLMOS, R.; JORGE-BOTANA, G.; BOTELLA, J. Can personality traits be measured analyzing written language? A meta-analytic study on computational methods. *Personality and Individual Differences*, v. 177, p. 110818, jul. 2021. ISSN 0191-8869. Disponível em: <https://doi.org/10.1016/j.paid.2021.110818>. Acesso em: 29 de out. de 2021.

MOSCATO, V.; PICARIELLO, A.; SPERLI, G. An emotional recommender system for music. *IEEE Intelligent Systems*, p. 1–1, 2020. ISSN 1541-1672, 1941-1294. Disponível em: <https://doi.org/10.1109/MIS.2020.3026000>. Acesso em: 29 de out. de 2021.

MUKTA, M. S. H.; ALI, M. E.; MAHMUD, J. Identifying and validating personality traits-based homophilies for an egocentric network. *Social Network Analysis and Mining*, v. 6, n. 1, p. 74, dez. 2016. ISSN 1869-5450, 1869-5469. Disponível em: <https://doi.org/10.1007/s13278-016-0383-4>. Acesso em: 29 de out. de 2021.

NAJAFIAN, S.; DELIC, A.; TKALČIĆ, M.; TINTAREV, N. Factors Influencing Privacy Concern for Explanations of Group Recommendation. In: *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. Utrecht Netherlands: Association for Computing Machinery (ACM), 2021. p. 14–23. ISBN 978-1-4503-8366-0. Disponível em: <https://doi.org/10.1145/3450613.3456845>. Acesso em: 29 de out. de 2021.

NALMPANTIS, O.; TJORTJIS, C. The 50/50 Recommender: A Method Incorporating Personality into Movie Recommender Systems. In: BORACCHI, G.; ILIADIS, L.; JAYNE, C.; LIKAS, A. (Ed.). *Engineering Applications of Neural Networks*. Cham: Springer International Publishing, 2017. v. 744, p. 498–507. ISBN 978-3-319-65171-2 978-3-319-65172-9. Series Title: Communications in Computer and Information Science. Disponível em: [https://doi.org/10.1007/978-3-319-65172-9\\_42](https://doi.org/10.1007/978-3-319-65172-9_42). Acesso em: 29 de out. de 2021.

NEEHAL, N.; MOTTALIB, M. A. Prediction of Preferred Personality for Friend Recommendation in Social Networks using Artificial Neural Network. In: *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. Cox'sBazar, Bangladesh: IEEE, 2019. p. 1–6. ISBN 978-1-5386-9111-3. Disponível em: <https://doi.org/10.1109/ECACE.2019.8679375>. Acesso em: 29 de out. de 2021.

NGUYEN, T. T.; HARPER, F. M.; TERVEEN, L.; KONSTAN, J. A. User Personality and User Satisfaction with Recommender Systems. *Information Systems Frontiers*, Springer, v. 20, n. 6, p. 1173–1189, dez. 2018. ISSN 1387-3326, 1572-9419. Disponível em: <https://doi.org/10.1007/s10796-017-9782-y>. Acesso em: 29 de out. de 2021.

NING, H.; DHELIM, S.; AUNG, N. PersoNet: Friend Recommendation System Based on Big-Five Personality Traits and Hybrid Filtering. *IEEE Transactions on Computational Social Systems*, v. 6, n. 3, p. 394–402, jun. 2019. ISSN 2329-924X, 2373-7476. Disponível em: <https://doi.org/10.1109/TCSS.2019.2903857>. Acesso em: 29 de out. de 2021.

NUNES, M. A. S. N. *Recommender Systems based on Personality Traits*. Tese (Doutorado) — Université Montpellier II — Sciences et Techniques du Languedoc, Montpellier, 2008. Disponível em: <https://tel.archives-ouvertes.fr/tel-00348370>. Acesso em: 29 de out. de 2021.

NUNES, M. A. S. N. *Recommender Systems based on Personality Traits: Could human psychological aspects influence the computer decision-making process?* [S.l.]: VDM Verlag, 2009. ISBN 978-3-639-16976-8.

NUNES, M. A. S. N. Computação Afetiva personalizando interfaces, interações e recomendações de produtos, serviços e pessoas em Ambientes computacionais. In: *Projetos e Pesquisas em Ciência da Computação no DCOMP/PROCC/UFS*. São Cristóvão - SE: UFS, 2012. p. 113–149. ISBN 978-85-7822-240-6. Disponível em: <http://almanaquesdacomputacao.com.br/gutanunes/publications/livroDcomp.pdf>. Acesso em: 29 de out. de 2021.

NUNES, M. A. S. N.; BEZERRA, J. S.; OLIVEIRA, A. A. PersonalityML: a markup language to standardize the user personality in recommender systems. *Revista Gestão, Inovação e Tecnologia*, v. 2, n. 3, p. 255–273, set. 2012. ISSN 2237-0722. Disponível em: <https://revistageintec.net/index.php/revista/article/view/50>. Acesso em: 29 de out. de 2021.

NUNES, M. A. S. N.; BEZERRA, J. S.; REINERT, D.; MORAES, D.; SILVA, É. P.; PEREIRA, A. J. S. Computação Afetiva e sua influência na personalização de Ambientes Educacionais: Gerando equipes compatíveis para uso em AVAs na EaD. In: *Educação e Ciberespaço: Estudos, Propostas e Desafios*. Aracaju - SE: Virtus, 2010. v. 1, p. 308–347. ISBN 978-85-64268-00-5. Disponível em: <http://almanaquesdacomputacao.com.br/gutanunes/publications/capitulo.pdf>. Acesso em: 29 de out. de 2021.

NUNES, M. A. S. N.; HU, R. Personality-based recommender systems: An overview. In: *Proceedings of the Sixth ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2012. p. 5–6. ISBN 9781450312707. Disponível em: <https://doi.org/10.1145/2365952.2365957>. Acesso em: 29 de out. de 2021.

OLIVEIRA, R. S. d. *Diversificação em sistemas de recomendação utilizando uma abordagem baseada em aspectos*. Tese (Doutorado) — Universidade Federal de Campina Grande (UFCG) — Programa de Pós-Graduação em Ciência da Computação, Campina Grande, Paraíba, Brasil, 2020. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/13223>. Acesso em: 29 de out. de 2021.

ONORI, M.; MICARELLI, A.; SANSONETTI, G. A Comparative Analysis of Personality-Based Music Recommender Systems. In: TKALČIĆ, M.; CAROLIS,

- B. D.; GEMMIS, M. de; KOŠIR, A. (Ed.). *Proceedings of the 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) co-located with 10th ACM Conference on Recommender Systems (RecSys)*. Boston, MA, USA: CEUR-WS.org, 2016. (CEUR Workshop Proceedings, v. 1680), p. 55–59. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-1680/paper8.pdf>. Acesso em: 29 de out. de 2021.
- PAIVA, F. A. P.; COSTA, J. A. F.; SILVA, C. R. M. A Personality-Based Recommender System for Semantic Searches in Vehicles Sales Portals. In: PISÓN, F. J. Martínez de; URRACA, R.; QUINTIÁN, H.; CORCHADO, E. (Ed.). *Hybrid Artificial Intelligent Systems*. Cham: Springer International Publishing, 2017. v. 10334, p. 600–612. ISBN 978-3-319-59649-5 978-3-319-59650-1. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-59650-1\\_51](https://doi.org/10.1007/978-3-319-59650-1_51). Acesso em: 29 de out. de 2021.
- PARK, Y. Recommender technologies and emerging applications. In: *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*. IGI Global, 2019. p. 458–470. Disponível em: <https://doi.org/10.4018/978-1-5225-2255-3.ch163>. Acesso em: 29 de out. de 2021.
- PATRO, S. G. K.; MISHRA, B. K.; PANDA, S. K.; KUMAR, R.; APOORVA, A. Hybrid social recommender systems for electronic commerce: A review. In: *Proceedings of the International Conference on Computer Science, Engineering and Applications (ICCSEA)*. Gunupur, India: IEEE, 2020. p. 1–6. ISBN 978-1-7281-5830-3. Disponível em: <https://doi.org/10.1109/ICCSEA49143.2020.9132881>. Acesso em: 29 de out. de 2021.
- PENNINGTON, J.; SOCHER, R.; MANNING, C. D. GloVe: Global Vectors for Word Representation. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014. p. 1532–1543. Disponível em: <https://doi.org/10.3115/v1/D14-1162>. Acesso em: 27 de dez. de 2021.
- POTASH, P.; RUMSHISKY, A. Recommender System Incorporating User Personality Profile through Analysis of Written Reviews. In: TKALČIČ, M.; CAROLIS, B. D.; GEMMIS, M. de; KOŠIR, A. (Ed.). *Proceedings of the 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) co-located with 10th ACM Conference on Recommender Systems (RecSys)*. Boston, MA, USA: CEUR-WS.org, 2016. (CEUR Workshop Proceedings, v. 1680), p. 60–66. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-1680/paper9.pdf>. Acesso em: 29 de out. de 2021.
- PRESSE, F. *Cambridge Analytica se declara culpada em caso de uso de dados do Facebook*. 2019. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2019/01/09/cambridge-analytica-se-declara-culpada-por-uso-de-dados-do-facebook.ghtml>. Acesso em: 29 de out. de 2021.
- QAMHIEH, M.; SAMMANEH, H.; DEMAIDI, M. N. PCRS: Personalized Career-Path Recommender System for Engineering Students. *IEEE Access*, v. 8, p. 214039–214049, 2020. ISSN 2169-3536. Disponível em: <https://doi.org/10.1109/ACCESS.2020.3040338>. Acesso em: 29 de out. de 2021.

- QUAN, Z. Collaborative filtering recommendation based on user personality. In: *Proceedings of the 6th International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII)*. Xi'an, China: IEEE, 2013. v. 3, p. 307–310. ISBN 978-1-4799-3985-5. ISSN 2155-1456. Disponível em: <https://doi.org/10.1109/ICIII.2013.6703579>. Acesso em: 29 de out. de 2021.
- QUIJANO-SANCHEZ, L.; RECIO-GARCIA, J. A.; DIAZ-AGUDO, B. Personality and Social Trust in Group Recommendations. In: *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*. Arras, France: IEEE, 2010. p. 121–126. ISBN 978-1-4244-8817-9. Disponível em: <https://doi.org/10.1109/ICTAI.2010.92>. Acesso em: 29 de out. de 2021.
- QUIJANO-SANCHEZ, L.; RECIO-GARCIA, J. A.; DIAZ-AGUDO, B. HappyMovie: A Facebook Application for Recommending Movies to Groups. In: *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*. Boca Raton, FL, USA: IEEE, 2011. p. 239–244. ISBN 978-1-4577-2068-0 978-0-7695-4596-7. Disponível em: <https://doi.org/10.1109/ICTAI.2011.44>. Acesso em: 29 de out. de 2021.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>. Acesso em: 29 de out. de 2021.
- RAMMSTEDT, B.; JOHN, O. P. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, Elsevier, v. 41, n. 1, p. 203–212, 2007. ISSN 0092-6566. Disponível em: <https://doi.org/10.1016/j.jrp.2006.02.001>. Acesso em: 29 de out. de 2021.
- RASMUSSEN, C. E. Gaussian processes in machine learning. In: BOUSQUET, O.; LUXBURG, U. von; RÄTSCH, G. (Ed.). *Advanced Lectures on Machine Learning: Summer School on Machine Learning*. Berlin, Heidelberg: Springer, 2003, (Lecture Notes in Computer Science, v. 3176). p. 63–71. ISBN 978-3-540-28650-9. Disponível em: [https://doi.org/10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4). Acesso em: 27 de dez. de 2021.
- RECIO-GARCIA, J. A.; JIMENEZ-DIAZ, G.; SANCHEZ-RUIZ, A. A.; DIAZ-AGUDO, B. Personality aware recommendations to groups. In: *Proceedings of the Third ACM Conference on Recommender Systems (RecSys)*. New York, New York, USA: ACM Press, 2009. p. 325. ISBN 978-1-60558-435-5. Disponível em: <https://doi.org/10.1145/1639714.1639779>. Acesso em: 29 de out. de 2021.
- RENDLE, S.; FREUDENTHALER, C.; GANTNER, Z.; SCHMIDT-THIEME, L. BPR: Bayesian Personalized Ranking from Implicit Feedback. In: *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*. Arlington, Virginia, USA: AUAI Press, 2009. p. 452–461. ISBN 9780974903958. Disponível em: <https://dl.acm.org/doi/10.5555/1795114.1795167>. Acesso em: 21 de dez. de 2021.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. (Ed.). *Recommender Systems Handbook*. Boston, MA: Springer US, 2015. ISBN 978-1-4899-7636-9 978-1-4899-7637-6. Disponível em: <https://doi.org/10.1007/978-1-4899-7637-6>. Acesso em: 29 de out. de 2021.

ROFFO, G. Towards Personality-Aware Recommendation. *arXiv:1607.05088 [cs]*, jul. 2016. ArXiv: 1607.05088. Disponível em: <http://arxiv.org/abs/1607.05088>. Acesso em: 29 de out. de 2021.

ROFFO, G.; VINCIARELLI, A. Personality in Computational Advertising: A Benchmark. In: TKALČIČ, M.; CAROLIS, B. D.; GEMMIS, M. de; KOŠIR, A. (Ed.). *Proceedings of the 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) co-located with 10th ACM Conference on Recommender Systems (RecSys)*. Boston, MA, USA: CEUR-WS.org, 2016. (CEUR Workshop Proceedings, v. 1680), p. 18–25. ISSN 1613-0073. Disponível em: <http://ceur-ws.org/Vol-1680/paper3.pdf>. Acesso em: 29 de out. de 2021.

ROSHCHINA, A.; CARDIFF, J.; ROSSO, P. TWIN: Personality-based Intelligent Recommender System. *Journal of Intelligent & Fuzzy Systems*, IOS Press, v. 28, n. 5, p. 2059–2071, jun. 2015. ISSN 1064-1246, 1875-8967. Disponível em: <https://doi.org/10.3233/IFS-141484>. Acesso em: 29 de out. de 2021.

RStudio Team. *RStudio: Integrated Development Environment for R*. Boston, MA, 2021. Disponível em: <http://www.rstudio.com/>. Acesso em: 29 de out. de 2021.

SANTOS, O. C.; SANEIRO, M.; SALMERON-MAJADAS, S.; BOTICARIO, J. G. A Methodological Approach to Eliciting Affective Educational Recommendations. In: *2014 IEEE 14th International Conference on Advanced Learning Technologies*. Athens, Greece: IEEE, 2014. p. 529–533. ISBN 978-1-4799-4038-7. Disponível em: <https://doi.org/10.1109/ICALT.2014.234>. Acesso em: 29 de out. de 2021.

SCHEDL, M.; MELENHORST, M.; LIEM, C. C. S.; MARTORELL, A.; MAYOR, Ó.; TKALČIČ, M. A personality-based adaptive system for visualizing classical music performances. In: *Proceedings of the 7th International Conference on Multimedia Systems*. Klagenfurt Austria: Association for Computing Machinery (ACM), 2016. p. 1–7. ISBN 978-1-4503-4297-1. Disponível em: <https://doi.org/10.1145/2910017.2910604>. Acesso em: 29 de out. de 2021.

SCHWARTZ, H. A.; EICHSTAEDT, J. C.; KERN, M. L.; DZIURZYNSKI, L.; RAMONES, S. M.; AGRAWAL, M.; SHAH, A.; KOSINSKI, M.; STILLWELL, D.; SELIGMAN, M. E. P.; UNGAR, L. H. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE*, v. 8, n. 9, p. e73791, set. 2013. ISSN 1932-6203. Disponível em: <https://doi.org/10.1371/journal.pone.0073791>. Acesso em: 29 de out. de 2021.

SCHWARTZ, S. H. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Academic Press, v. 25, p. 1–65, 1992. ISSN 0065-2601. Disponível em: [https://doi.org/10.1016/S0065-2601\(08\)60281-6](https://doi.org/10.1016/S0065-2601(08)60281-6). Acesso em: 29 de out. de 2021.

SCHWARTZ, S. H. A proposal for measuring value orientations across nations. *Questionnaire package of the European Social Survey*, p. 259–290, 2003. Disponível em: [https://www.researchgate.net/publication/312444842\\_A\\_proposal\\_for\\_measuring\\_value\\_orientations\\_across\\_nations](https://www.researchgate.net/publication/312444842_A_proposal_for_measuring_value_orientations_across_nations). Acesso em: 29 de out. de 2021.

SHAYEGAN, M. J.; VALIZADEH, M. A Recommender System based on the analysis of personality traits in Telegram social network. *arXiv:2010.00643 [cs]*, out. 2020. ArXiv: 2010.00643. Disponível em: <http://arxiv.org/abs/2010.00643>. Acesso em: 29 de out. de 2021.

SPINELLI, S.; De Toffoli, A.; DINNELLA, C.; LAUREATI, M.; PAGLIARINI, E.; BENDINI, A.; BRAGHIERI, A.; Gallina Toschi, T.; SINESIO, F.; TORRI, L.; GASPERI, F.; ENDRIZZI, I.; MAGLI, M.; BORGOGNO, M.; di Salvo, R.; FAVOTTO, S.; PRESCOTT, J.; MONTELEONE, E. Personality traits and gender influence liking and choice of food pungency. *Food Quality and Preference*, v. 66, p. 113–126, 2018. ISSN 0950-3293. Disponível em: <https://doi.org/10.1016/j.foodqual.2018.01.014>. Acesso em: 27 de dez. de 2021.

SRIVASTAVA, A.; BALA, P. K.; KUMAR, B. Transfer learning for resolving sparsity problem in recommender systems: human values approach. *Journal of Information Systems and Technology Management (JISTEM)*, v. 14, n. 3, p. 323–337, 2017. ISSN 1807-1775. Disponível em: <https://doi.org/10.4301/s1807-17752017000300002>. Acesso em: 29 de out. de 2021.

SRIVASTAVA, A.; BALA, P. K.; KUMAR, B. New perspectives on gray sheep behavior in e-commerce recommendations. *Journal of Retailing and Consumer Services*, Elsevier, v. 53, p. 101764, 2020. ISSN 0969-6989. Disponível em: <https://doi.org/10.1016/j.jretconser.2019.02.018>. Acesso em: 29 de out. de 2021.

SUÁREZ-GARCÍA, E.; LANDIN, A.; VALCARCE, D.; BARREIRO, Á. Term Association Measures for Memory-based Recommender Systems. In: *Proceedings of the 5th Spanish Conference on Information Retrieval*. Zaragoza Spain: Association for Computing Machinery (ACM), 2018. p. 1–8. ISBN 978-1-4503-6543-7. Disponível em: <https://doi.org/10.1145/3230599.3230606>. Acesso em: 29 de out. de 2021.

SUN, J.; REN, D.; XU, D. Leveraging User Personality and Tag Information for One Class Collaborative Filtering. In: HONG, R.; CHENG, W.-H.; YAMASAKI, T.; WANG, M.; NGO, C.-W. (Ed.). *Advances in Multimedia Information Processing – PCM 2018*. Cham: Springer International Publishing, 2018. v. 11164, p. 830–840. ISBN 978-3-030-00775-1 978-3-030-00776-8. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-030-00776-8\\_76](https://doi.org/10.1007/978-3-030-00776-8_76). Acesso em: 29 de out. de 2021.

TANASESCU, V.; JONES, C. B.; COLOMBO, G.; CHORLEY, M. J.; ALLEN, S. M.; WHITAKER, R. M. The Personality of Venues: Places and the Five-Factors ('Big Five') Model of Personality. In: *2013 Fourth International Conference on Computing for Geospatial Research and Application*. San Jose, CA, USA: IEEE, 2013. p. 76–81. ISBN 978-0-7695-5012-1. Disponível em: <https://doi.org/10.1109/COMGEO.2013.12>. Acesso em: 29 de out. de 2021.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, v. 29, n. 1, p. 24–54, 2010. Disponível em: <https://doi.org/10.1177/0261927X09351676>. Acesso em: 29 de out. de 2021.

TING, T. L.; VARATHAN, K. D. JOB RECOMMENDATION USING FACEBOOK PERSONALITY SCORES. *Malaysian Journal of Computer Science*, v. 31, n. 4, p. 311–331, out. 2018. ISSN 0127-9084. Disponível em: <https://doi.org/10.22452/mjcs.vol31no4.5>. Acesso em: 29 de out. de 2021.

TKALČIČ, M. Emotions and personality in recommender systems: Tutorial. In: *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys)*. New York, NY, USA: Association for Computing Machinery (ACM), 2018. p. 535–536. ISBN 9781450359016. Disponível em: <https://doi.org/10.1145/3240323.3241619>. Acesso em: 29 de out. de 2021.

TKALČIČ, M.; CHEN, L. Personality and Recommender Systems. In: *Recommender Systems Handbook*. Boston, MA: Springer US, 2015. p. 715–739. ISBN 978-1-4899-7636-9 978-1-4899-7637-6. Disponível em: [https://doi.org/10.1007/978-1-4899-7637-6\\_21](https://doi.org/10.1007/978-1-4899-7637-6_21). Acesso em: 29 de out. de 2021.

TKALČIČ, M.; KUNAVR, M.; KOŠIR, A.; TASIČ, J. Addressing the new user problem with a personality based user similarity measure. In: RICCI, F.; SEMERARO, G.; GEMMIS, M. de; LOPS, P.; MASTHOFF, J.; GRASSO, F.; HAM, J. (Ed.). *Joint Proceedings of the Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA) and the 2nd Workshop on User Models for Motivational Systems: The affective and the rational routes to persuasion (UMMS) — Workshops at the 19th International Conference on User Modeling, Adaptation and Personalization (UMAP)*. Girona, Spain: CEUR-WS.org, 2011, (CEUR Workshop Proceedings, v. 740). p. 106–111. Disponível em: [http://ceur-ws.org/Vol-740/UMMS2011\\_paper6.pdf](http://ceur-ws.org/Vol-740/UMMS2011_paper6.pdf). Acesso em: 29 de out. de 2021.

TKALČIČ, M.; KUNAVR, M.; TASIC, J.; KOŠIR, A. Personality based user similarity measure for a collaborative recommender system. In: PETER, C.; AXELROD, E. C. an L.; AGIUS, H.; AFZAL, S.; BALAAM, M. (Ed.). *Proceedings of the 5th International Workshop on Emotion in Human-Computer Interaction - Real world challenges*. Cambridge, UK: Fraunhofer Verlag, 2009. p. 30–37. ISBN 978-3-8396-0088-7. Disponível em: [https://publica.fraunhofer.de/eprints/urn\\_nbn\\_de\\_0011-n-1134433.pdf](https://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-1134433.pdf). Acesso em: 29 de out. de 2021.

TOMMASEL, A.; CORBELLINI, A.; GODOY, D.; SCHIAFFINO, S. Exploring the role of personality traits in followee recommendation. *Online Information Review*, v. 39, n. 6, p. 812–830, out. 2015. ISSN 1468-4527. Disponível em: <https://doi.org/10.1108/OIR-04-2015-0107>. Acesso em: 29 de out. de 2021.

TOMMASEL, A.; CORBELLINI, A.; GODOY, D.; SCHIAFFINO, S. Personality-aware followee recommendation algorithms: An empirical analysis. *Engineering Applications of Artificial Intelligence*, v. 51, p. 24–36, maio 2016. ISSN 0952-1976. Disponível em: <https://doi.org/10.1016/j.engappai.2016.01.016>. Acesso em: 29 de out. de 2021.

TOMMASEL, A.; CORBELLINI, A.; GODOY, D. L.; SCHIAFFINO, S. N. On the Role of Personality Traits in Followee Recommendation Algorithms. In: *Argentine Symposium on Artificial Intelligence (ASAI) — XLIV Jornadas Argentinas de Informática e Investigación Operativa (JAIIO)*. Rosario, Argentina: SEDICI, 2015. p. 105–112. ISSN 2451-7585.

Disponível em: <http://sedici.unlp.edu.ar/handle/10915/52104>. Acesso em: 29 de out. de 2021.

TRAN, T. N. T.; FELFERNIG, A.; TINTAREV, N. Humanized Recommender Systems: State-of-the-art and Research Issues. *ACM Transactions on Interactive Intelligent Systems*, v. 11, n. 2, p. 1–41, jul. 2021. ISSN 2160-6455, 2160-6463. Disponível em: <https://doi.org/10.1145/3446906>. Acesso em: 29 de out. de 2021.

TREVISAN, L. F.; CAZELLA, S. C.; NUNES, M. A. S. N. Aplicando Traços de Personalidade e Contextos em Sistemas de Recomendação para TV Digital: um facilitador do processo de ensino-aprendizagem. *Anais do XXII SBIE - XVII WIE*, Aracaju, p. 1792–1799, 2011. ISSN 2176-4301. Disponível em: <http://almanaquesdacomputacao.com.br/gutanunes/publications/trevisan.pdf>. Acesso em: 29 de out. de 2021.

UDDIN, M. F.; BANERJEE, S.; LEE, J. Recommender System Framework for Academic Choices: Personality Based Recommendation Engine (PBRE). In: *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*. Pittsburgh, PA, USA: IEEE, 2016. p. 476–483. ISBN 978-1-5090-3207-5. Disponível em: <https://doi.org/10.1109/IRI.2016.70>. Acesso em: 29 de out. de 2021.

UFER, D.; LIN, W.; ORTEGA, D. L. Personality traits and preferences for specialty coffee: Results from a coffee shop field experiment. *Food Research International*, v. 125, p. 108504, 2019. ISSN 0963-9969. Disponível em: <https://doi.org/10.1016/j.foodres.2019.108504>. Acesso em: 27 de dez. de 2021.

VARGAS, S.; CASTELLS, P. Rank and relevance in novelty and diversity metrics for recommender systems. In: *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys)*. Chicago, Illinois, USA: ACM Press, 2011. p. 109. ISBN 978-1-4503-0683-6. Disponível em: <https://doi.org/10.1145/2043932.2043955>. Acesso em: 29 de out. de 2021.

VINCIARELLI, A.; MOHAMMADI, G. A Survey of Personality Computing. *IEEE Transactions on Affective Computing*, v. 5, n. 3, p. 273–291, jul. 2014. ISSN 1949-3045. Disponível em: <https://doi.org/10.1109/TAFFC.2014.2330816>. Acesso em: 29 de out. de 2021.

WANG, H.; ZUO, Y.; LI, H.; WU, J. Cross-domain recommendation with user personality. *Knowledge-Based Systems*, v. 213, p. 106664, fev. 2021. ISSN 0950-7051. Disponível em: <https://doi.org/10.1016/j.knosys.2020.106664>. Acesso em: 29 de out. de 2021.

WANG, J. A Collaborative Filtering Systems based on Personality Information:. In: . Xi'an, China: [s.n.], 2015. Disponível em: <https://doi.org/10.2991/iiicec-15.2015.163>. Acesso em: 29 de out. de 2021.

WANG, W.; CHEN, J.; WANG, J.; CHEN, J.; LIU, J.; GONG, Z. Trust-Enhanced Collaborative Filtering for Personalized Point of Interests Recommendation. *IEEE Transactions on Industrial Informatics*, v. 16, n. 9, p. 6124–6132, set. 2020. ISSN 1551-3203, 1941-0050. Disponível em: <https://doi.org/10.1109/TII.2019.2958696>. Acesso em: 29 de out. de 2021.

- WU, W. Implicit acquisition of user personality for augmenting recommender systems. In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion*. New York, NY, USA: Association for Computing Machinery (ACM), 2017. (IUI '17 Companion), p. 201–204. ISBN 9781450348935. Disponível em: <https://doi.org/10.1145/3030024.3038287>. Acesso em: 29 de out. de 2021.
- WU, W.; CHEN, L. Implicit acquisition of user personality for augmenting movie recommendations. In: RICCI, F.; BONTCHEVA, K.; CONLAN, O.; LAWLESS, S. (Ed.). *International Conference on User Modeling, Adaptation, and Personalization (UMAP)*. Cham: Springer International Publishing, 2015. p. 302–314. ISBN 978-3-319-20267-9. Disponível em: [https://doi.org/10.1007/978-3-319-20267-9\\_25](https://doi.org/10.1007/978-3-319-20267-9_25). Acesso em: 29 de out. de 2021.
- WU, W.; CHEN, L.; HE, L. Using personality to adjust diversity in recommender systems. In: *Proceedings of the 24th ACM Conference on Hypertext and Social Media - HT '13*. Paris, France: ACM Press, 2013. p. 225–229. ISBN 978-1-4503-1967-6. Disponível em: <https://doi.org/10.1145/2481492.2481521>. Acesso em: 29 de out. de 2021.
- WU, W.; CHEN, L.; ZHAO, Y. Personalizing recommendation diversity based on user personality. *User Modeling and User-Adapted Interaction (UMUAI)*, v. 28, n. 3, p. 237–276, ago. 2018. ISSN 0924-1868, 1573-1391. Disponível em: <https://doi.org/10.1007/s11257-018-9205-x>. Acesso em: 29 de out. de 2021.
- XIA, F.; ASABERE, N. Y.; LIU, H.; CHEN, Z.; WANG, W. Socially Aware Conference Participant Recommendation With Personality Traits. *IEEE Systems Journal*, v. 11, n. 4, p. 2255–2266, dez. 2017. ISSN 1932-8184, 1937-9234, 2373-7816. Disponível em: <https://doi.org/10.1109/JSYST.2014.2342375>. Acesso em: 29 de out. de 2021.
- XIAO, P.; FAN, Y.; DU, Y. A Personality-Aware Followee Recommendation Model Based on Text Semantics and Sentiment Analysis. In: HUANG, X.; JIANG, J.; ZHAO, D.; FENG, Y.; HONG, Y. (Ed.). *Natural Language Processing and Chinese Computing*. Cham: Springer International Publishing, 2018. v. 10619, p. 503–514. ISBN 978-3-319-73618-1. Series Title: Lecture Notes in Computer Science. Disponível em: [https://doi.org/10.1007/978-3-319-73618-1\\_42](https://doi.org/10.1007/978-3-319-73618-1_42). Acesso em: 29 de out. de 2021.
- YAKHCHI, S.; BEHESHTI, A.; GHAFARI, S. M.; ORGUN, M. A. Enabling the analysis of personality aspects in recommender systems. *CoRR*, abs/2001.04825, 2020. Disponível em: <https://arxiv.org/abs/2001.04825>. Acesso em: 29 de out. de 2021.
- YANG, H.-C.; HUANG, Z.-R. Mining personality traits from social messages for game recommender systems. *Knowledge-Based Systems*, Elsevier, v. 165, p. 157–168, fev. 2019. ISSN 0950-7051. Disponível em: <https://doi.org/10.1016/j.knosys.2018.11.025>. Acesso em: 29 de out. de 2021.
- YANG, H.-C.; LEE, C.-H.; YEH, C.-Y. Mining personality traits from social text messages. In: *Proceedings of the 7th Multidisciplinary in International Social Networks Conference and The 3rd International Conference on Economics, Management and Technology*. New York, NY, USA: Association for Computing Machinery (ACM), 2020. p. 1–5. ISBN 9781450389457. Disponível em: <https://doi.org/10.1145/3429395.3429412>. Acesso em: 29 de out. de 2021.

YANG, H.-C.; LIN, C. S.; HUANG, Z.-R.; TSAI, T.-H. Text Mining on Player Personality for Game Recommendation. In: *Proceedings of the 4th Multidisciplinary International Social Networks Conference on ZZZ - MISNC '17*. Bangkok, Thailand: ACM Press, 2017. p. 1–6. ISBN 978-1-4503-4881-2. Disponível em: <https://doi.org/10.1145/3092090.3092132>. Acesso em: 29 de out. de 2021.

YI, M.-Y.; LEE, O.-J.; JUNG, J. J. MBTI-Based Collaborative Recommendation System: A Case Study of Webtoon Contents. In: VINH, P. C.; ALAGAR, V. (Ed.). *International Conference on Context-Aware Systems and Applications (ICCASA)*. Cham: Springer International Publishing, 2016. p. 101–110. ISBN 978-3-319-29236-6. Disponível em: [https://doi.org/10.1007/978-3-319-29236-6\\_11](https://doi.org/10.1007/978-3-319-29236-6_11). Acesso em: 29 de out. de 2021.

ZHANG, Z.; ZHAO, X.; WANG, G.; BI, X. A New Point-of-Interest Classification Model with an Extreme Learning Machine. *Cognitive Computation*, v. 10, n. 6, p. 951–964, dez. 2018. ISSN 1866-9956, 1866-9964. Disponível em: <https://doi.org/10.1007/s12559-018-9599-0>. Acesso em: 29 de out. de 2021.

ZHENG, Y.; SUBRAMANIYAN, A. Personality-Aware Collaborative Learning: Models and Explanations. In: BAROLLI, L.; TAKIZAWA, M.; XHAFI, F.; ENOKIDO, T. (Ed.). *Advanced Information Networking and Applications*. Cham: Springer International Publishing, 2020. v. 926, p. 631–642. ISBN 978-3-030-15031-0 978-3-030-15032-7. Series Title: Advances in Intelligent Systems and Computing. Disponível em: [https://doi.org/10.1007/978-3-030-15032-7\\_53](https://doi.org/10.1007/978-3-030-15032-7_53). Acesso em: 29 de out. de 2021.

ZHOU, J.-L.; FU, Y.; LU, H.; SUN, C.-J. From Popularity to Personality — A Heuristic Music Recommendation Method for Niche Market. *Journal of Computer Science and Technology*, v. 26, n. 5, p. 816–822, set. 2011. ISSN 1000-9000, 1860-4749. Disponível em: <https://doi.org/10.1007/s11390-011-0180-5>. Acesso em: 29 de out. de 2021.

ZHOU, T.; KUSCSIK, Z.; LIU, J.-G.; MEDO, M.; WAKELING, J. R.; ZHANG, Y.-C. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, National Academy of Sciences, v. 107, n. 10, p. 4511–4515, 2010. ISSN 0027-8424. Disponível em: <https://doi.org/10.1073/pnas.1000488107>. Acesso em: 29 de out. de 2021.

# Apêndice A

## Publicações

Ao longo da realização da pesquisa de doutorado descrita neste documento, algumas contribuições foram divulgadas por meio de publicações em periódicos e em conferências. Neste apêndice, encontram-se listados os artigos publicados.

1. AGUIAR, J. J. B. Considerando Estilos de Aprendizagem, Emoções e Personalidade em Informática na Educação. *Informática na Educação: Teoria & Prática*, v. 20, n. 2 mai/ago, p. 85–102, set. 2017. ISSN 1982-1654, 1516-084X. Disponível em: <https://doi.org/10.22456/1982-1654.65333>. Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado no periódico científico editado pelo programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (UFRGS).
- **Resumo:** Muitas vezes tratam-se turmas homogeneamente e as TDIC (Tecnologias Digitais de Informação e Comunicação) são utilizadas pelos alunos desconsiderando que eles possuem particularidades (tais como: estilos de aprendizagem, emoções e traços de personalidade) que influenciam em sua aquisição de conhecimento. Neste trabalho, são apresentados conceitos e reflexões visando incentivar discussões e estudos teóricos ou empíricos que considerem essas particularidades dos alunos durante o processo de ensino e aprendizagem. É apresentada também uma breve análise de dados de alunos de computação, em diferentes níveis de ensino, os quais responderam a questionários visando identificar perfis referentes a seus estilos de aprendizagem e a seus traços de personalidade. A análise corrobora com a ideia da heterogeneidade de perfis, mesmo considerando alunos da mesma área/nível de ensino. Percebe-se que ainda há várias possibilidades de pesquisas sobre aliar as TDIC e as particularidades de alunos objetivando a melhoria do processo de ensino e aprendizagem.

2. AGUIAR, J.; BARBOSA, A.; FECHINE, J.; COSTA, E. Um Estudo sobre a Influência das Dimensões do Modelo Felder-Silverman na Recomendação de Recursos Educacionais baseada nos Estilos de Aprendizagem dos Alunos. In: *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*. Recife-PE: SBC (Sociedade Brasileira de Computação), 2017. v. 28, p. 1277-1286. Disponível em: <https://doi.org/10.5753/cbie.sbie.2017.1277>. Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado nos anais do SBIE 2017 — evento do VI Congresso Brasileiro de Informática na Educação (CBIE 2017) — e apresentado no dia 01 de novembro de 2017, na Sessão Técnica 7 do SBIE/CBIE, realizado na Universidade Federal de Pernambuco (UFPE).
- **Resumo:** A teoria dos Estilos de Aprendizagem (EA) pode ser utilizada para construir o perfil dos usuários em Sistemas de Recomendação Educacionais (SRE). Neste artigo é apresentado um estudo sobre a influência das quatro dimensões do modelo de EA de Felder-Silverman (FSLSM) na construção de perfis de usuários de SRE. Para isso, analisou-se a possibilidade de utilizar apenas alguma(s) dimensão(ões) para simplificar esta construção, mantendo, ainda, as recomendações personalizadas acuradas. Os resultados apontaram que as dimensões do FSLSM não têm a mesma influência na recomendação e que, embora não tenha sido possível destacar precisamente aquelas de maior influência, há indícios de que o uso de metade das dimensões seja suficiente.

3. AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Estratégia baseada em Personalidade e Tendências para Recomendação de Objetos de Aprendizagem usando Algoritmo Genético. In: *I Workshop Latino-Americano de Trabalhos em Andamento em Computação (WLATAC) - Anais*. 1. ed. São Paulo-SP: SBC (Sociedade Brasileira de Computação), 2018. v. 1, p. 62-67. ISBN 978-85-7669-460-1. Disponível em: [http://portaldeconteudo.sbc.org.br/index.php/wlatac\\_clei/issue/view/282/wlatac\\_clei](http://portaldeconteudo.sbc.org.br/index.php/wlatac_clei/issue/view/282/wlatac_clei). Acesso em: 29 de out. de 2021.

- **Informações complementares:** Capítulo publicado no livro referente aos anais do I Workshop Latino-Americano de Trabalhos em Andamento em Computação (WLATAC 2018), evento satélite do CLEI-LACLO 2018 (XLIV Conferência Latino-americana de Informática & XIII Conferência Latino-americana de Tecnologias de Aprendizagem), realizado durante os dias 01 a 05 de outubro de 2018.
- **Resumo:** Neste capítulo é apresentada uma pesquisa que propõe um modelo para recomendação de Objetos de Aprendizagem, usando estratégias de recomendação baseadas em conceitos de Personalidade (Traços de Personalidade e Estilos de Aprendizagem) e Filtragem Colaborativa baseada em Tendências. Tais estratégias foram combinadas utilizando Algoritmo Genético. Os resultados do estágio atual da pesquisa indicam que, com a estratégia híbrida proposta, são obtidas

recomendações mais acuradas em comparação aos algoritmos que o compõem isoladamente, além de recomendações mais acuradas que as providas por outras abordagens tradicionais de recomendação.

4. AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Recomendação de Objetos de Aprendizagem utilizando Filtragem Colaborativa baseada em Tendências e em Estilos de Aprendizagem. In: *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*. Fortaleza–CE: SBC (Sociedade Brasileira de Computação), 2018. v. 29, p. 1423–1432. Disponível em: <https://doi.org/10.5753/cbie.sbie.2018.1423>. Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado nos anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018) — evento que integra o VII Congresso Brasileiro de Informática na Educação (CBIE 2018) — e apresentado no dia 01 de novembro de 2018, na Sessão Técnica 34 do SBIE/CBIE, realizado durante os dias 29 de outubro a 01 de novembro de 2018.
- **Resumo:** Há pesquisas que indicam a importância de considerar particularidades, relativas ao processo de aprendizagem dos estudantes, ao elaborar estratégias para Sistemas de Recomendação de recursos educacionais. Nesta pesquisa, é proposta uma estratégia de recomendação personalizada referente a Objetos de Aprendizagem, com base nos Estilos de Aprendizagem dos estudantes e na Filtragem Colaborativa baseada em Tendências — combinadas por meio de um Algoritmo Genético. Uma avaliação experimental (utilizando um conjunto de dados referentes a estudantes de Computação) indicou que a estratégia proposta proporcionou resultados melhores em comparação a outras abordagens de recomendação.

5. AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Experimentando a Influência dos Traços de Personalidade do Modelo Big Five na Recomendação de Recursos Educacionais. In: *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)*. Brasília–DF: SBC (Sociedade Brasileira de Computação), 2019. v. 30, p. 1711–1720. Disponível em: <https://doi.org/10.5753/cbie.sbie.2019.1711>. Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado nos anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019) — evento que integra o VIII Congresso Brasileiro de Informática na Educação (CBIE 2019) — e apresentado no dia 11 de novembro de 2019, na Sessão Técnica 20 do SBIE/CBIE, realizado durante os dias 11 a 14 de novembro de 2019.
- **Resumo:** Pesquisas recentes têm usado a teoria dos Traços de Personalidade para construir o perfil dos usuários em Sistemas de Recomendação. Neste artigo, é apresentado um estudo experimental sobre a influência dos traços do modelo *Big Five* na construção de perfis de usuários de Sistemas de Recomendação Educacionais. Particularmente, analisou-se a possibilidade de utilizar

apenas algum(ns) traço(s) para simplificar esta construção, mantendo, ainda, as recomendações personalizadas acuradas. Embora não tenha sido possível destacar precisamente os traços do modelo *Big Five* com maior influência, os resultados apontaram que os cinco traços não têm a mesma influência na re-comendação, havendo indícios de que o uso do traço Abertura seja suficiente.

6. AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Estudo de Desempenho de Algoritmos de Filtragem Colaborativa para Sistemas de Recomendação Educacionais aplicando Agrupamento Prévio de Usuários com Traços de Personalidade Similares. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)*. Natal–RN: SBC (Sociedade Brasileira de Computação), 2020. v. 31, p. 1082–1091. Disponível em: <https://doi.org/10.5753/cbie.sbie.2020.1082>. Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado nos anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020) — evento que integra o IX Congresso Brasileiro de Informática na Educação (CBIE 2020) — e apresentado no dia 25 de novembro de 2020, na Sessão Técnica ‘Lagoa de Aratuba’ do SBIE/CBIE, realizado, no formato on-line, durante os dias 24 a 28 de novembro de 2020.
- **Resumo:** Pesquisas têm usado a teoria dos Traços de Personalidade (TP) ao aplicar Filtragem Colaborativa (FC) em Sistemas de Recomendação Educacionais. A ideia levantada nesta pesquisa é que todos os algoritmos de FC lidam, implicitamente, com os TP dos usuários, e, ao agrupar explicitamente os usuários, baseando-se nesses TP, o desempenho dos algoritmos pode melhorar. Neste artigo, é apresentado um estudo experimental sobre a acurácia de algoritmos de FC aplicando um agrupamento prévio dos usuários (maioritariamente discentes de computação), baseando-se em seus TP (modelo *Big Five*). Os resultados indicaram que, com tal estratégia, no domínio educacional, há a possibilidade de melhoria da acurácia de algoritmos de FC baseados em vizinhança.

7. AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Collaborative Filtering Strategy for Product Recommendation Using Personality Characteristics of Customers. In: *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia 2020)*. São Luís–MA (Brazil): Association for Computing Machinery (ACM), 2020. p. 157–164. ISBN 978-1-4503-8196-3. Disponível em: <https://doi.org/10.1145/3428658.3430969>. Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado nos anais do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2020) e apresentado no dia 02 de dezembro de 2020, na Sessão Técnica 6 do WebMedia, realizado, no formato on-line, durante os dias 30 de novembro a 4 de dezembro de 2020.

- **Resumo:** *Research indicates that people can receive more useful product recommendations if the filtering process considers their personality. In this paper, we propose a hybrid strategy for Recommender Systems (using matrix factorization and personality-based neighborhood) to recommend the best products calculated for a particular customer (user). The proposed user profile used in the definition of the neighborhood involves these three personality models: Big Five (or OCEAN, or Five-Factor Model), Needs, and Values. We experimented with data from more than 10,000 Amazon customers. We inferred their personality characteristics from the analysis of reviews via IBM Watson Personality Insights. The results indicated that the proposed strategy's performance was better than that of the state-of-the-art algorithms analyzed. Besides, there was no statistical difference between using only the Big Five model or using it together with the Needs and Values models.*

8. AGUIAR, J. J. B.; ARAÚJO, J. M. F. R.; COSTA, E. d. B. Estudo Comparativo de Abordagens para Sistemas de Recomendação baseados em Personalidade com uso do serviço IBM Watson Personality Insights. *Revista Ibérica de Sistemas e Tecnologias de Informação (RISTI)*, Porto (Portugal), v. 40, p. 73–88, 2020. ISSN 1646-9895. Disponível em: (<https://doi.org/10.17013/risti.40.73-88>). Acesso em: 29 de out. de 2021.

- **Informações complementares:** Artigo publicado, em dezembro de 2020, na 40ª edição da Revista Ibérica de Sistemas e Tecnologias de Informação (RISTI).
- **Resumo:** Considerar a personalidade dos usuários nos Sistemas de Recomendação pode proporcionar resultados mais relevantes. Neste estudo, foi analisado se, com os avanços na detecção da personalidade (inferência sem usar questionários), as abordagens de Filtragem Colaborativa baseadas em personalidade continuam melhorando a acurácia das abordagens tradicionais (baseadas essencialmente em *ratings*). Outrossim, foi analisado se há diferenças ao aplicar modelos distintos de personalidade. No experimento, foram considerados 1058 usuários do TripAdvisor e 10889 clientes da Amazon, com características de personalidade inferidas via *IBM Watson Personality Insights*. Os resultados indicaram a possibilidade de melhorar a acurácia ao empregar uma abordagem usando dados inferidos concernentes a algum dos modelos de personalidade analisados (*Big Five*, *Needs* e *Values*). Ademais, o modelo *Values* proporcionou resultados equivalentes ao *Big Five* (sem facetas); e, em termos gerais, não houve melhoria ao usar o *Big Five* incluindo dados das suas facetas (nem ao incluir dados dos outros modelos).

# **Apêndice B**

## **Slides**

Nas seguintes páginas deste apêndice, encontram-se os slides utilizados durante a sessão pública de defesa desta tese de doutorado (realizada no dia 17 de dezembro de 2021, às 8 horas e 30 minutos).

## Características de Personalidade Identificadas via Textos aplicadas à Recomendação por Filtragem Colaborativa

Janderson Jason Barbosa Aguiar

Orientadores:

Dr.<sup>a</sup> Joseana Macêdo Fachine Régis de Araújo  
Dr. Evandro de Barros Costa

Campina Grande  
Dezembro – 2021

### Roteiro

- Motivação
- Objetivos
  - Objetivo Geral
  - Objetivos Específicos
- Pesquisas Relacionadas
- Abordagens Elaboradas
- Investigação Experimental
- Considerações Finais

### Motivação (1)

- Sistemas de Recomendação (SR)
  - Apoio ao acesso de conteúdos mais adequados ao perfil de cada indivíduo
    - Filtragem Colaborativa (FC)
    - Filtragem Baseada em Conteúdo (FBC)
    - Filtragem Híbrida
    - ...



“Como **melhorar as recomendações** para oferecer informações, produtos ou serviços **mais personalizados** para os usuários?”

## Motivação (2)

- *Personality Computing*
  - Variados modelos e características de personalidade
    - Questionários
    - *Automatic Personality Recognition (APR)*
      - APR baseado em texto

A personalidade pode influenciar o processo de tomada de decisão

Likes/Ratings influenciados pela personalidade

## Motivação (3)

- Recomendação baseada em Personalidade

Interesses e padrões comportamentais são similares entre pessoas com personalidade similar

Indivíduos similares quanto à personalidade atribuindo likes/ratings similares ao avaliarem os mesmos itens

Algoritmos de FC influenciados pela personalidade

Aspecto motivador para realizar uma investigação com foco na acurácia

## Motivação (4)

- Recomendação baseada em Personalidade

De que forma o conhecimento das **características de personalidade** dos usuários de SR, identificadas automaticamente **via textos**, pode contribuir **para melhorar**, em cenários gerais, a acurácia dos algoritmos de **recomendação**?

## Objetivos (1)

- **Objetivo Geral**
  - Investigar estratégias para a **melhoria da acurácia dos algoritmos de FC**, enfocando na “colaboração” (dados de opinião) dos usuários similares quanto à **personalidade**
    - Identificada via **APR baseado em texto**

## Objetivos (2)

- **Objetivo Específico O1**
  - Rever pesquisas iniciais de referência em Recomendação baseada em Personalidade
- **Objetivo Específico O2**
  - Analisar a acurácia das predições de um algoritmo de FC ao desconsiderar a “opinião” de usuários menos similares quanto à personalidade de um usuário-alvo de recomendação

## Objetivos (3)

- **Objetivo Específico O3**
  - Analisar a influência dos traços de personalidade (de maneira individual, agrupada e ponderada)
- **Objetivo Específico O4**
  - Propor e avaliar estratégias de recomendação baseadas nas características de personalidade dos usuários e suas preferências anteriores

## Pesquisas Relacionadas

Pesquisa	Modelo de personalidade	Identificação da personalidade	Método de recomendação	Domínio
Nunes (2008)	<i>Big Five</i>	NEO-IPIP; TIPI	<i>Personality neighborhood</i>	Não específico
Tkalčič et al. (2009)	<i>Big Five</i>	IPIP 50-version	<i>Personality neighborhood</i>	Não específico
Hu e Pu (2011)	<i>Big Five</i>	TIPI	<i>Personality neighborhood</i>	Não específico
Roshchina, Cardiff e Rosso (2015)	<i>Big Five</i>	APR (MAIRESSE et al., 2007)	<i>Personality neighborhood</i>	Turismo
Srivastava, Bala e Kumar (2017)	<i>Values</i>	APR ( <i>Watson Personality Insights</i> )	<i>Personality neighborhood</i>	Não específico
Srivastava, Bala e Kumar (2020)	<i>Big Five; Values; Needs</i>	APR ( <i>Watson Personality Insights</i> )	<i>Personality neighborhood</i>	Produtos
Dhelim et al. (2021)	<i>Big Five</i>	TIPI	<i>Hybrid personality filtering</i>	Artigos de notícias
ESTA PESQUISA	<i>Big Five; Values; Needs</i>	APR ( <i>Watson Personality Insights</i> )	<i>Personality neighborhood; Hybrid personality filtering</i>	Não específico

Defesa de Tese de Doutorado — Janderson Aguiar

10

## Abordagens Elaboradas (1)

### • Abordagens

#### – Referentes ao objetivo **O1**

- Variações de FC, baseadas nas pesquisas de Tkalčič et al. (2009) e Hu e Pu (2011), considerando os modelos *Big Five* (com e sem facetas), *Needs* e *Values*

#### – Referentes ao objetivo **O2**

- Abordagem "tps": *clustering* (com *k-means*) baseando-se em características de personalidade para algoritmos de FC (UserKNN, ItemKNN, SMF, BMF e SVD++)

#### – Referentes ao objetivo **O3**

- Abordagens de FC baseada em personalidade variando os traços do *Big Five* (OCEAN)
  - Exemplos: 020-060-000-020-000, 010-010-040-020-020, ...
- Variações considerando características dos três modelos
  - Algoritmo Genético para considerar as 52 características

Defesa de Tese de Doutorado — Janderson Aguiar

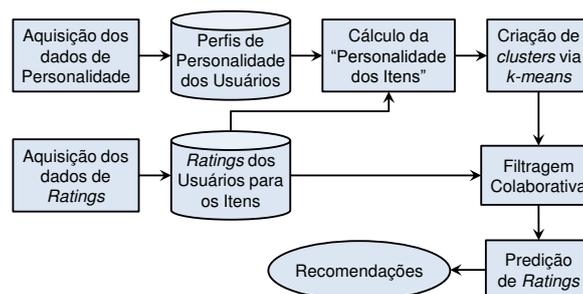
11

## Abordagens Elaboradas (2)

### • Abordagens

#### – Referentes ao objetivo **O4**

- Estratégia "itps"

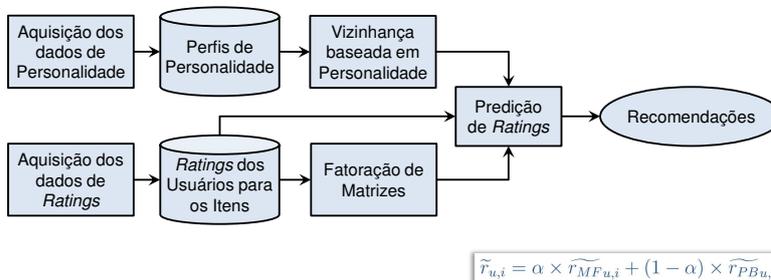


Defesa de Tese de Doutorado — Janderson Aguiar

12

## Abordagens Elaboradas (3)

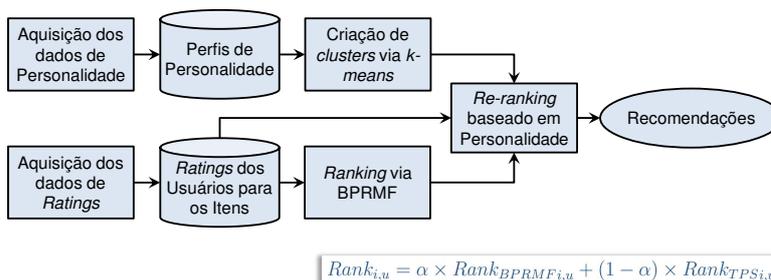
- Abordagens
  - Referentes ao objetivo **O4**
    - Estratégia **weighted**



$$\tilde{r}_{u,i} = \alpha \times \tilde{r}_{MFu,i} + (1 - \alpha) \times \tilde{r}_{PBu,i}$$

## Abordagens Elaboradas (4)

- Abordagens
  - Referentes ao objetivo **O4**
    - Estratégia **re-ranking**



$$Rank_{i,u} = \alpha \times Rank_{BPRMF_{i,u}} + (1 - \alpha) \times Rank_{TPS_{i,u}}$$

## Investigação Experimental (1)

- Características gerais dos experimentos
  - Conjuntos de dados:



Conjunto de dados	Quantidade de itens	Quantidade de usuários	Quantidade de ratings	Média de ratings por item	Média de ratings por usuário
TripAdvisor	192	641	1.250	6,51	1,95
TripAdvisorCompleto	26.844	1.058	31.645	1,18	29,91
AmazonFoods	4.306	10.889	106.371	24,70	9,77
AmazonFoodsCompleto	29.962	11.765	145.660	4,86	12,38

## Investigação Experimental (2)

- Características gerais dos experimentos
  - *Random Subsampling*
    - 2/3 dos dados para treinamento e 1/3 dos dados para teste
    - Dados divididos aleatoriamente por 50 vezes
  - Principais métricas: RMSE e *F-Measure*
  - Testes estatísticos ( $\alpha=5\%$ )
    - Testes paramétricos: ANOVA e T
    - Testes não paramétricos: Kruskal-Wallis e Mann-Whitney U

## Investigação Experimental (3)

- Principais Resultados
  - Referentes ao objetivo **O1**

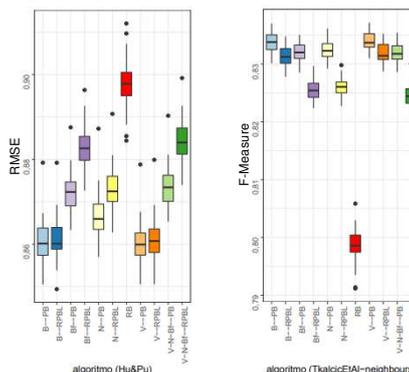
**QP1: Os resultados obtidos pelos primeiros estudos relevantes que propuseram estratégias de recomendação via FC baseadas em traços de personalidade (identificados via questionários) são similares aos obtidos utilizando-se APR e conjuntos de dados maiores?**

**Abordagens *personality-based* proporcionaram, no mínimo, acurácia similar às abordagens *rating-based***

## Investigação Experimental (4)

- Principais Resultados
  - Referentes ao objetivo **O1**

**Exemplo de cenário com maior acurácia das abordagens *personality-based* (a abordagem usando o modelo *Big Five* sem facetas não tem desempenho inferior às outras abordagens)**



## Investigação Experimental (5)

- Principais Resultados
  - Referentes ao objetivo **O2**

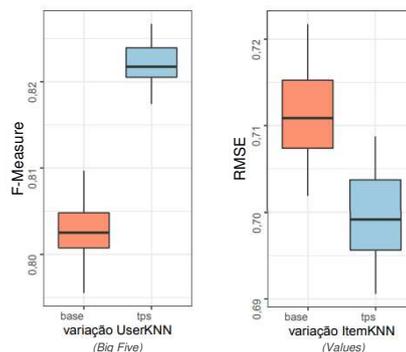
**QP2: A acurácia das predições de um algoritmo de FC é afetada se o algoritmo for executado focando na opinião de usuários similares quanto à personalidade (identificada via APR)?**

**É possível haver efeito na acurácia das predições de um algoritmo de FC ao desconsiderar a opinião de usuários menos similares a um usuário-alvo quanto à personalidade**

## Investigação Experimental (6)

- Principais Resultados
  - Referentes ao objetivo **O2**

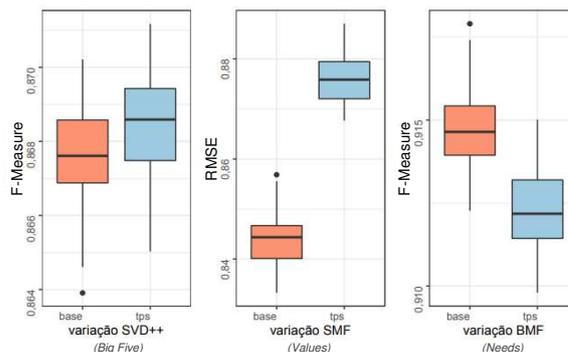
**Efeito geralmente positivo na acurácia dos algoritmos baseados em vizinhança**



## Investigação Experimental (7)

- Principais Resultados
  - Referentes ao objetivo **O2**

**Efeito negativo na acurácia em alguns cenários com algoritmos baseados em modelo**



## Investigação Experimental (8)

- Principais Resultados
  - Referentes ao objetivo **O3**

**QP3: Os traços de personalidade (identificados via APR) influenciam de maneira diferente no resultado da recomendação? É possível obter melhores resultados ao ignorar algum(ns) traço(s) ou ao ponderá-los?**

**Os traços do *Big Five* podem influenciar de maneira diferente na acurácia dos algoritmos de FC baseada em personalidade**

## Investigação Experimental (9)

- Principais Resultados
  - Referentes ao objetivo **O3**

**As variações ignorando traços ou ponderando-os proporcionaram geralmente redução de desempenho quando comparadas à variação tradicional (020-020-020-020)**

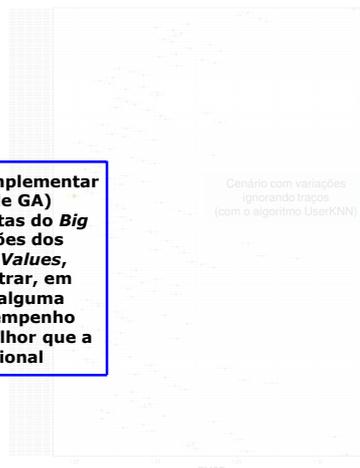


## Investigação Experimental (10)

- Principais Resultados
  - Referentes ao objetivo **O3**

**As variações ignorando traços ou ponderando-os proporcionaram geralmente redução de desempenho quando comparadas à variação tradicional (020-020-020-020)**

**OBS.: em análise complementar (incluindo uso de GA) abrangendo as facetas do *Big Five* e as dimensões dos modelos *Needs* e *Values*, foi possível encontrar, em alguns cenários, alguma variação com desempenho estatisticamente melhor que a variação tradicional**



## Investigação Experimental (11)

- Principais Resultados
  - Referentes ao objetivo **O4**

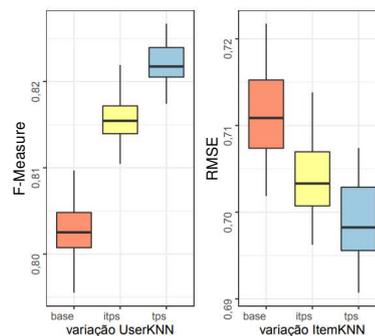
QP4: Baseando-se nas análises referentes às questões de pesquisa QP1, QP2 e QP3, é possível incorporar estratégias aos algoritmos de FC (utilizando explicitamente dados de personalidade) e obter recomendações mais relevantes?

É possível obter efeitos positivos na acurácia de algoritmos de FC ao aplicar as estratégias propostas (com personalidade identificada via APR baseado em texto)

## Investigação Experimental (12)

- Principais Resultados
  - Referentes ao objetivo **O4**

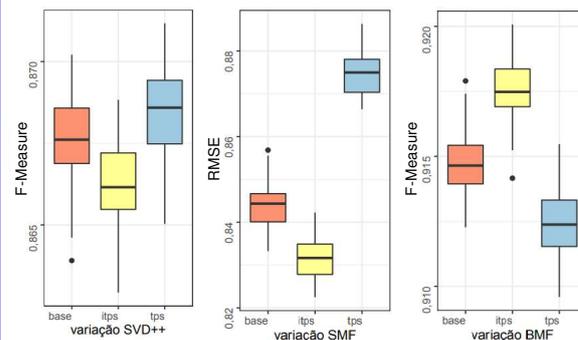
A estratégia "itps" proporcionou, nos algoritmos baseados em vizinhança, resultados geralmente melhores (no mínimo similares) em comparação ao algoritmo base, mas por vezes não superou "tps"



## Investigação Experimental (13)

- Principais Resultados
  - Referentes ao objetivo **O4**

Nos algoritmos baseados em modelo, por vezes foram observados casos de efeito positivo de "itps" em relação a SMF e BMF (diferentemente de "tps", com efeito negativo) e foram observados casos de efeito negativo de "itps" em relação a SVD++ (enquanto "tps" proporcionou resultados no mínimo similares ao algoritmo base)





## Considerações Finais (1)

- Conclusões
  - É pertinente considerar **características de personalidade** (identificadas via APR baseado em texto) em abordagens convencionais de FC
    - Possibilidade de efeitos positivos na acurácia (aliados a ganhos esperados em outros aspectos, como novidade)
      - O estado-da-arte, comumente usando o *Big Five* e questionários, destaca o uso de personalidade para obter melhorias em relação à **diversidade das recomendações**, em relação à **acurácia em cenários específicos** de SR (como situações de *cold-start*) e em relação à **personality matching** (referente a categorias de domínios específicos)
  - As informações apresentadas na tese podem **aprimorar pesquisas** em SR
  - Publicações em periódicos e conferências
    - IETP, SBIE/CBIE, WLATAC/CLEI, WEBMEDIA, RISTI

## Considerações Finais (2)

- Limitações
  - Necessidade de avanços em APR baseado em texto
    - Escassez de ferramentas para português do Brasil considerando além dos traços principais do *Big Five*
    - Serviço *Watson Personality Insights* em processo de finalização pela IBM
  - Ameaças à validade referentes aos conjuntos de dados utilizados
    - Usuários com menos de 600 palavras desconsiderados
  - ...

## Considerações Finais (3)

- Pesquisas Futuras
  - Utilizar outros conjuntos de dados
  - Estudar novos fatores e opções específicas dos algoritmos de FC
  - Investigar além de FC
  - Aplicar recomendação *cross-domain*
  - Analisar além da acurácia
    - Novidade
    - Diversidade
    - Serendipidade
    - Explicabilidade
  - ...

## Apêndice C

### Exemplo referente a APR baseado em textos de um usuário

Neste apêndice, é apresentado um exemplo referente ao reconhecimento automático de personalidade via textos, utilizado na investigação experimental desta tese. No Quadro C.1, são apresentados os textos (em língua inglesa) dos *reviews* de determinado usuário do conjunto de dados da Amazon (*Fine Foods*). A partir desses textos, foram obtidos, via *Watson Personality Insights*, os seguintes valores representativos da personalidade de tal usuário (segundo o formalismo da Equação 4.1 — Capítulo 4).

- |                    |                       |                       |                       |
|--------------------|-----------------------|-----------------------|-----------------------|
| • $t_1 = 84, 34\%$ | • $f_9 = 66, 59\%$    | • $f_{22} = 68, 43\%$ | • $v_5 = 82, 60\%$    |
| • $t_2 = 64, 60\%$ | • $f_{10} = 53, 89\%$ | • $f_{23} = 72, 48\%$ | • $n_1 = 60, 16\%$    |
| • $t_3 = 50, 40\%$ | • $f_{11} = 56, 19\%$ | • $f_{24} = 60, 47\%$ | • $n_2 = 75, 10\%$    |
| • $t_4 = 65, 74\%$ | • $f_{12} = 74, 25\%$ | • $f_{25} = 52, 80\%$ | • $n_3 = 83, 51\%$    |
| • $t_5 = 52, 13\%$ | • $f_{13} = 54, 84\%$ | • $f_{26} = 52, 44\%$ | • $n_4 = 50, 03\%$    |
| • $f_1 = 49, 84\%$ | • $f_{14} = 59, 16\%$ | • $f_{27} = 47, 12\%$ | • $n_5 = 80, 19\%$    |
| • $f_2 = 79, 03\%$ | • $f_{15} = 55, 54\%$ | • $f_{28} = 55, 66\%$ | • $n_6 = 59, 10\%$    |
| • $f_3 = 65, 02\%$ | • $f_{16} = 53, 47\%$ | • $f_{29} = 59, 91\%$ | • $n_7 = 65, 23\%$    |
| • $f_4 = 71, 07\%$ | • $f_{17} = 48, 16\%$ | • $f_{30} = 40, 80\%$ | • $n_8 = 71, 46\%$    |
| • $f_5 = 74, 48\%$ | • $f_{18} = 33, 66\%$ | • $v_1 = 56, 97\%$    | • $n_9 = 69, 08\%$    |
| • $f_6 = 66, 10\%$ | • $f_{19} = 69, 76\%$ | • $v_2 = 75, 96\%$    | • $n_{10} = 61, 49\%$ |
| • $f_7 = 68, 14\%$ | • $f_{20} = 68, 96\%$ | • $v_3 = 66, 70\%$    | • $n_{11} = 67, 22\%$ |
| • $f_8 = 60, 40\%$ | • $f_{21} = 45, 30\%$ | • $v_4 = 62, 24\%$    | • $n_{12} = 70, 08\%$ |

Quadro C.1: Textos dos reviews de um usuário Amazon (A11LNPG39A2ZV4).

<p><i>Flat coffee taste, artificial vanilla, blah. Of all the pods I've tried in my Senseo, these are the worst.</i></p>
<p><i>I brought these Newtons Fruit Thins Fig and Honey into my office to share with my colleagues during an afternoon break. The first thing we noticed is the lovely cinnamon-and-cookie smell when you open the package. These sweet mild cookies with tiny bits of fig call out for a cold glass of milk or a good cup of tea. I prefer the older fruity-Newtons, but these are a really good alternative for crunchy cookie fans, mixed cookie plates, or those looking for a less fruity Newton that still has plenty of sweet, cinnamon, cookie goodness.</i></p>
<p><i>There's more about Bob Marley on these coffee packets than brewing instructions. There are quotes, inspiration, attitude, very nice rendering of a lion, and the man himself holding his impressive locks. But machine compatibility? Light/med/dark roast? Recommended cup size. Ah, no. I found that one pod fit the 2-pod holder of my Senseo. Two Marley pods were too large for the 2-pod holder of the Senseo. I brewed two separate cups, using two pods this morning, but there was no caffeine jolt in this coffee. A single pod gives a delicious, but slightly weak cup. Two pods give a very "Starbucks-y" cup of coffee with strong flavor, but without their jolt of caffeine. In fact, I needed a third cup of Marley coffee to accomplish what a perfectly ordinary spoonful of Maxwell House Instant Coffee can do any morning of the week when I am in a hurry (but at a cost of around two dollars). If these folks are serious about coffee, they need to take a look at their competition; more flavor and data, less music. Unless you're running a café.</i></p>
<p><i>Using my Aerobie 80R08 AeroPress Coffee and Espresso Maker, this Starbucks vanilla coffee made such a fabulous cup of coffee! The coffee is a light roast with no bitterness (especially when made in the AeroPress), and the vanilla notes are definitely present, but without seeming artificial or overwhelming. My friend drank it plain black, and thought it was a great cup of coffee. I make mine with a double-strength coffee base, warmed milk, and raw sugar, a recipe for a little bit of caffeinated heaven. I don't buy flavored coffees very often, but I'll definitely be putting this one into my regular rotation.</i></p>
<p><i>IZZE is soooooo good! It's a delicious beverage, with subtle genuine fruit flavor, just enough natural sweetness, and moderate carbonation. Sipped from the can, or poured over ice, it's very refreshing, and not at all cloying. On the afternoon that my order of IZZE arrived from Amazon, a 16-year old family friend stopped by. I offered him a can, and he said, "I've had it before, and it's great!" I'd quite happily serve this at summer parties as an alternative to traditional sodas, wine, or even lemonade. Since it doesn't contain refined sugars, stimulants such as caffeine, or artificial sweeteners, it should appeal to my guests who are avoiding those ingredients. ETA: I've since tried the Sparkling Blackberry, and it is even more delicious, and a gorgeous rosy shade that looks beautiful in a glass of ice.</i></p>
<p><i>This coffee smells great first thing in the morning, if you occasionally enjoy a flavored coffee. The chocolate is a rich addition to the good, medium-bodied coffee, with plenty of morning caffeine. I enjoy my coffee with plenty of hot milk and a bit of raw sugar (like a latte), made with my AeroPress Coffee and Espresso Maker. This made a fantastic cup of coffee with just the right touch of chocolate, and no artificial flavor notes. I brewed the coffee the way I normally do with two huge Aeropress scoops of coffee, and water at 175 degrees for 30 seconds; then I lowered the plunger. Mmmmmmm. A perfect cup of extra-good latte, and a shot of chocolate flavoring.</i></p>
<p><i>Oh bliss, mmm, oh, mmm, oh, excuse me! Was I talking with my mouth full? I have received a package of Scharffen Berger Semisweet Chocolate in the mail to review, and it is both my favorite type of eating and baking chocolate (of most brands). The Scharffen Berger does not disappoint. It bakes beautifully when chopped into chunks and used in my go-to cookie and scone recipes. The chocolate has lovely gooey melty flops when you bite into the chunks (I'm eating a cookie right now), and the chocolate is a glorious, subtle, hits-you-slowly essence. It's the sophisticated version of Mom's cookies. In scones, the chocolate is perfect with raspberry jam. I'd happily make my Death By Chocolate Cake - or any other pure chocolate recipe - with these stellar bars. They really have an after dinner quality to them. So why am I eating the other half of a leftover bar in the afternoon? Well, as I said, Semisweet is my favorite kind of chocolate and Scharffen Berger does not disappoint! A perfect treat with a cup of coffee.</i></p>

Fonte: autoria própria — baseando-se no conjunto de dados AmazonFoods.

# Apêndice D

## Resultados de testes estatísticos

Neste apêndice, são apresentados os resultados dos testes estatísticos referentes aos experimentos apresentados no Capítulo 5. Nas seguintes tabelas, os valores destacados em amarelo são maiores que o nível de significância 0,05, indicando não ser possível rejeitar a hipótese nula em questão.

Na Tabela D.1, os destaques em amarelo ilustram que, com os dados oriundos do TripAdvisor, não houve diferença estatística (para ambas as métricas) em relação às variações (abordagem *rating-based* e abordagens *personality-based*). Com os dados oriundos da Amazon, houve diferença estatística.

Nas Tabelas D.2, D.3 e D.4, apresenta-se que, com os dados oriundos do TripAdvisor, houve diferença estatística em poucos cenários (referentes à métrica RMSE e aos algoritmos UserKNN e ItemKNN; além do algoritmo SMF, especificamente no TripAdvisorCompleto). Com os dados oriundos da Amazon, houve diferença estatística na maioria dos cenários (exceções destacadas em amarelo, geralmente envolvendo o algoritmo SVD++).

Na Tabela D.5, os destaques em amarelo ilustram que, com os dados oriundos do TripAdvisor, geralmente não houve diferença estatística de desempenho em relação às variações analisadas (com exceção de três cenários — referentes à métrica RMSE e aos algoritmos UserKNN, ItemKNN e SMF, especificamente com o TripAdvisorCompleto).

Na Tabela D.6, os destaques em amarelo ilustram que não houve diferença estatística em poucos cenários (envolvendo a métrica *F-Measure* e os algoritmos BMF e ItemKNN).

Na Tabela D.7, os destaques em amarelo ilustram que, com os dados oriundos do TripAdvisor, comumente não houve diferença estatística de desempenho em relação às

variações analisadas (houve diferença em cenários considerando a métrica RMSE e os algoritmos UserKNN e ItemKNN). Em relação aos dados oriundos da Amazon, houve dois cenários sem diferença estatística (envolvendo o algoritmo ItemKNN). Na Tabela D.7, não há os pares referentes ao algoritmo baseline e a variação “tps”, pois isto já é apresentado na Tabela D.2.

Na Tabela D.8, a ausência de destaques em amarelo representa que, em todos os cenários analisados, houve diferença estatística.

Na Tabela D.9, os destaques em amarelo ilustram que, com o conjunto de dados TripAdvisor, não houve diferença estatística em relação às métricas analisadas referentes à acurácia.

Tabela D.1: Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP1.

Conjunto de dados	Varição de FC	Métrica	ANOVA: $Pr(>F)$	Kruskal-Wallis: $p$ -value
TripAdvisor	Hu&Pu	RMSE	1,00E+00	1,00E+00
TripAdvisor	Hu&Pu	F-Measure	9,96E-01	9,84E-01
TripAdvisor	TkalčičEtAl-public	RMSE	1,00E+00	1,00E+00
TripAdvisor	TkalčičEtAl-public	F-Measure	1,00E+00	1,00E+00
TripAdvisor	TkalčičEtAl-neighbours	RMSE	1,00E+00	1,00E+00
TripAdvisor	TkalčičEtAl-neighbours	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	Hu&Pu	RMSE	9,98E-01	9,98E-01
TripAdvisorCompleto	Hu&Pu	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	TkalčičEtAl-public	RMSE	1,00E+00	1,00E+00
TripAdvisorCompleto	TkalčičEtAl-public	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	TkalčičEtAl-neighbours	RMSE	1,00E+00	1,00E+00
TripAdvisorCompleto	TkalčičEtAl-neighbours	F-Measure	1,00E+00	1,00E+00
AmazonFoods	Hu&Pu	RMSE	1,52E-218	3,64E-92
AmazonFoods	Hu&Pu	F-Measure	1,88E-186	3,53E-79
AmazonFoods	TkalčičEtAl-public	RMSE	2,25E-91	1,25E-26
AmazonFoods	TkalčičEtAl-public	F-Measure	1,48E-201	2,80E-36
AmazonFoods	TkalčičEtAl-neighbours	RMSE	5,35E-293	2,46E-72
AmazonFoods	TkalčičEtAl-neighbours	F-Measure	0,00E+00	7,16E-86
AmazonFoodsCompleto	Hu&Pu	RMSE	1,92E-174	6,03E-83
AmazonFoodsCompleto	Hu&Pu	F-Measure	1,75E-164	1,07E-75
AmazonFoodsCompleto	TkalčičEtAl-public	RMSE	2,18E-12	3,76E-21
AmazonFoodsCompleto	TkalčičEtAl-public	F-Measure	1,99E-94	1,59E-32
AmazonFoodsCompleto	TkalčičEtAl-neighbours	RMSE	1,51E-86	5,75E-55
AmazonFoodsCompleto	TkalčičEtAl-neighbours	F-Measure	2,77E-264	2,50E-77

Fonte: autoria própria.

Tabela D.2: P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP2 (para o modelo *Big Five*).

Conjunto de dados	Algoritmo de FC	Métrica	T	Mann–Whitney U
TripAdvisor	UserKNN	RMSE	8,51E-03	8,54E-03
TripAdvisor	UserKNN	F-Measure	1,91E-01	1,89E-01
TripAdvisor	ItemKNN	RMSE	1,11E-02	4,76E-03
TripAdvisor	ItemKNN	F-Measure	2,32E-01	2,15E-01
TripAdvisor	SMF	RMSE	8,51E-01	7,91E-01
TripAdvisor	SMF	F-Measure	1,00E+00	1,00E+00
TripAdvisor	BMF	RMSE	3,93E-01	3,79E-01
TripAdvisor	BMF	F-Measure	7,36E-02	8,86E-02
TripAdvisor	SVD++	RMSE	5,59E-01	5,10E-01
TripAdvisor	SVD++	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	UserKNN	RMSE	2,42E-09	7,14E-08
TripAdvisorCompleto	UserKNN	F-Measure	2,93E-01	2,64E-01
TripAdvisorCompleto	ItemKNN	RMSE	7,82E-08	1,97E-06
TripAdvisorCompleto	ItemKNN	F-Measure	2,97E-01	2,43E-01
TripAdvisorCompleto	SMF	RMSE	3,66E-05	1,12E-04
TripAdvisorCompleto	SMF	F-Measure	8,25E-01	7,33E-01
TripAdvisorCompleto	BMF	RMSE	9,17E-01	8,99E-01
TripAdvisorCompleto	BMF	F-Measure	1,64E-01	1,47E-01
TripAdvisorCompleto	SVD++	RMSE	2,54E-01	1,79E-01
TripAdvisorCompleto	SVD++	F-Measure	9,83E-01	9,97E-01
AmazonFoods	UserKNN	RMSE	4,76E-77	7,06E-18
AmazonFoods	UserKNN	F-Measure	2,29E-58	7,06E-18
AmazonFoods	ItemKNN	RMSE	1,56E-22	4,90E-16
AmazonFoods	ItemKNN	F-Measure	8,86E-04	1,35E-03
AmazonFoods	SMF	RMSE	9,24E-50	7,06E-18
AmazonFoods	SMF	F-Measure	3,08E-40	7,49E-18
AmazonFoods	BMF	RMSE	3,01E-44	7,05E-18
AmazonFoods	BMF	F-Measure	3,80E-16	3,09E-13
AmazonFoods	SVD++	RMSE	4,88E-05	3,37E-05
AmazonFoods	SVD++	F-Measure	3,25E-03	5,07E-03
AmazonFoodsCompleto	UserKNN	RMSE	3,68E-84	7,06E-18
AmazonFoodsCompleto	UserKNN	F-Measure	2,64E-54	7,06E-18
AmazonFoodsCompleto	ItemKNN	RMSE	4,82E-16	1,27E-12
AmazonFoodsCompleto	ItemKNN	F-Measure	3,51E-02	5,36E-02
AmazonFoodsCompleto	SMF	RMSE	1,34E-31	1,36E-17
AmazonFoodsCompleto	SMF	F-Measure	9,71E-37	1,01E-17
AmazonFoodsCompleto	BMF	RMSE	1,07E-33	1,01E-17
AmazonFoodsCompleto	BMF	F-Measure	6,30E-12	1,05E-09
AmazonFoodsCompleto	SVD++	RMSE	3,85E-05	1,44E-04
AmazonFoodsCompleto	SVD++	F-Measure	2,32E-01	3,13E-01

Fonte: autoria própria.

Tabela D.3: P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP2 (para o modelo *Values*).

Conjunto de dados	Algoritmo de FC	Métrica	T	Mann–Whitney U
TripAdvisor	UserKNN	RMSE	5,20E-03	6,40E-03
TripAdvisor	UserKNN	F-Measure	4,16E-01	4,06E-01
TripAdvisor	ItemKNN	RMSE	6,04E-03	2,18E-03
TripAdvisor	ItemKNN	F-Measure	5,48E-01	6,37E-01
TripAdvisor	SMF	RMSE	5,47E-01	5,15E-01
TripAdvisor	SMF	F-Measure	1,00E+00	1,00E+00
TripAdvisor	BMF	RMSE	6,40E-01	6,05E-01
TripAdvisor	BMF	F-Measure	9,88E-02	1,11E-01
TripAdvisor	SVD++	RMSE	8,61E-01	8,28E-01
TripAdvisor	SVD++	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	UserKNN	RMSE	1,73E-06	1,62E-05
TripAdvisorCompleto	UserKNN	F-Measure	3,88E-01	3,19E-01
TripAdvisorCompleto	ItemKNN	RMSE	2,21E-05	4,97E-05
TripAdvisorCompleto	ItemKNN	F-Measure	2,71E-01	2,24E-01
TripAdvisorCompleto	SMF	RMSE	4,59E-03	3,54E-03
TripAdvisorCompleto	SMF	F-Measure	8,21E-01	7,36E-01
TripAdvisorCompleto	BMF	RMSE	9,44E-01	9,20E-01
TripAdvisorCompleto	BMF	F-Measure	1,74E-01	1,89E-01
TripAdvisorCompleto	SVD++	RMSE	5,81E-01	4,40E-01
TripAdvisorCompleto	SVD++	F-Measure	9,61E-01	9,40E-01
AmazonFoods	UserKNN	RMSE	1,96E-34	7,07E-18
AmazonFoods	UserKNN	F-Measure	2,22E-28	7,06E-18
AmazonFoods	ItemKNN	RMSE	6,03E-21	3,57E-15
AmazonFoods	ItemKNN	F-Measure	2,40E-03	5,18E-03
AmazonFoods	SMF	RMSE	2,29E-51	7,06E-18
AmazonFoods	SMF	F-Measure	1,09E-39	7,50E-18
AmazonFoods	BMF	RMSE	7,01E-45	7,06E-18
AmazonFoods	BMF	F-Measure	1,33E-16	3,25E-13
AmazonFoods	SVD++	RMSE	3,30E-01	3,85E-01
AmazonFoods	SVD++	F-Measure	1,90E-02	2,55E-02
AmazonFoodsCompleto	UserKNN	RMSE	1,16E-78	7,06E-18
AmazonFoodsCompleto	UserKNN	F-Measure	8,86E-48	7,06E-18
AmazonFoodsCompleto	ItemKNN	RMSE	1,38E-14	9,23E-12
AmazonFoodsCompleto	ItemKNN	F-Measure	2,08E-04	2,55E-04
AmazonFoodsCompleto	SMF	RMSE	7,17E-37	7,50E-18
AmazonFoodsCompleto	SMF	F-Measure	6,54E-35	1,01E-17
AmazonFoodsCompleto	BMF	RMSE	3,85E-35	8,46E-18
AmazonFoodsCompleto	BMF	F-Measure	1,69E-13	4,81E-11
AmazonFoodsCompleto	SVD++	RMSE	4,93E-01	4,00E-01
AmazonFoodsCompleto	SVD++	F-Measure	8,79E-01	9,37E-01

Fonte: autoria própria.

Tabela D.4: P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP2 (para o modelo *Needs*).

Conjunto de dados	Algoritmo de FC	Métrica	T	Mann–Whitney U
TripAdvisor	UserKNN	RMSE	1,39E-02	1,48E-02
TripAdvisor	UserKNN	F-Measure	4,84E-01	5,12E-01
TripAdvisor	ItemKNN	RMSE	1,03E-02	4,32E-03
TripAdvisor	ItemKNN	F-Measure	6,99E-01	6,22E-01
TripAdvisor	SMF	RMSE	7,74E-01	7,12E-01
TripAdvisor	SMF	F-Measure	1,00E+00	1,00E+00
TripAdvisor	BMF	RMSE	4,36E-01	3,68E-01
TripAdvisor	BMF	F-Measure	9,90E-02	1,37E-01
TripAdvisor	SVD++	RMSE	6,34E-01	5,60E-01
TripAdvisor	SVD++	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	UserKNN	RMSE	6,25E-07	6,11E-06
TripAdvisorCompleto	UserKNN	F-Measure	2,80E-01	2,75E-01
TripAdvisorCompleto	ItemKNN	RMSE	7,87E-06	2,57E-05
TripAdvisorCompleto	ItemKNN	F-Measure	3,27E-01	2,90E-01
TripAdvisorCompleto	SMF	RMSE	1,38E-03	1,32E-03
TripAdvisorCompleto	SMF	F-Measure	8,21E-01	7,30E-01
TripAdvisorCompleto	BMF	RMSE	9,76E-01	9,56E-01
TripAdvisorCompleto	BMF	F-Measure	1,37E-01	1,03E-01
TripAdvisorCompleto	SVD++	RMSE	4,65E-01	3,59E-01
TripAdvisorCompleto	SVD++	F-Measure	9,71E-01	9,78E-01
AmazonFoods	UserKNN	RMSE	9,09E-41	7,07E-18
AmazonFoods	UserKNN	F-Measure	2,60E-34	7,06E-18
AmazonFoods	ItemKNN	RMSE	7,03E-27	3,74E-17
AmazonFoods	ItemKNN	F-Measure	2,47E-04	4,67E-04
AmazonFoods	SMF	RMSE	2,44E-42	7,06E-18
AmazonFoods	SMF	F-Measure	2,55E-35	1,14E-17
AmazonFoods	BMF	RMSE	3,62E-39	7,06E-18
AmazonFoods	BMF	F-Measure	6,90E-10	9,13E-09
AmazonFoods	SVD++	RMSE	7,14E-03	4,76E-03
AmazonFoods	SVD++	F-Measure	8,87E-05	1,23E-04
AmazonFoodsCompleto	UserKNN	RMSE	9,58E-87	7,06E-18
AmazonFoodsCompleto	UserKNN	F-Measure	1,43E-56	7,06E-18
AmazonFoodsCompleto	ItemKNN	RMSE	3,28E-21	3,19E-15
AmazonFoodsCompleto	ItemKNN	F-Measure	9,39E-06	1,98E-05
AmazonFoodsCompleto	SMF	RMSE	1,64E-25	8,48E-17
AmazonFoodsCompleto	SMF	F-Measure	2,96E-30	2,46E-17
AmazonFoodsCompleto	BMF	RMSE	2,17E-28	6,15E-17
AmazonFoodsCompleto	BMF	F-Measure	1,62E-06	8,44E-06
AmazonFoodsCompleto	SVD++	RMSE	1,35E-02	1,54E-02
AmazonFoodsCompleto	SVD++	F-Measure	4,64E-02	7,87E-02

Fonte: autoria própria.

Tabela D.5: Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP3 (para variações omitindo traços).

Conjunto de dados	Algoritmo de FC	Métrica	ANOVA: $Pr(>F)$	Kruskal-Wallis: $p$ -value
TripAdvisor	UserKNN	RMSE	1,00E+00	1,00E+00
TripAdvisor	UserKNN	F-Measure	1,00E+00	1,00E+00
TripAdvisor	ItemKNN	RMSE	1,00E+00	1,00E+00
TripAdvisor	ItemKNN	F-Measure	1,00E+00	1,00E+00
TripAdvisor	SMF	RMSE	1,00E+00	1,00E+00
TripAdvisor	SMF	F-Measure	1,00E+00	1,00E+00
TripAdvisor	BMF	RMSE	1,00E+00	9,94E-01
TripAdvisor	BMF	F-Measure	1,00E+00	1,00E+00
TripAdvisor	SVD++	RMSE	1,00E+00	1,00E+00
TripAdvisor	SVD++	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	UserKNN	RMSE	1,35E-14	5,27E-17
TripAdvisorCompleto	UserKNN	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	ItemKNN	RMSE	6,57E-11	5,33E-15
TripAdvisorCompleto	ItemKNN	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	SMF	RMSE	7,72E-06	1,11E-10
TripAdvisorCompleto	SMF	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	BMF	RMSE	1,00E+00	1,00E+00
TripAdvisorCompleto	BMF	F-Measure	1,00E+00	1,00E+00
TripAdvisorCompleto	SVD++	RMSE	1,00E+00	1,00E+00
TripAdvisorCompleto	SVD++	F-Measure	1,00E+00	1,00E+00
AmazonFoods	UserKNN	RMSE	0,00E+00	0,00E+00
AmazonFoods	UserKNN	F-Measure	0,00E+00	0,00E+00
AmazonFoods	ItemKNN	RMSE	0,00E+00	0,00E+00
AmazonFoods	ItemKNN	F-Measure	4,77E-79	1,68E-62
AmazonFoods	SMF	RMSE	0,00E+00	0,00E+00
AmazonFoods	SMF	F-Measure	0,00E+00	1,15E-281
AmazonFoods	BMF	RMSE	0,00E+00	0,00E+00
AmazonFoods	BMF	F-Measure	5,63E-271	3,50E-205
AmazonFoods	SVD++	RMSE	0,00E+00	0,00E+00
AmazonFoods	SVD++	F-Measure	2,49E-196	9,25E-161
AmazonFoodsCompleto	UserKNN	RMSE	0,00E+00	0,00E+00
AmazonFoodsCompleto	UserKNN	F-Measure	0,00E+00	0,00E+00
AmazonFoodsCompleto	ItemKNN	RMSE	0,00E+00	0,00E+00
AmazonFoodsCompleto	ItemKNN	F-Measure	3,00E-179	1,16E-153
AmazonFoodsCompleto	SMF	RMSE	0,00E+00	0,00E+00
AmazonFoodsCompleto	SMF	F-Measure	0,00E+00	5,52E-303
AmazonFoodsCompleto	BMF	RMSE	0,00E+00	4,16E-246
AmazonFoodsCompleto	BMF	F-Measure	1,38E-247	1,20E-173
AmazonFoodsCompleto	SVD++	RMSE	0,00E+00	0,00E+00
AmazonFoodsCompleto	SVD++	F-Measure	1,11E-187	8,03E-161

Fonte: autoria própria.

Tabela D.6: Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP3 (para variações sem omitir traços).

Conjunto de dados	Algoritmo de FC	Métrica	ANOVA: $Pr(>F)$	Kruskal-Wallis: $p$ -value
AmazonFoods	UserKNN	RMSE	0,00E+00	0,00E+00
AmazonFoods	UserKNN	F-Measure	2,66E-303	9,41E-227
AmazonFoods	ItemKNN	RMSE	2,18E-32	4,49E-25
AmazonFoods	ItemKNN	F-Measure	1,00E+00	1,00E+00
AmazonFoods	SMF	RMSE	5,12E-80	1,19E-62
AmazonFoods	SMF	F-Measure	4,92E-05	5,72E-04
AmazonFoods	BMF	RMSE	1,97E-19	8,21E-16
AmazonFoods	BMF	F-Measure	1,43E-01	1,83E-01
AmazonFoods	SVD++	RMSE	1,50E-126	4,51E-122
AmazonFoods	SVD++	F-Measure	8,49E-20	1,04E-15
AmazonFoodsCompleto	UserKNN	RMSE	0,00E+00	0,00E+00
AmazonFoodsCompleto	UserKNN	F-Measure	4,08E-285	4,44E-215
AmazonFoodsCompleto	ItemKNN	RMSE	1,14E-52	4,00E-41
AmazonFoodsCompleto	ItemKNN	F-Measure	1,93E-02	9,45E-03
AmazonFoodsCompleto	SMF	RMSE	2,01E-113	4,66E-91
AmazonFoodsCompleto	SMF	F-Measure	1,28E-03	6,64E-03
AmazonFoodsCompleto	BMF	RMSE	5,33E-03	4,49E-03
AmazonFoodsCompleto	BMF	F-Measure	7,58E-01	9,70E-01
AmazonFoodsCompleto	SVD++	RMSE	2,40E-142	1,56E-115
AmazonFoodsCompleto	SVD++	F-Measure	1,91E-10	9,06E-10
Personality2018	SMF	RMSE	0,00E+00	0,00E+00
Personality2018	SMF	F-Measure	0,00E+00	0,00E+00
Personality2018	BMF	RMSE	0,00E+00	0,00E+00
Personality2018	BMF	F-Measure	0,00E+00	0,00E+00

Fonte: autoria própria.

Tabela D.7: P-valores resultantes dos testes T e Mann–Whitney U para o experimento referente à QP4 (para estratégia “itps”).

Conjunto de dados	Par de Algoritmos	Métrica	T	Mann–Whitney U
TripAdvisor	BMF e BMF_itps	RMSE	8,96E-01	9,01E-01
TripAdvisor	BMF e BMF_itps	F-Measure	4,24E-01	4,24E-01
TripAdvisor	BMF_tps e BMF_itps	RMSE	4,69E-01	4,32E-01
TripAdvisor	BMF_tps e BMF_itps	F-Measure	3,21E-01	2,84E-01
TripAdvisor	SMF e SMF_itps	RMSE	8,20E-01	7,64E-01
TripAdvisor	SMF e SMF_itps	F-Measure	1,00E+00	1,00E+00
TripAdvisor	SMF_tps e SMF_itps	RMSE	9,69E-01	9,53E-01
TripAdvisor	SMF_tps e SMF_itps	F-Measure	1,00E+00	1,00E+00
TripAdvisor	SVD++ e SVD++_itps	RMSE	9,59E-01	9,40E-01
TripAdvisor	SVD++ e SVD++_itps	F-Measure	1,00E+00	1,00E+00
TripAdvisor	SVD++_tps e SVD++_itps	RMSE	5,23E-01	4,48E-01
TripAdvisor	SVD++_tps e SVD++_itps	F-Measure	1,00E+00	1,00E+00
TripAdvisor	UserKNN e UserKNN_itps	RMSE	6,46E-06	8,05E-06
TripAdvisor	UserKNN e UserKNN_itps	F-Measure	5,21E-01	4,97E-01
TripAdvisor	UserKNN_tps e UserKNN_itps	RMSE	3,53E-02	3,02E-02
TripAdvisor	UserKNN_tps e UserKNN_itps	F-Measure	4,84E-01	4,30E-01
TripAdvisor	ItemKNN e ItemKNN_itps	RMSE	1,11E-06	2,21E-07
TripAdvisor	ItemKNN e ItemKNN_itps	F-Measure	4,97E-01	5,08E-01
TripAdvisor	ItemKNN_tps e ItemKNN_itps	RMSE	1,13E-02	6,27E-03
TripAdvisor	ItemKNN_tps e ItemKNN_itps	F-Measure	5,98E-01	6,59E-01
AmazonFoods	BMF e BMF_itps	RMSE	5,10E-01	5,21E-01
AmazonFoods	BMF e BMF_itps	F-Measure	6,85E-19	7,84E-14
AmazonFoods	BMF_tps e BMF_itps	RMSE	8,38E-46	7,06E-18
AmazonFoods	BMF_tps e BMF_itps	F-Measure	9,09E-38	8,98E-18
AmazonFoods	SMF e SMF_itps	RMSE	2,02E-20	5,53E-15
AmazonFoods	SMF e SMF_itps	F-Measure	1,41E-11	3,94E-10
AmazonFoods	SMF_tps e SMF_itps	RMSE	5,63E-65	7,07E-18
AmazonFoods	SMF_tps e SMF_itps	F-Measure	5,25E-50	7,05E-18
AmazonFoods	SVD++ e SVD++_itps	RMSE	1,96E-20	6,34E-15
AmazonFoods	SVD++ e SVD++_itps	F-Measure	6,90E-06	2,20E-05
AmazonFoods	SVD++_tps e SVD++_itps	RMSE	5,96E-29	5,31E-17
AmazonFoods	SVD++_tps e SVD++_itps	F-Measure	7,50E-13	6,49E-11
AmazonFoods	UserKNN e UserKNN_itps	RMSE	1,68E-57	7,06E-18
AmazonFoods	UserKNN e UserKNN_itps	F-Measure	4,68E-43	7,06E-18
AmazonFoods	UserKNN_tps e UserKNN_itps	RMSE	3,91E-40	8,46E-18
AmazonFoods	UserKNN_tps e UserKNN_itps	F-Measure	7,75E-23	2,11E-15
AmazonFoods	ItemKNN e ItemKNN_itps	RMSE	7,22E-11	2,58E-09
AmazonFoods	ItemKNN e ItemKNN_itps	F-Measure	1,74E-02	3,49E-02
AmazonFoods	ItemKNN_tps e ItemKNN_itps	RMSE	2,15E-07	2,54E-06
AmazonFoods	ItemKNN_tps e ItemKNN_itps	F-Measure	2,28E-01	2,06E-01
AmazonFoodsCompleto	BMF e BMF_itps	RMSE	2,21E-49	7,06E-18
AmazonFoodsCompleto	BMF e BMF_itps	F-Measure	1,71E-20	1,73E-15
AmazonFoodsCompleto	BMF_tps e BMF_itps	RMSE	1,49E-19	9,28E-15
AmazonFoodsCompleto	BMF_tps e BMF_itps	F-Measure	7,82E-36	7,05E-18
AmazonFoodsCompleto	SMF e SMF_itps	RMSE	1,04E-28	2,07E-17
AmazonFoodsCompleto	SMF e SMF_itps	F-Measure	3,70E-20	1,18E-14
AmazonFoodsCompleto	SMF_tps e SMF_itps	RMSE	2,53E-57	7,06E-18
AmazonFoodsCompleto	SMF_tps e SMF_itps	F-Measure	1,69E-52	7,05E-18
AmazonFoodsCompleto	SVD++ e SVD++_itps	RMSE	6,21E-50	7,06E-18
AmazonFoodsCompleto	SVD++ e SVD++_itps	F-Measure	1,04E-07	1,31E-07
AmazonFoodsCompleto	SVD++_tps e SVD++_itps	RMSE	1,16E-55	7,07E-18
AmazonFoodsCompleto	SVD++_tps e SVD++_itps	F-Measure	1,91E-10	1,48E-09
AmazonFoodsCompleto	UserKNN e UserKNN_itps	RMSE	3,96E-65	7,06E-18
AmazonFoodsCompleto	UserKNN e UserKNN_itps	F-Measure	1,24E-37	7,06E-18
AmazonFoodsCompleto	UserKNN_tps e UserKNN_itps	RMSE	3,49E-47	7,06E-18
AmazonFoodsCompleto	UserKNN_tps e UserKNN_itps	F-Measure	1,25E-23	4,12E-16
AmazonFoodsCompleto	ItemKNN e ItemKNN_itps	RMSE	2,40E-16	1,41E-12
AmazonFoodsCompleto	ItemKNN e ItemKNN_itps	F-Measure	6,64E-08	1,58E-07
AmazonFoodsCompleto	ItemKNN_tps e ItemKNN_itps	RMSE	9,56E-01	9,48E-01
AmazonFoodsCompleto	ItemKNN_tps e ItemKNN_itps	F-Measure	8,20E-04	7,30E-04

Fonte: autoria própria.

Tabela D.8: Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP4 (para estratégia *weighted*).

Conjunto de dados	Algoritmos	Métrica	ANOVA: $Pr(>F)$	Kruskal-Wallis: $p$ -value
TripAdvisor	BMF+PB	RMSE	6,73E-52	4,30E-41
TripAdvisor	BMF+PB	F-Measure	4,51E-12	7,71E-12
TripAdvisor	SMF+PB	RMSE	2,84E-71	1,20E-52
TripAdvisor	SMF+PB	F-Measure	1,23E-15	4,44E-14
TripAdvisorCompleto	BMF+PB	RMSE	2,93E-26	2,94E-21
TripAdvisorCompleto	BMF+PB	F-Measure	8,81E-07	2,74E-05
TripAdvisorCompleto	SMF+PB	RMSE	6,59E-185	2,63E-76
TripAdvisorCompleto	SMF+PB	F-Measure	4,68E-174	1,03E-72
AmazonFoods	BMF+PB	RMSE	0,00E+00	1,82E-102
AmazonFoods	BMF+PB	F-Measure	0,00E+00	2,05E-98
AmazonFoods	SMF+PB	RMSE	0,00E+00	3,59E-106
AmazonFoods	SMF+PB	F-Measure	0,00E+00	2,64E-104
AmazonFoodsCompleto	BMF+PB	RMSE	0,00E+00	8,40E-100
AmazonFoodsCompleto	BMF+PB	F-Measure	0,00E+00	1,07E-101
AmazonFoodsCompleto	SMF+PB	RMSE	0,00E+00	1,61E-106
AmazonFoodsCompleto	SMF+PB	F-Measure	0,00E+00	2,22E-106

Fonte: autoria própria.

Tabela D.9: Resultados dos testes ANOVA e Kruskal-Wallis para o experimento referente à QP4 (para estratégia *re-ranking*).

Conjunto de dados	Métrica	ANOVA: $Pr(>F)$	Kruskal-Wallis: $p$ -value
TripAdvisor	F-Measure	7,74E-01	8,14E-01
TripAdvisor	MAP	9,26E-01	9,75E-01
TripAdvisor	MRR	8,65E-01	9,11E-01
TripAdvisor	NDCG	9,79E-01	9,56E-01
TripAdvisor	MSI@1	3,76E-57	8,43E-32
TripAdvisor	MSI@5	8,52E-71	3,03E-39
TripAdvisor	MSI@10	2,36E-78	5,03E-42
TripAdvisorCompleto	F-Measure	3,13E-06	6,50E-06
TripAdvisorCompleto	MAP	1,82E-08	1,33E-07
TripAdvisorCompleto	MRR	5,90E-09	5,72E-08
TripAdvisorCompleto	NDCG	1,02E-02	3,48E-02
TripAdvisorCompleto	MSI@1	6,79E-97	3,24E-40
TripAdvisorCompleto	MSI@5	5,31E-104	2,45E-46
TripAdvisorCompleto	MSI@10	1,45E-75	9,34E-42
AmazonFoods	F-Measure	5,19E-37	1,25E-29
AmazonFoods	MAP	7,60E-27	4,28E-21
AmazonFoods	MRR	1,79E-25	1,19E-19
AmazonFoods	NDCG	2,30E-15	8,45E-13
AmazonFoods	MSI@1	8,25E-216	5,76E-49
AmazonFoods	MSI@5	4,37E-258	4,77E-64
AmazonFoods	MSI@10	1,17E-252	6,08E-64
AmazonFoodsCompleto	F-Measure	4,33E-120	2,46E-53
AmazonFoodsCompleto	MAP	2,40E-113	1,47E-49
AmazonFoodsCompleto	MRR	2,68E-134	3,77E-45
AmazonFoodsCompleto	NDCG	7,48E-84	1,43E-41
AmazonFoodsCompleto	MSI@1	6,55E-125	1,39E-43
AmazonFoodsCompleto	MSI@5	9,46E-183	3,68E-60
AmazonFoodsCompleto	MSI@10	1,15E-200	1,17E-60

Fonte: autoria própria.

# Apêndice E

## *Boxplots* e análises referentes à QP2

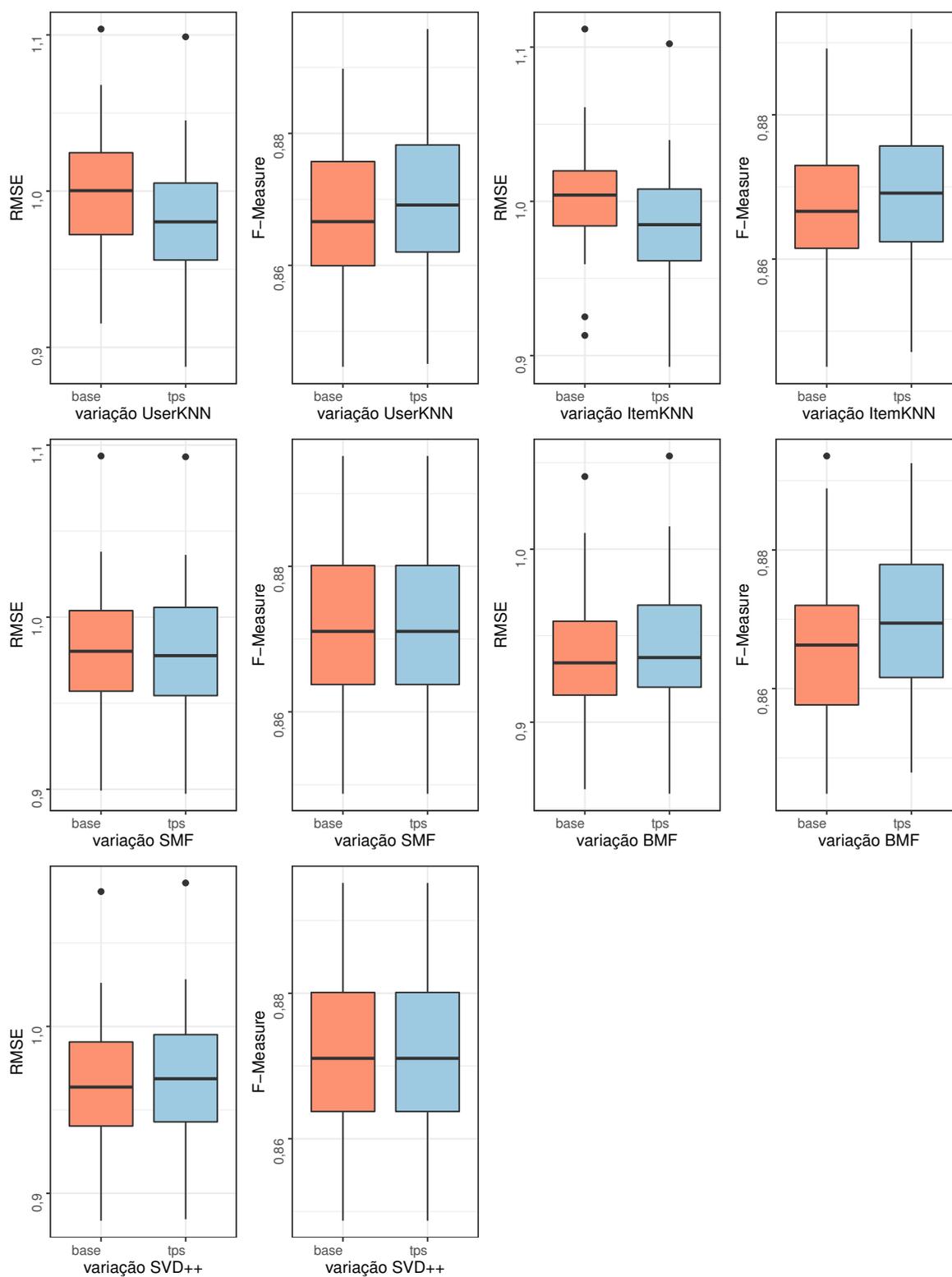
Neste apêndice, encontram-se os *boxplots* ilustrando os resultados referentes à QP2, além de comentários referentes à análise estatística (baseando-se nos resultados de testes T e testes Mann-Whitney U, apresentados no Apêndice D). Para melhor compreensão deste apêndice, deve-se ler a Subseção 5.2.2.

Considerando o modelo *Big Five*, têm-se os *boxplots* da Figura E.1 (conjunto de dados TripAdvisor), da Figura E.2 (conjunto de dados TripAdvisorCompleto), da Figura E.3 (conjunto de dados AmazonFoods) e da Figura E.4 (conjunto de dados AmazonFoodsCompleto).

Comparando em relação ao UserKNN, para TripAdvisor e TripAdvisorCompleto, houve diferença estatística para a métrica RMSE, mas não para *F-Measure*; nas ocorrências de diferença estatística, percebe-se melhor desempenho (menor RMSE) da abordagem considerando personalidade (modelo *Big Five*). Ainda em relação ao UserKNN, para AmazonFoods e AmazonFoodsCompleto, houve diferença estatística tanto para a métrica RMSE quanto para *F-Measure*, com melhor desempenho (menor RMSE e maior *F-Measure*) da abordagem considerando personalidade (modelo *Big Five*).

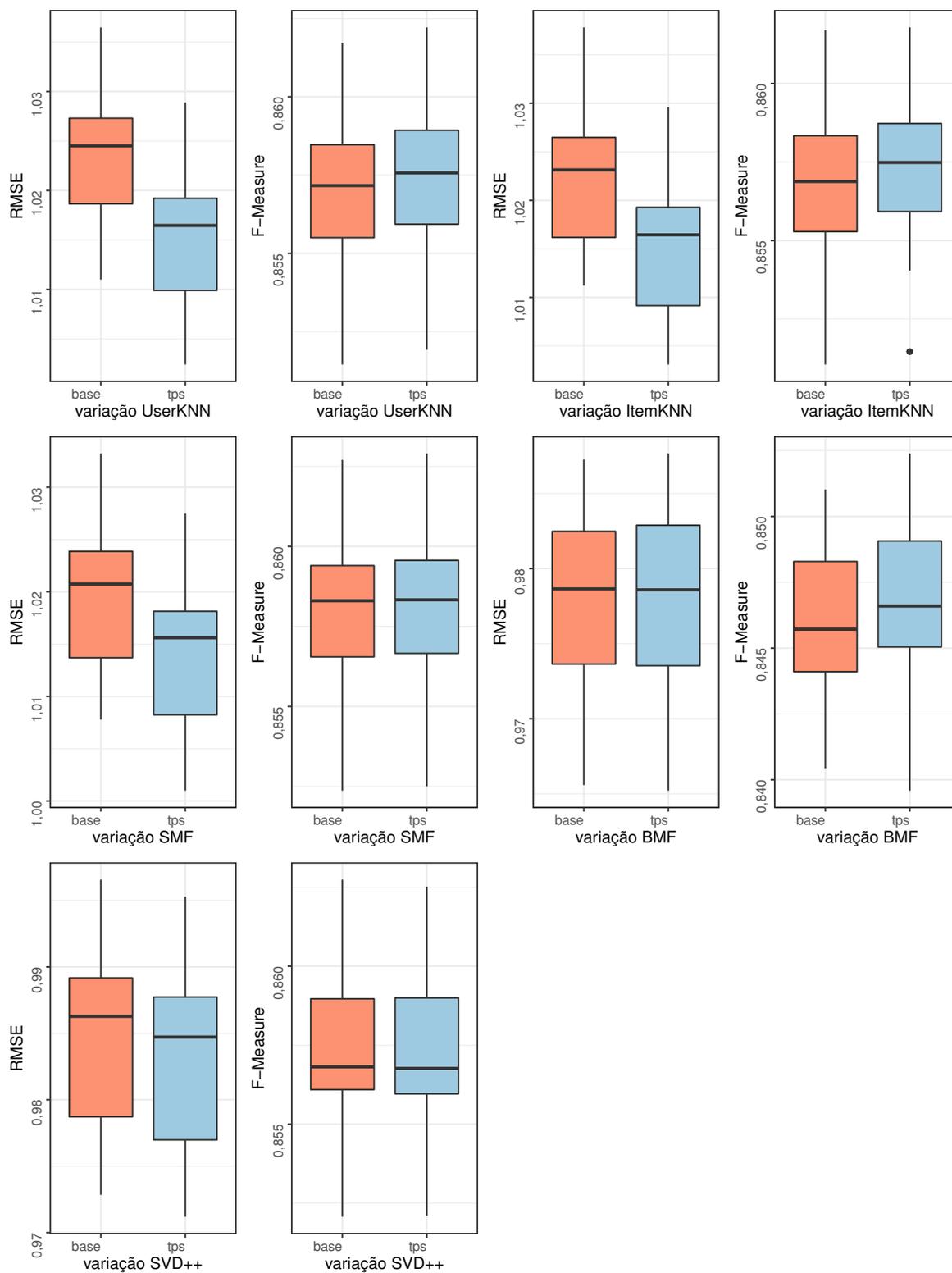
Comparando em relação ao ItemKNN, similarmente ao que foi comentado em relação ao UserKNN, houve diferença estatística para a métrica RMSE, mas não para *F-Measure*, tanto para TripAdvisor quanto TripAdvisorCompleto, sendo o menor RMSE da abordagem considerando personalidade (modelo *Big Five*).

Figura E.1: *Boxplots* referentes à QP2 (modelo *Big Five*) com o conjunto de dados TripAdvisor.



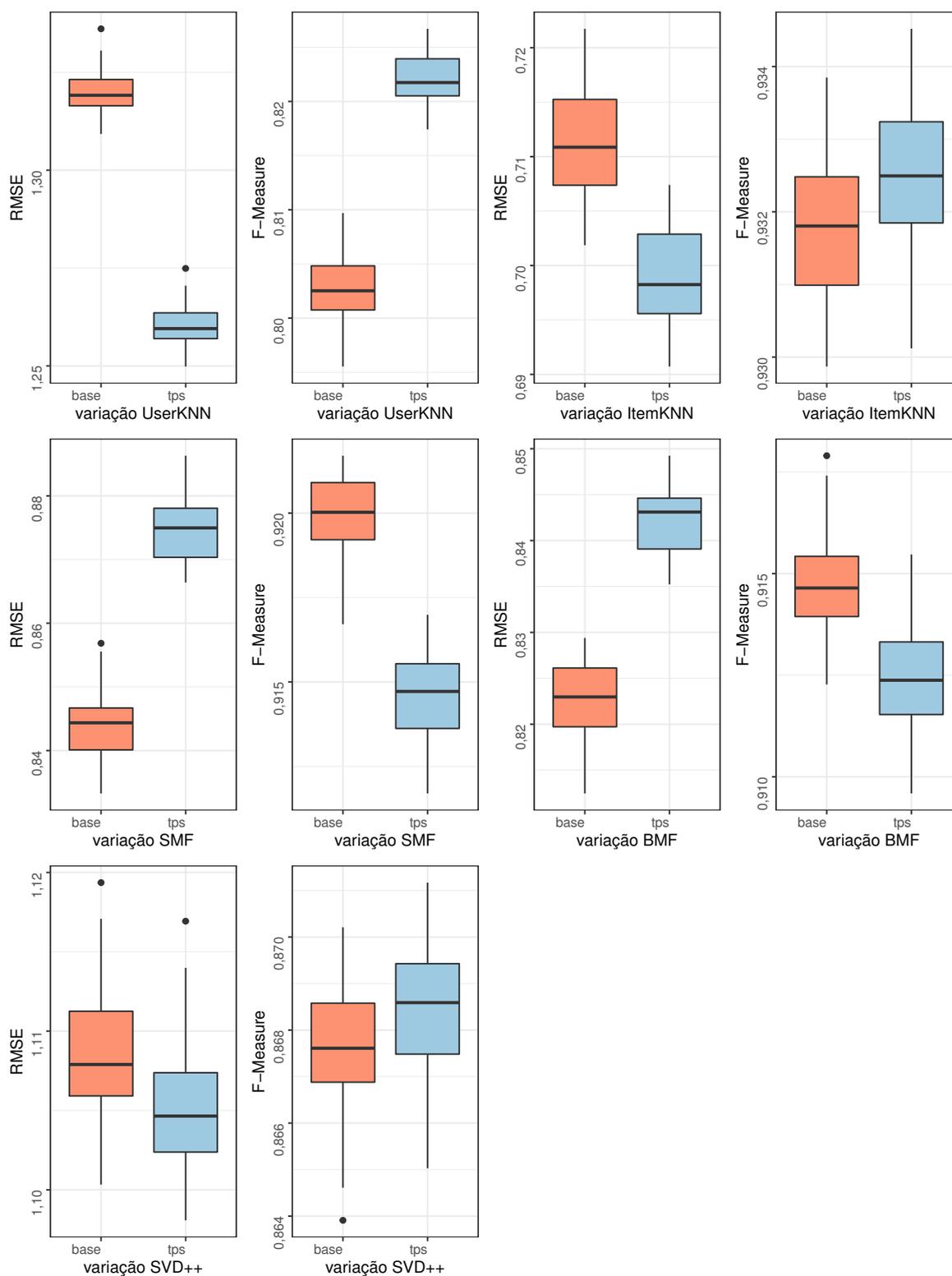
Fonte: autoria própria.

Figura E.2: *Boxplots* referentes à QP2 (modelo *Big Five*) com o conjunto de dados TripAdvisorCompleto.



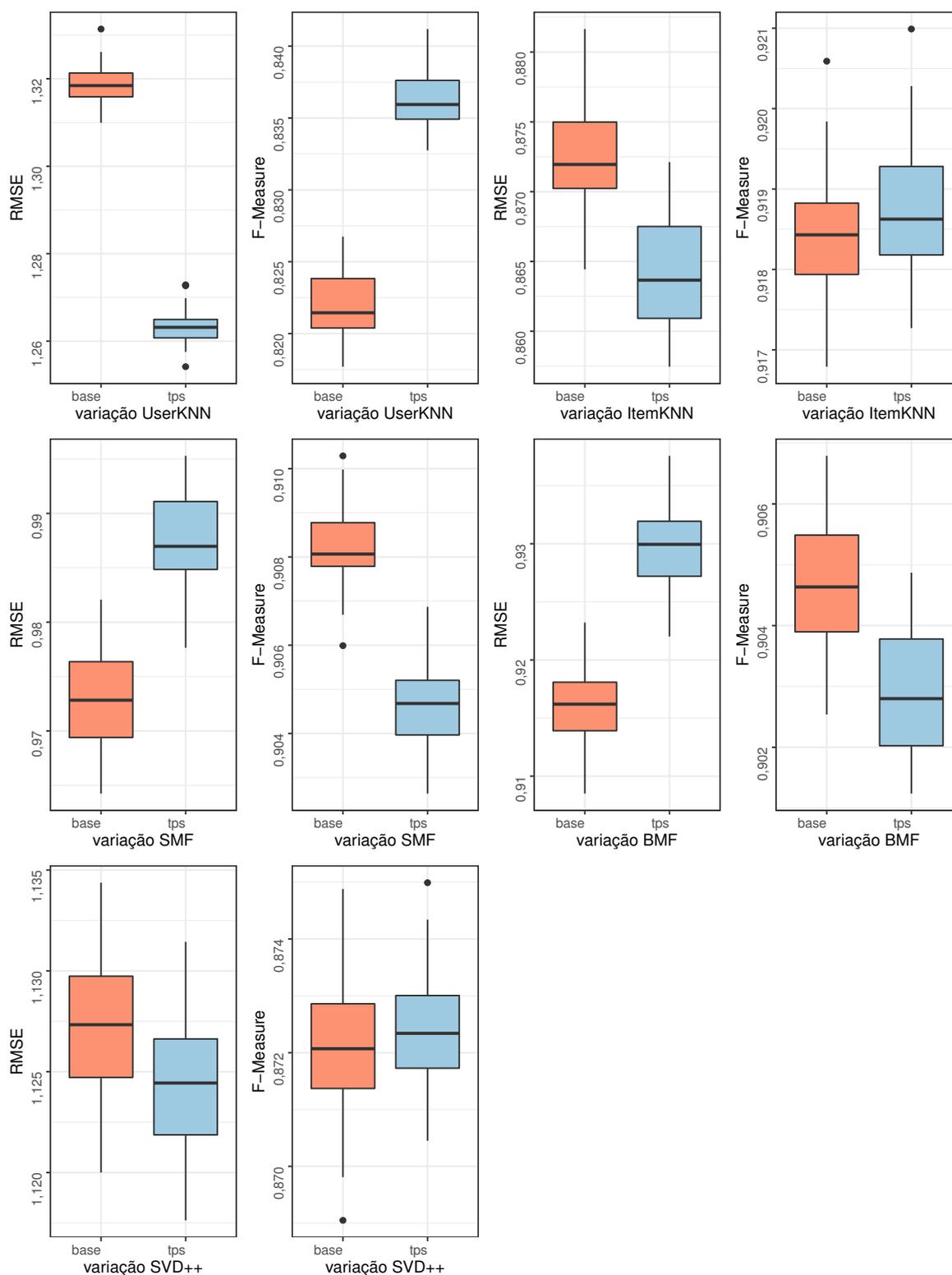
Fonte: autoria própria.

Figura E.3: *Boxplots* referentes à QP2 (modelo *Big Five*) com o conjunto de dados AmazonFoods.



Fonte: autoria própria.

Figura E.4: *Boxplots* referentes à QP2 (modelo *Big Five*) com o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Ainda em relação ao ItemKNN, para AmazonFoods e AmazonFoodsCompleto, também percebe-se, em geral, melhor desempenho (menor RMSE e maior *F-Measure*) da abordagem considerando personalidade (modelo *Big Five*): tanto para o conjunto de dados reduzido quanto para o conjunto de dados completo, houve diferença estatística para a métrica RMSE; e também houve diferença estatística para o conjunto de dados reduzido em relação a *F-Measure*. Todavia, para *F-Measure* no conjunto de dados completo, o p-valor foi de 0,054 no teste U, enquanto foi obtido p-valor 0,035 no teste T (ou seja, avaliações diferentes ao analisar com  $\alpha$  igual a 5%, e considerando certo desvio na normalidade dos dados, não daria para confirmar a diferença estatística).

Comparando em relação ao SMF, para TripAdvisor e TripAdvisorCompleto, só houve diferença estatística para a métrica RMSE, no conjunto de dados completo; nessa ocorrência, o melhor desempenho (menor RMSE) foi da abordagem considerando personalidade (modelo *Big Five*). Ainda em relação ao SMF, tanto para a métrica RMSE quanto para *F-Measure* e tanto para AmazonFoods e AmazonFoodsCompleto, houve diferença estatística; entretanto, a abordagem considerando personalidade (modelo *Big Five*) proporcionou pior desempenho (maior RMSE e menor *F-Measure*).

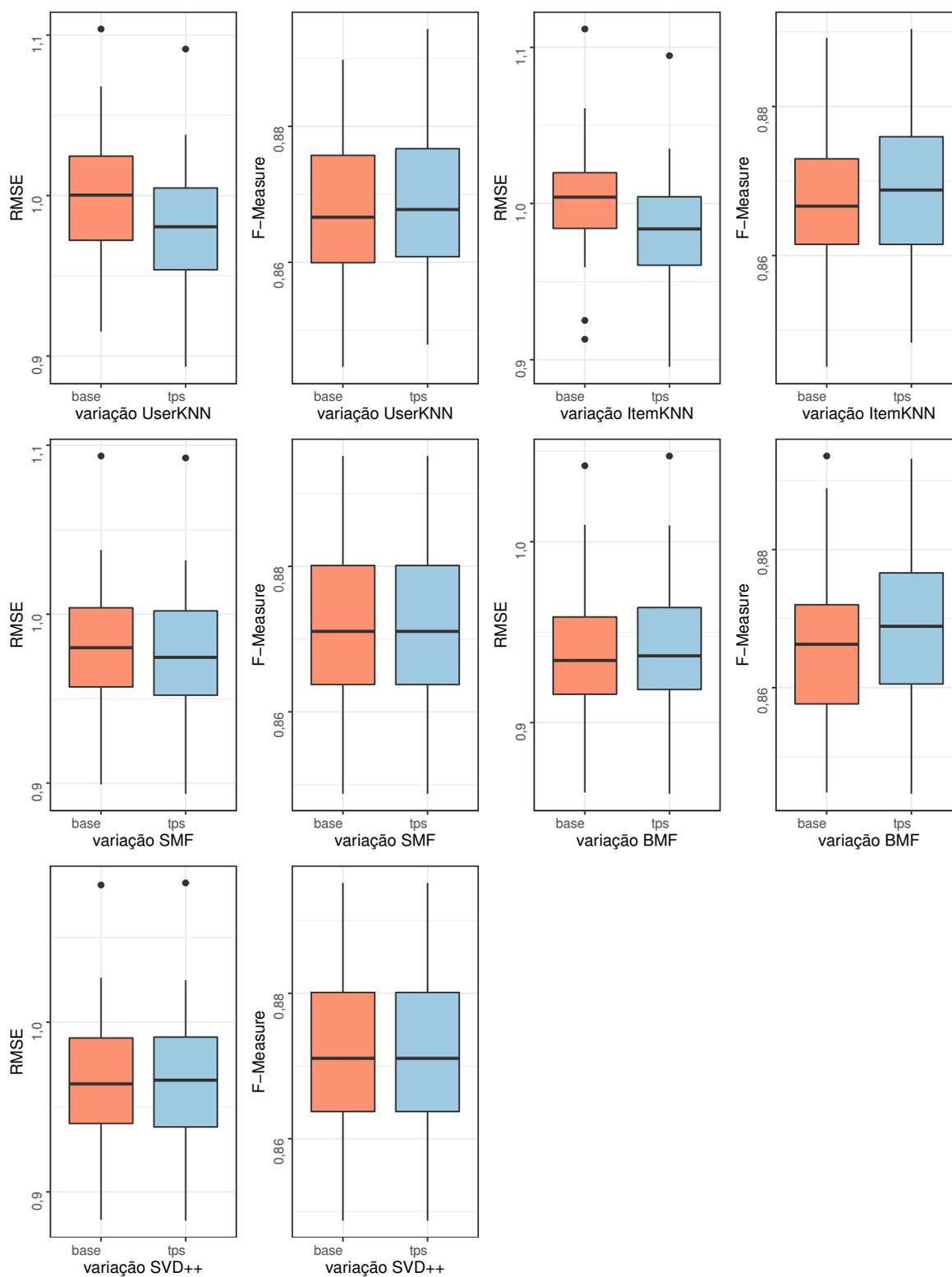
Comparando em relação ao BMF, para TripAdvisor e TripAdvisorCompleto, não houve diferença estatística para ambas as métricas. Para AmazonFoods e AmazonFoodsCompleto, o resultado foi similar ao que foi comentado com o SMF (pior desempenho para a abordagem considerando personalidade via *Big Five*).

Comparando em relação ao SVD++, para TripAdvisor e TripAdvisorCompleto, não houve diferença estatística para ambas as métricas. Para AmazonFoodsCompleto, com a métrica *F-Measure*, também não houve diferença estatística; entretanto, para RMSE, e para ambas as métricas com o AmazonFoods, houve diferença estatística, com melhor desempenho para a abordagem considerando personalidade (modelo *Big Five*).

Considerando o modelo *Values*, têm-se os *boxplots* da Figura E.5 (conjunto de dados TripAdvisor), da Figura E.6 (conjunto de dados TripAdvisorCompleto), da Figura E.7 (conjunto de dados AmazonFoods) e da Figura E.8 (conjunto de dados AmazonFoodsCompleto).

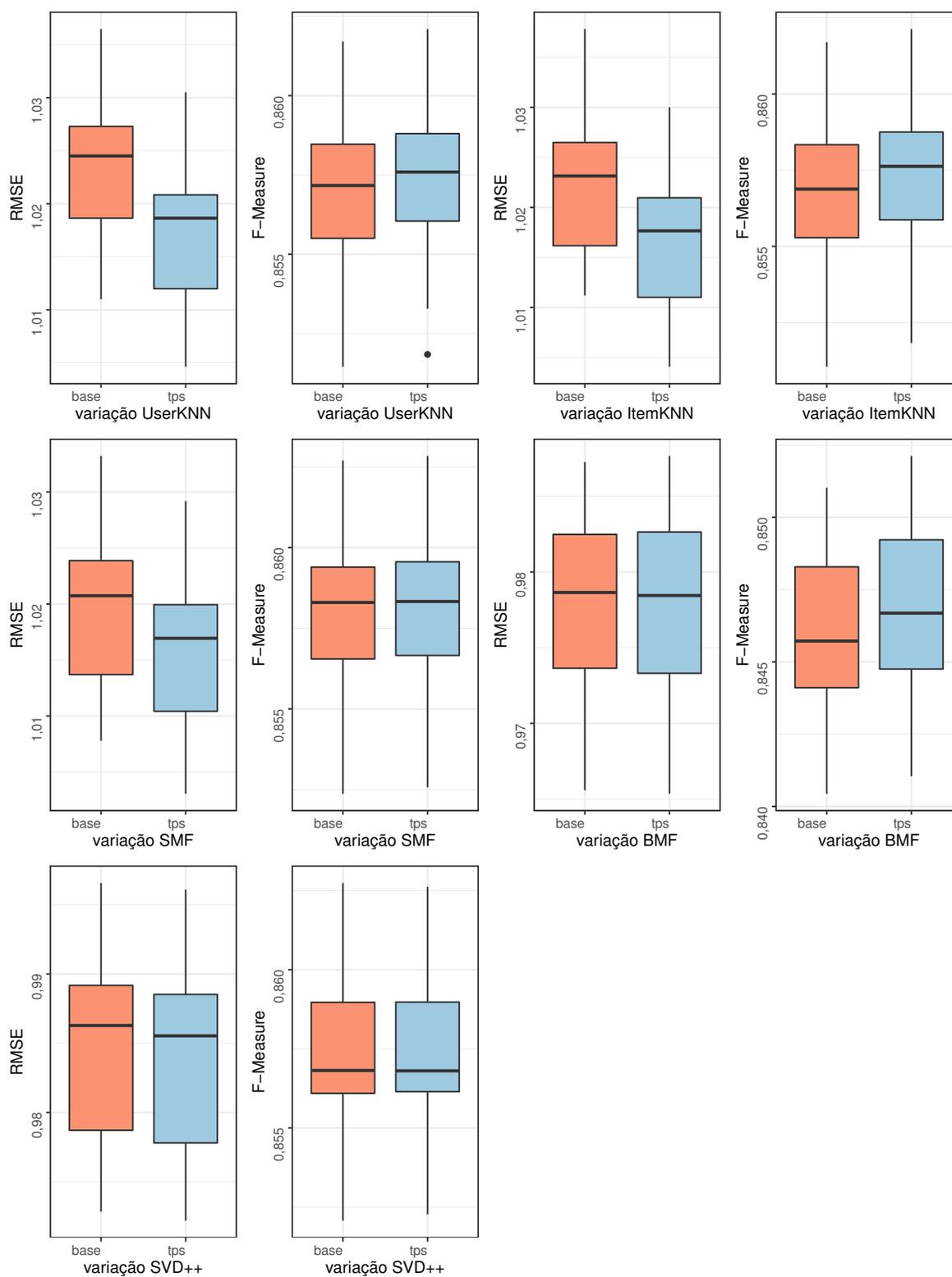
Comparando em relação ao UserKNN e ao ItemKNN, apenas não houve diferença estatística para a métrica *F-Measure* para TripAdvisor e TripAdvisorCompleto. Nas ocorrências de diferença estatística, percebeu-se melhor desempenho (menor RMSE e maior *F-Measure*) da abordagem considerando personalidade (modelo *Values*).

Figura E.5: *Boxplots* referentes à QP2 (modelo *Values*) com o conjunto de dados TripAdvisor.



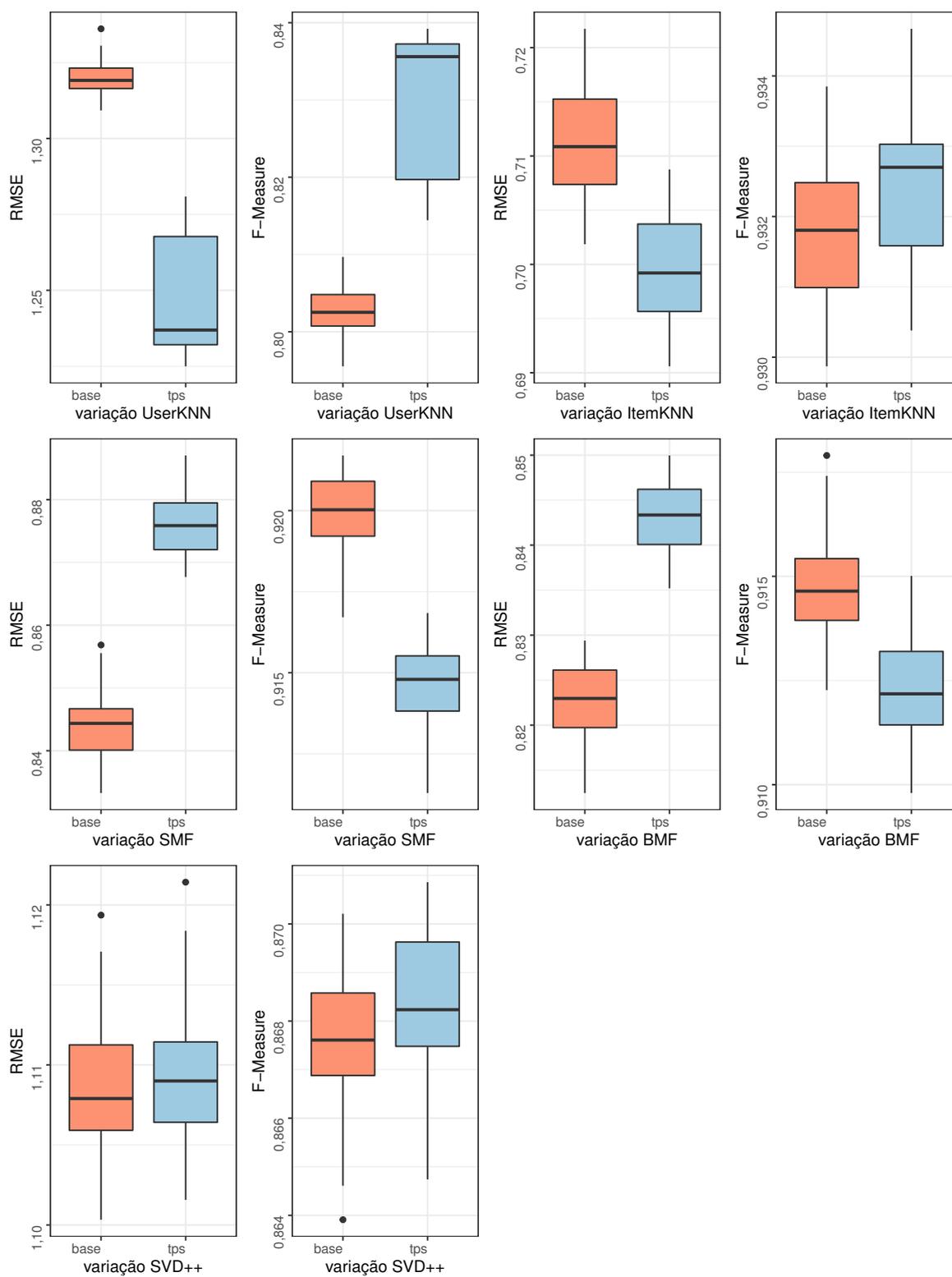
Fonte: autoria própria.

Figura E.6: *Boxplots* referentes à QP2 (modelo *Values*) com o conjunto de dados TripAdvisorCompleto.



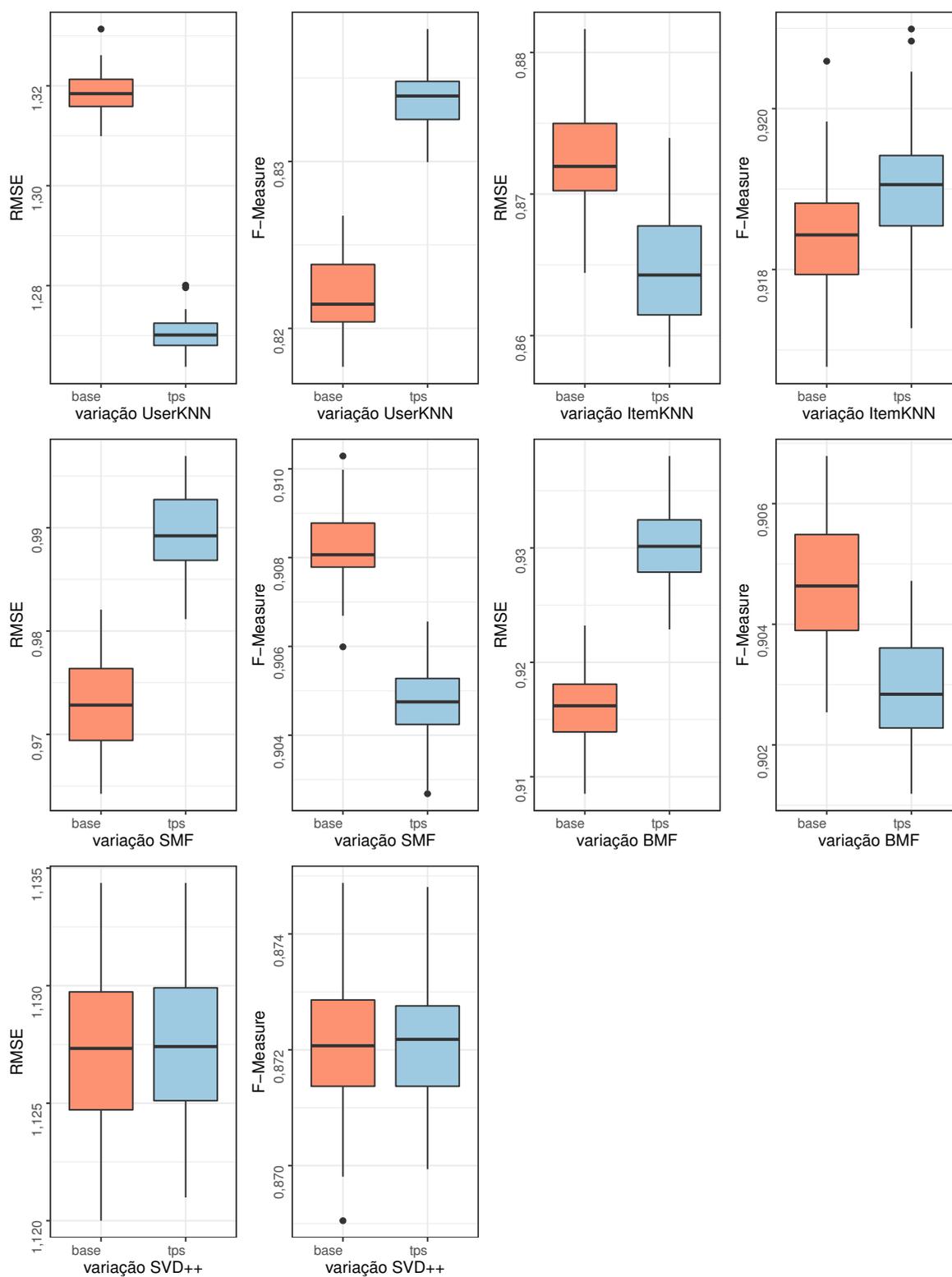
Fonte: autoria própria.

Figura E.7: *Boxplots* referentes à QP2 (modelo *Values*) com o conjunto de dados AmazonFoods.



Fonte: autoria própria.

Figura E.8: *Boxplots* referentes à QP2 (modelo *Values*) com o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Comparando em relação ao SMF, para TripAdvisor e TripAdvisorCompleto, só houve diferença estatística para a métrica RMSE, no conjunto de dados completo; nessa ocorrência, o melhor desempenho (menor RMSE) foi da abordagem considerando personalidade (modelo *Values*). Ainda em relação ao SMF, tanto para AmazonFoods quanto AmazonFoodsCompleto, e para ambas as métricas, houve diferença estatística; entretanto, a abordagem considerando personalidade (modelo *Values*) proporcionou pior desempenho (maior RMSE e menor *F-Measure*).

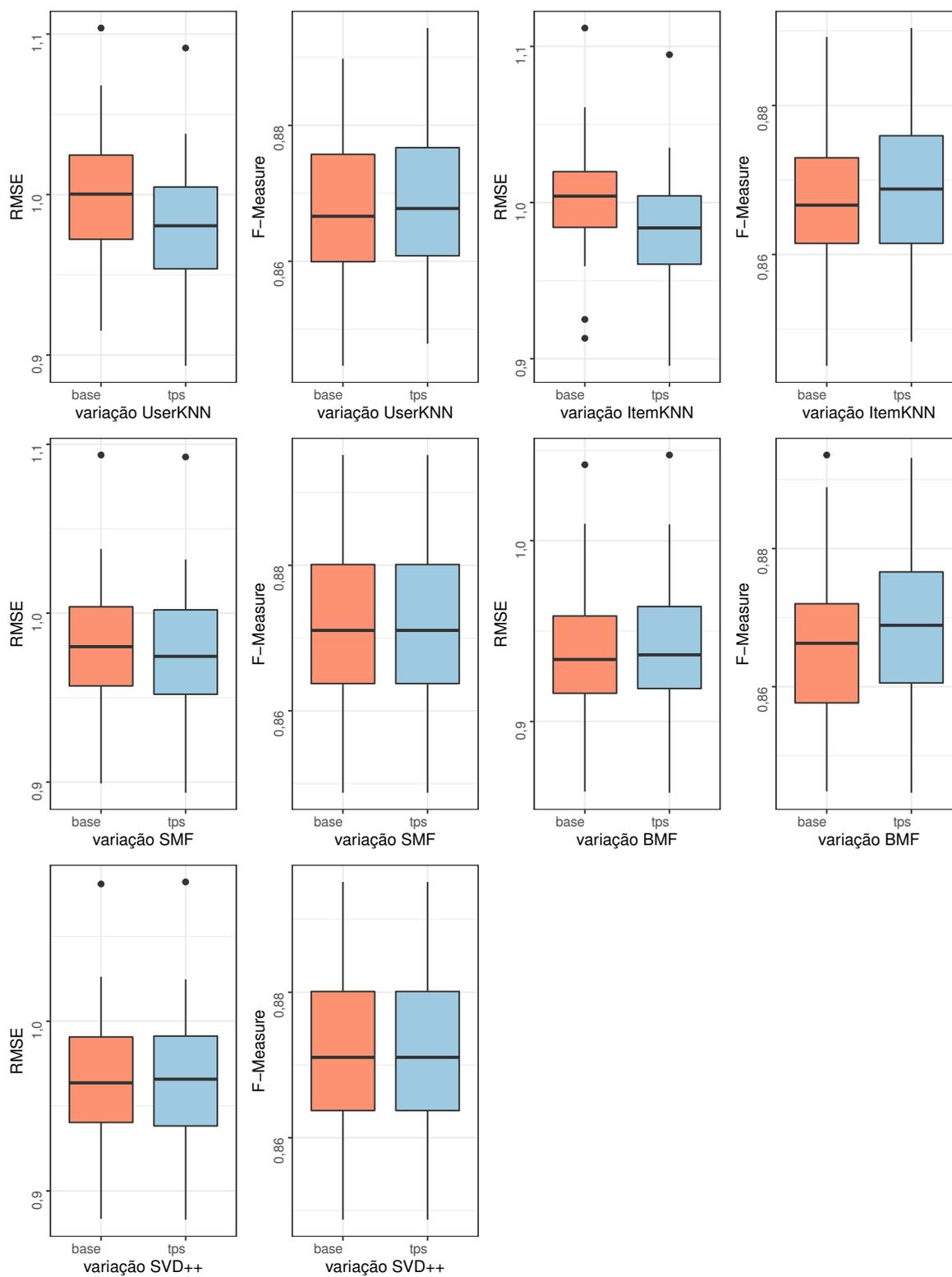
Comparando em relação ao BMF, para TripAdvisor e TripAdvisorCompleto, não houve diferença estatística para ambas as métricas. Para AmazonFoods e AmazonFoodsCompleto, o resultado foi similar ao que foi comentado com o SMF (pior desempenho para a abordagem considerando o modelo *Values*).

Comparando em relação ao SVD++, não houve diferença estatística para a maioria dos casos. Só houve diferença estatística para a métrica *F-Measure* no conjunto de dados AmazonFoods (com melhor desempenho para a abordagem considerando personalidade via modelo *Values*).

Considerando o modelo *Needs*, têm-se os *boxplots* da Figura E.9 (conjunto de dados TripAdvisor), da Figura E.10 (conjunto de dados TripAdvisorCompleto), da Figura E.11 (conjunto de dados AmazonFoods) e da Figura E.12 (conjunto de dados AmazonFoodsCompleto).

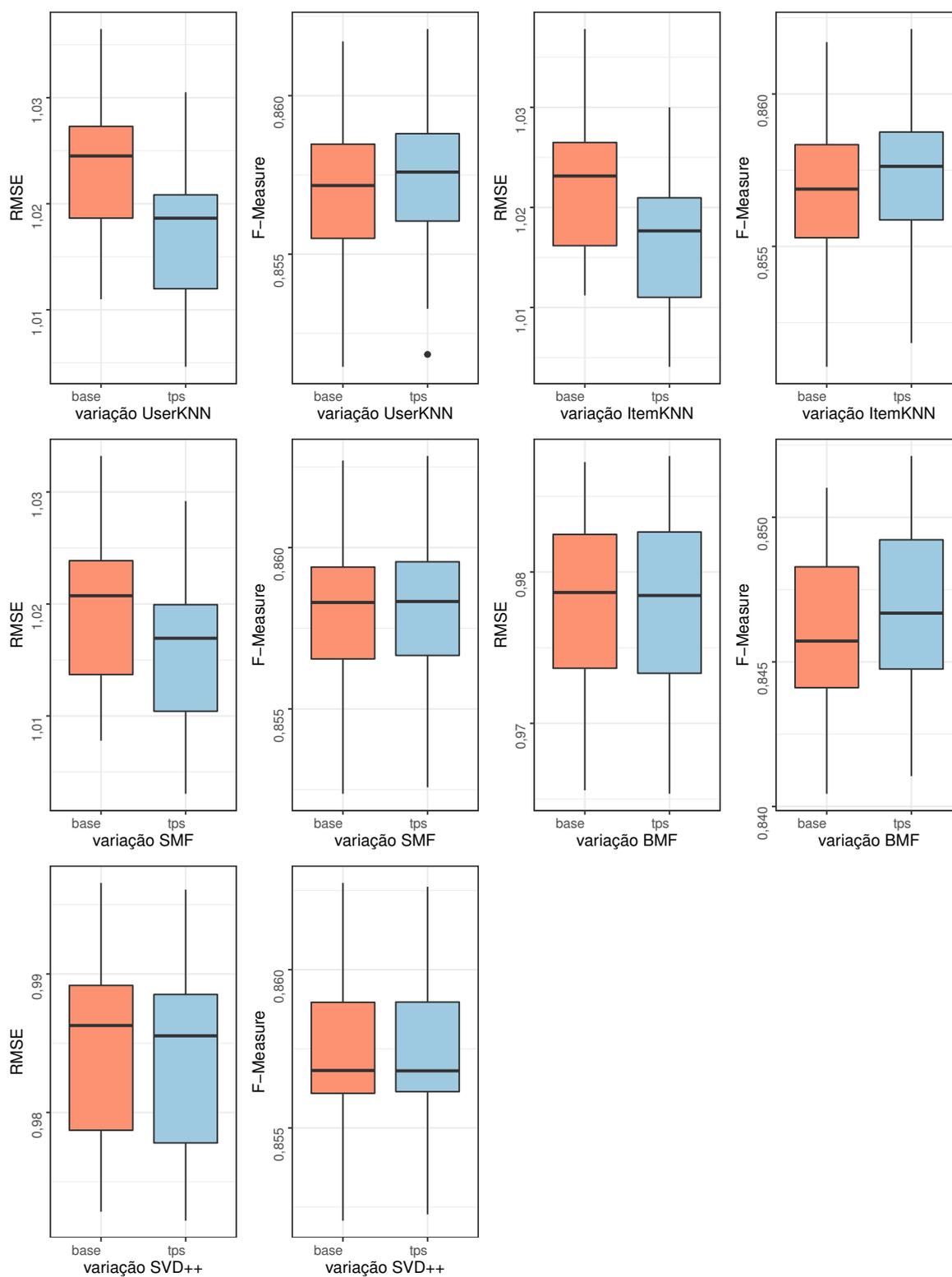
Comparando em relação ao UserKNN e ao ItemKNN, não houve diferença estatística para a métrica *F-Measure*, para TripAdvisor e TripAdvisorCompleto. Nas ocorrências de diferença estatística, o melhor desempenho (menor RMSE e maior *F-Measure*) foi da abordagem considerando o modelo *Needs*.

Comparando em relação ao SMF, para TripAdvisor e TripAdvisorCompleto, só houve diferença estatística para a métrica RMSE, no conjunto de dados completo; nessa ocorrência, o melhor desempenho (menor RMSE) foi da abordagem considerando personalidade (modelo *Needs*). Para AmazonFoods e AmazonFoodsCompleto, houve diferença estatística para ambas as métricas (com a abordagem considerando o modelo *Needs* proporcionando pior desempenho — maior RMSE e menor *F-Measure*).

Figura E.9: *Boxplots* referentes à QP2 (modelo *Needs*) com o conjunto de dados TripAdvisor.

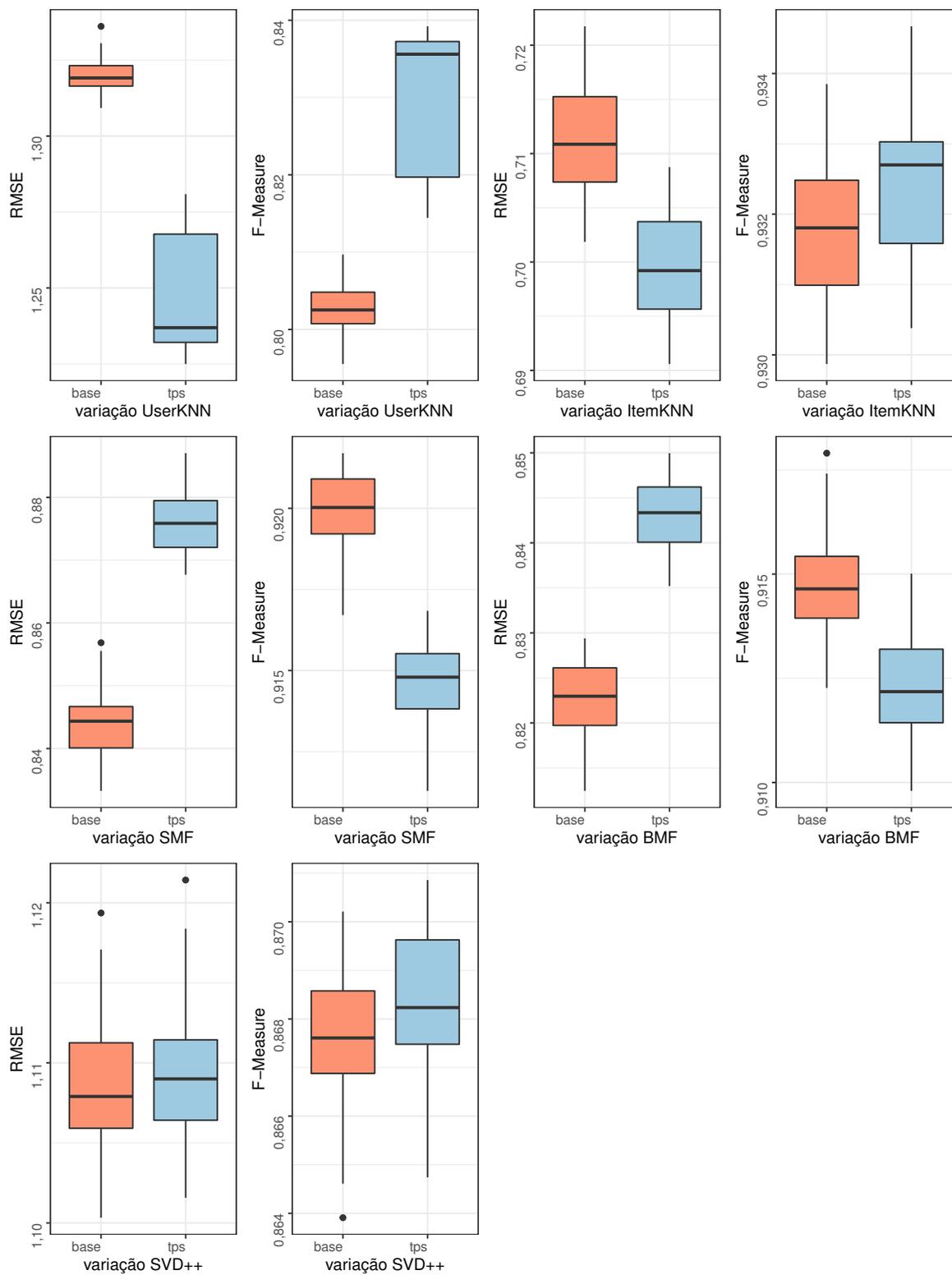
Fonte: autoria própria.

Figura E.10: *Boxplots* referentes à QP2 (modelo *Needs*) com o conjunto de dados TripAdvisorCompleto.



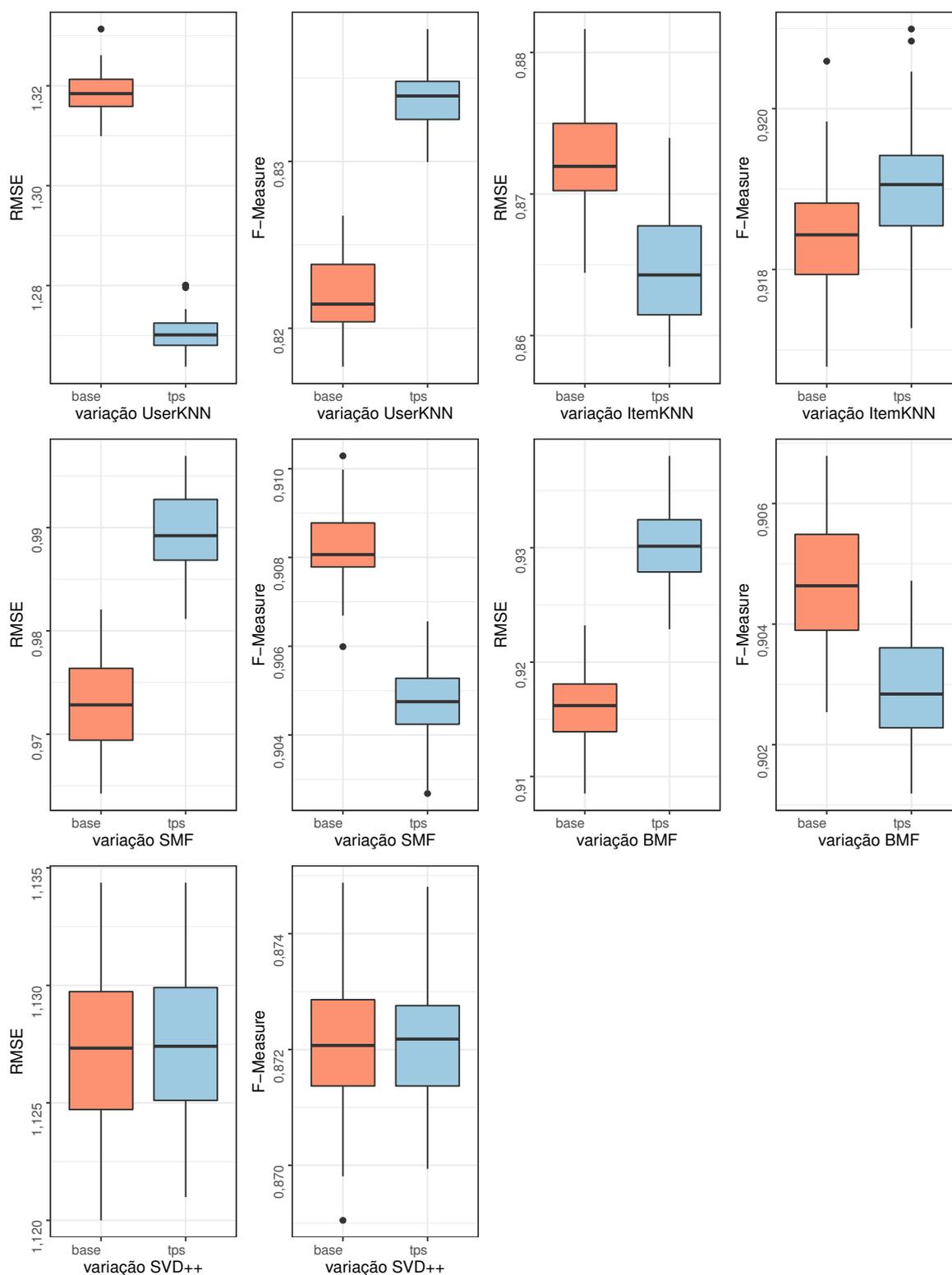
Fonte: autoria própria.

Figura E.11: *Boxplots* referentes à QP2 (modelo *Needs*) com o conjunto de dados AmazonFoods.



Fonte: autoria própria.

Figura E.12: *Boxplots* referentes à QP2 (modelo *Needs*) com o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Comparando em relação ao BMF, para TripAdvisor e TripAdvisorCompleto, não houve diferença estatística para ambas as métricas. Para AmazonFoods e AmazonFoodsCompleto, o resultado foi similar ao que foi comentado com o SMF (pior desempenho para a abordagem considerando o modelo *Needs*).

Comparando em relação ao SVD++, para TripAdvisor e TripAdvisorCompleto, não houve diferença estatística para ambas as métricas. Para AmazonFoods e AmazonFoodsCompleto, houve diferença estatística para a métrica RMSE (pior desempenho para a abordagem considerando o modelo *Needs*); e também houve diferença estatística para o conjunto de dados reduzido em relação a *F-Measure* (melhor desempenho para a abordagem considerando o modelo *Needs*). Todavia, para *F-Measure* no conjunto de dados completo, o p-valor foi de 0,079 no teste U, enquanto foi obtido p-valor 0,046 no teste T (ou seja, avaliações diferentes ao analisar com  $\alpha$  igual a 5%, e considerando certo desvio na normalidade dos dados, não daria para confirmar a diferença estatística).

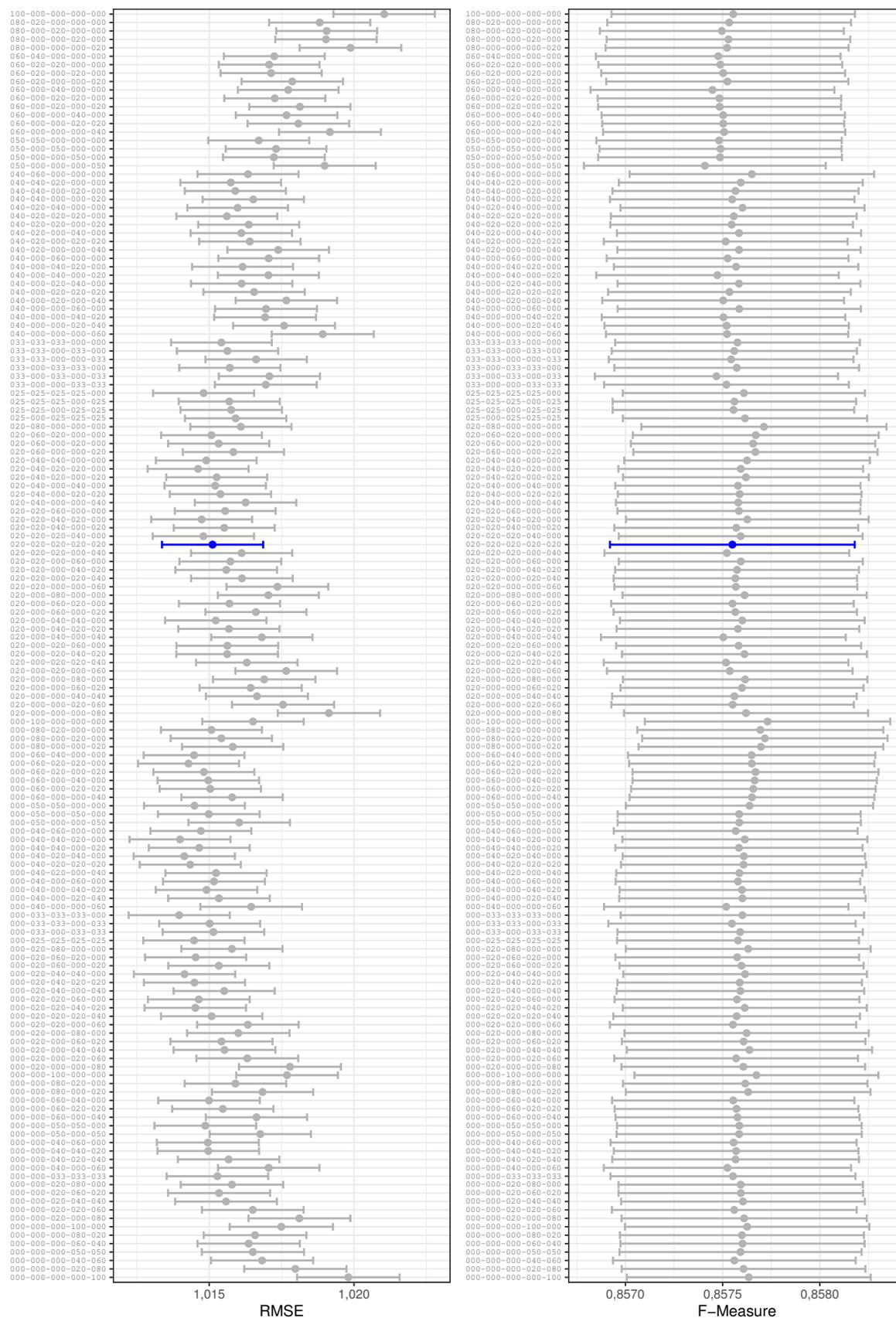
# Apêndice F

## Intervalos de confiança referentes à QP3

Neste apêndice, encontram-se os gráficos ilustrando os intervalos de confiança referentes à QP3 — variações de uso dos traços do *Big Five*, considerando alguns algoritmos base de FC. Os intervalos de confiança destacados em azul representam a variação tradicional (com todos os traços do modelo, e com o mesmo peso para todos).

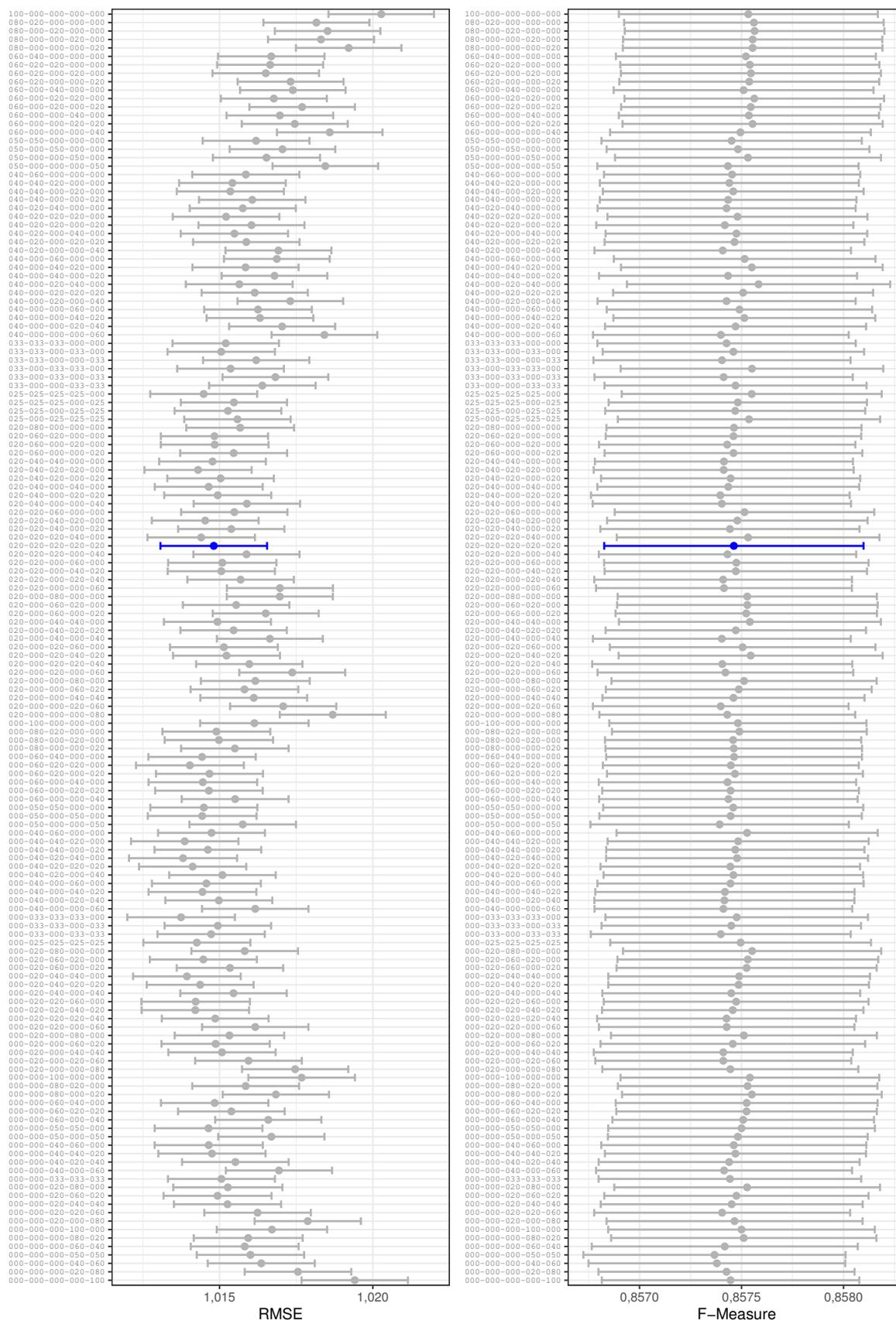
Para melhor compreensão deste apêndice, deve-se ler a Seção 5.2.3. Em síntese, a variação tradicional (020-020-020-020-020), em geral, não proporcionou desempenho significativamente pior em comparação às demais variações.

Figura F.1: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo UserKNN e o conjunto de dados TripAdvisorCompleto.



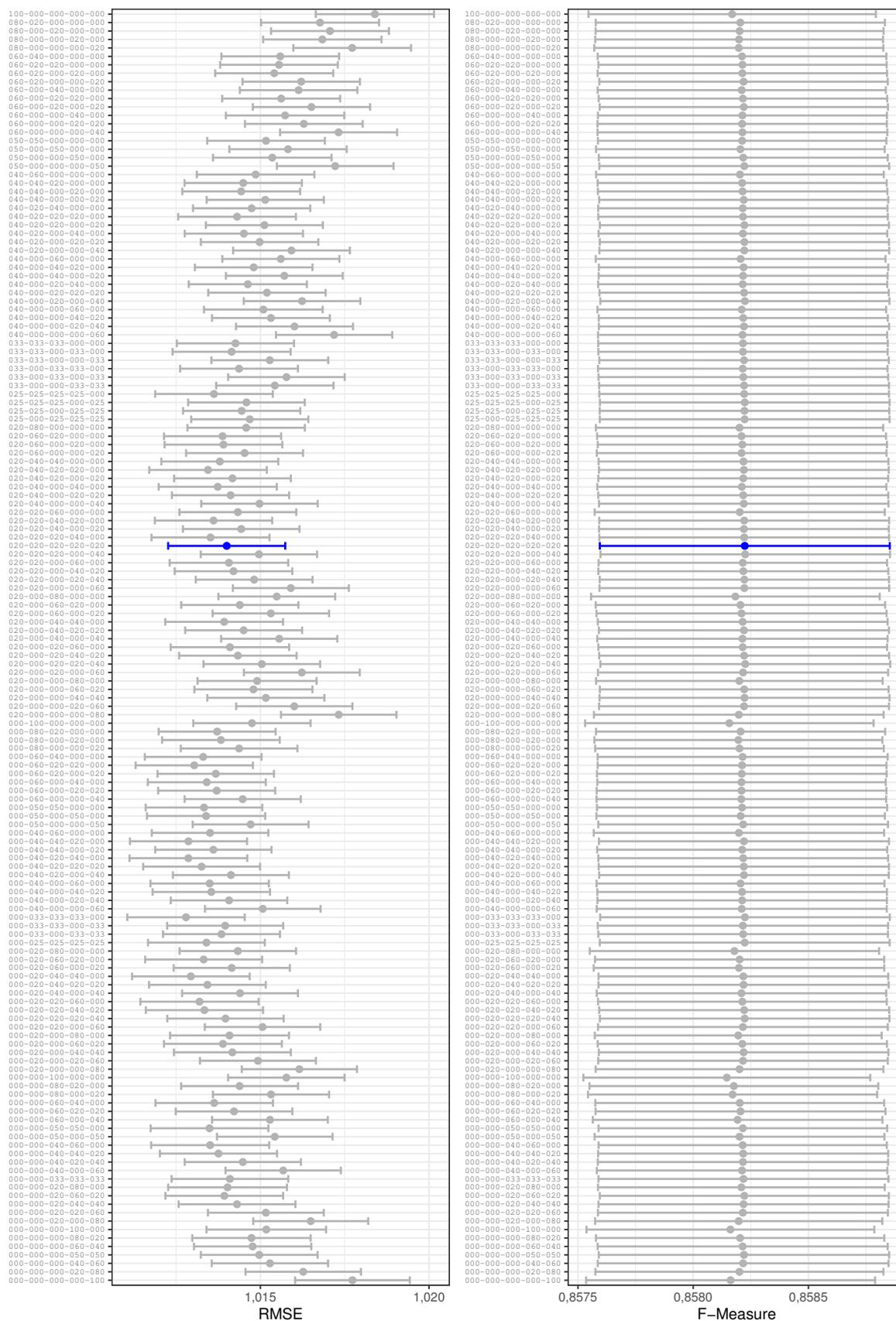
Fonte: autoria própria.

Figura F.2: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo ItemKNN e o conjunto de dados TripAdvisorCompleto.



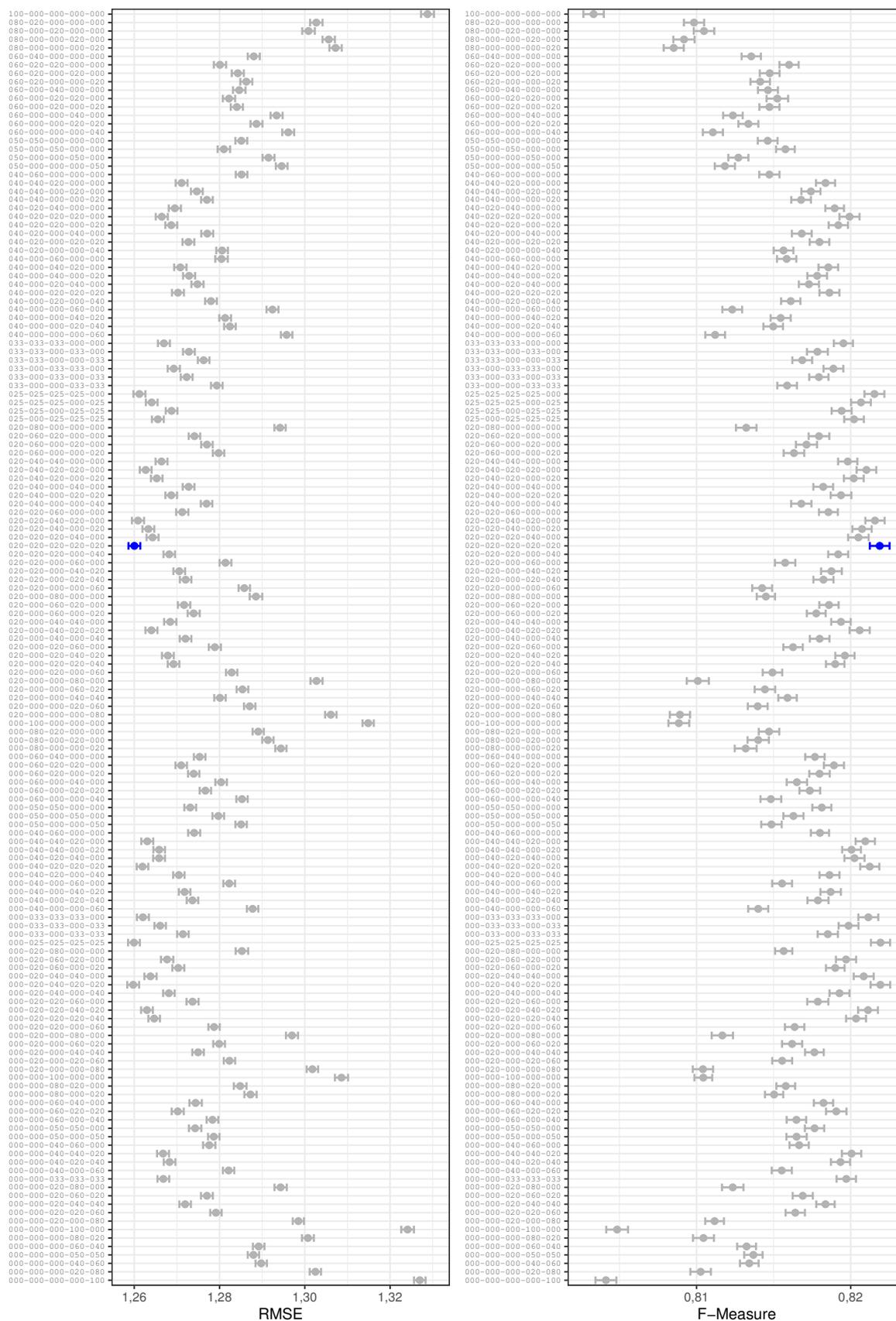
Fonte: autoria própria.

Figura F.3: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SMF e o conjunto de dados TripAdvisorCompleto.



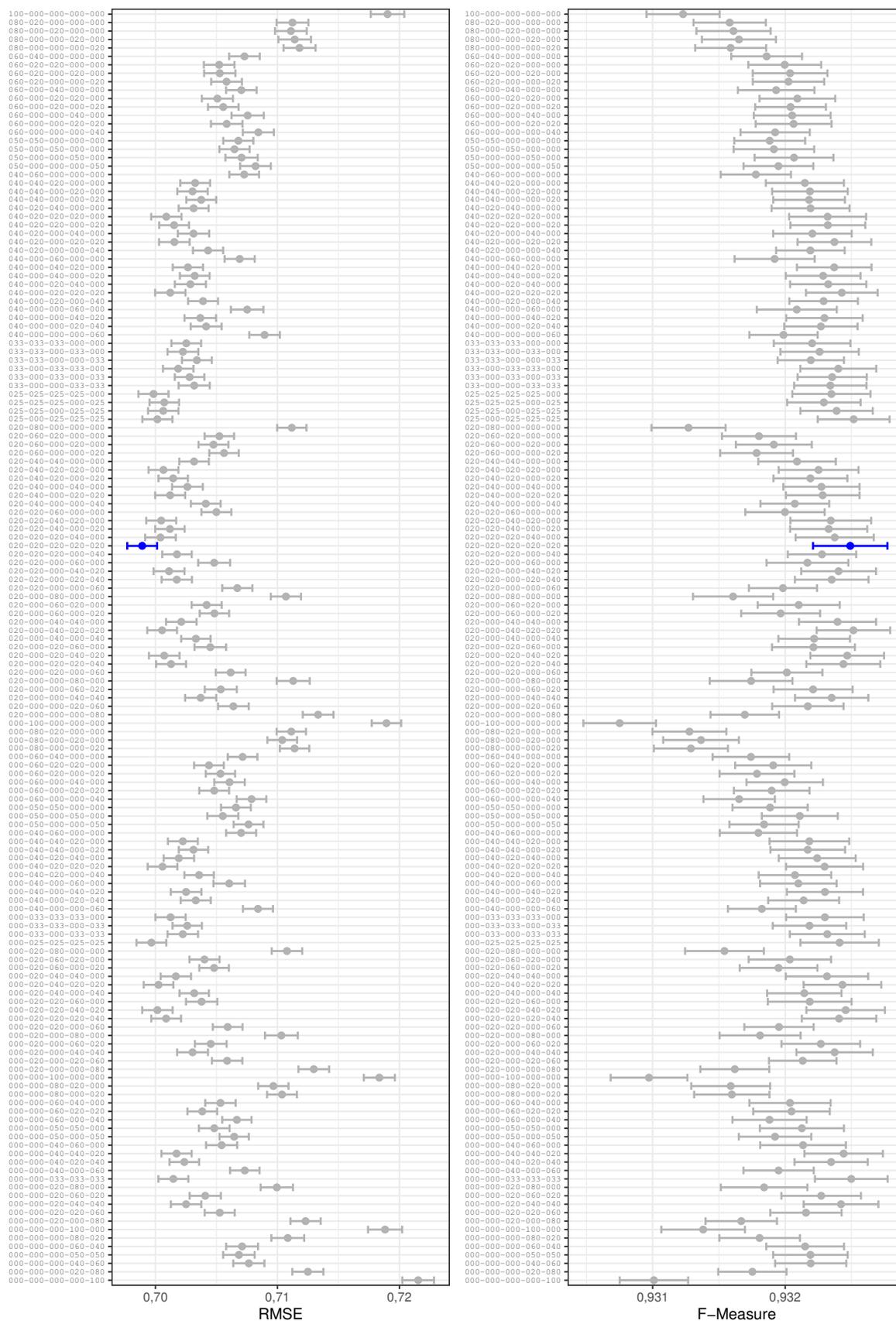
Fonte: autoria própria.

Figura F.4: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo UserKNN e o conjunto de dados AmazonFoods.



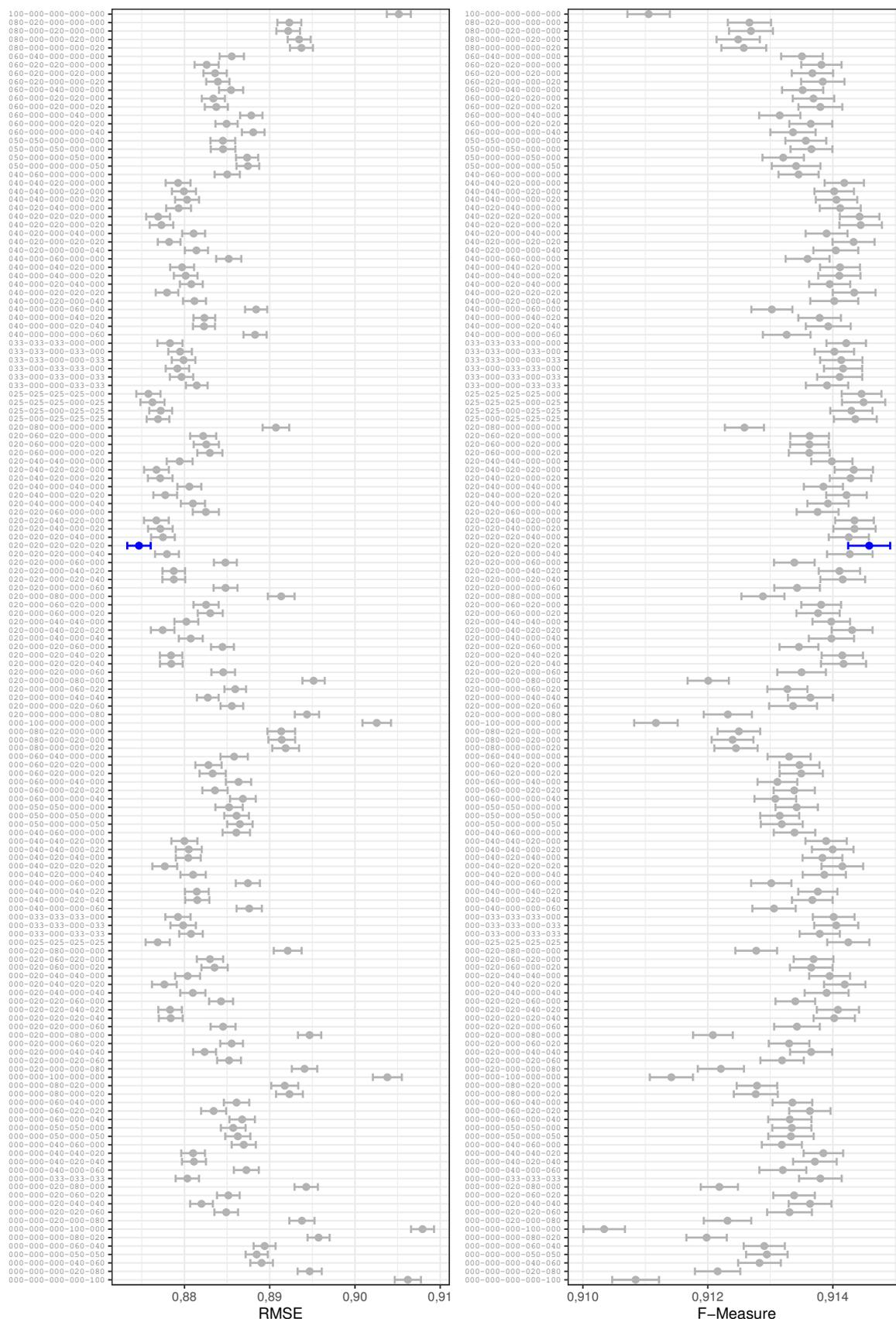
Fonte: autoria própria.

Figura F.5: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo ItemKNN e o conjunto de dados AmazonFoods.



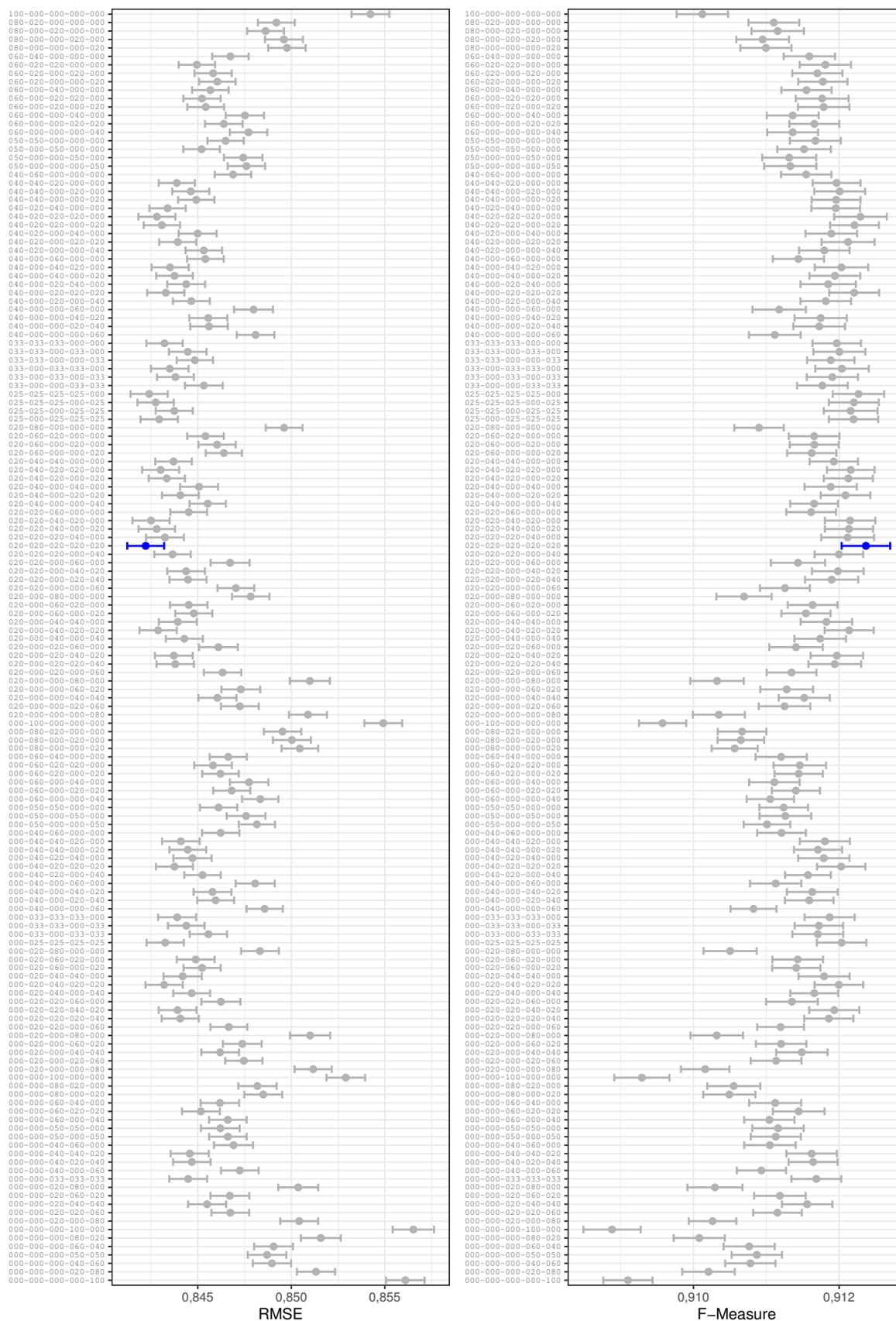
Fonte: autoria própria.

Figura F.6: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SMF e o conjunto de dados AmazonFoods.



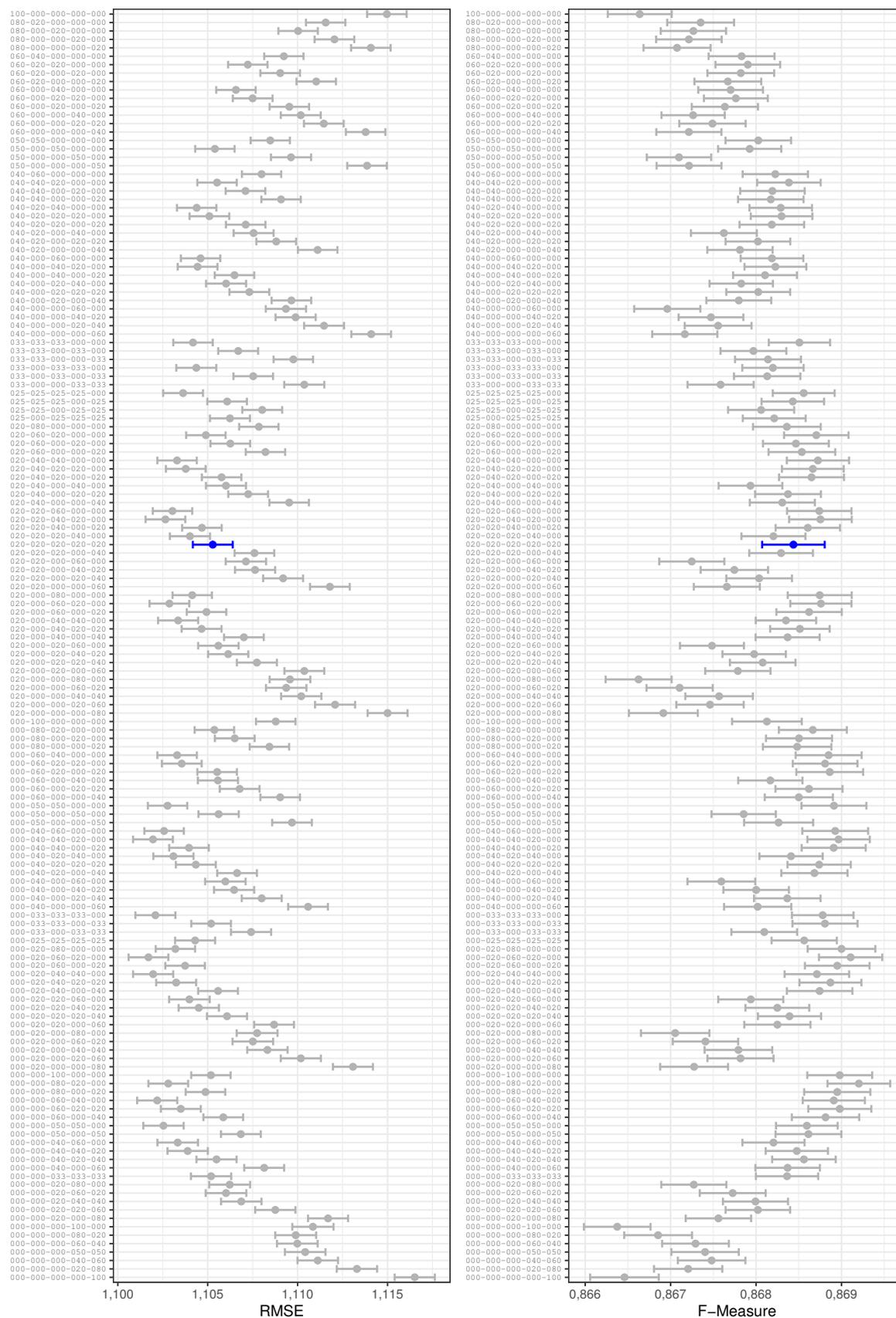
Fonte: autoria própria.

Figura F.7: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo BMF e o conjunto de dados AmazonFoods.



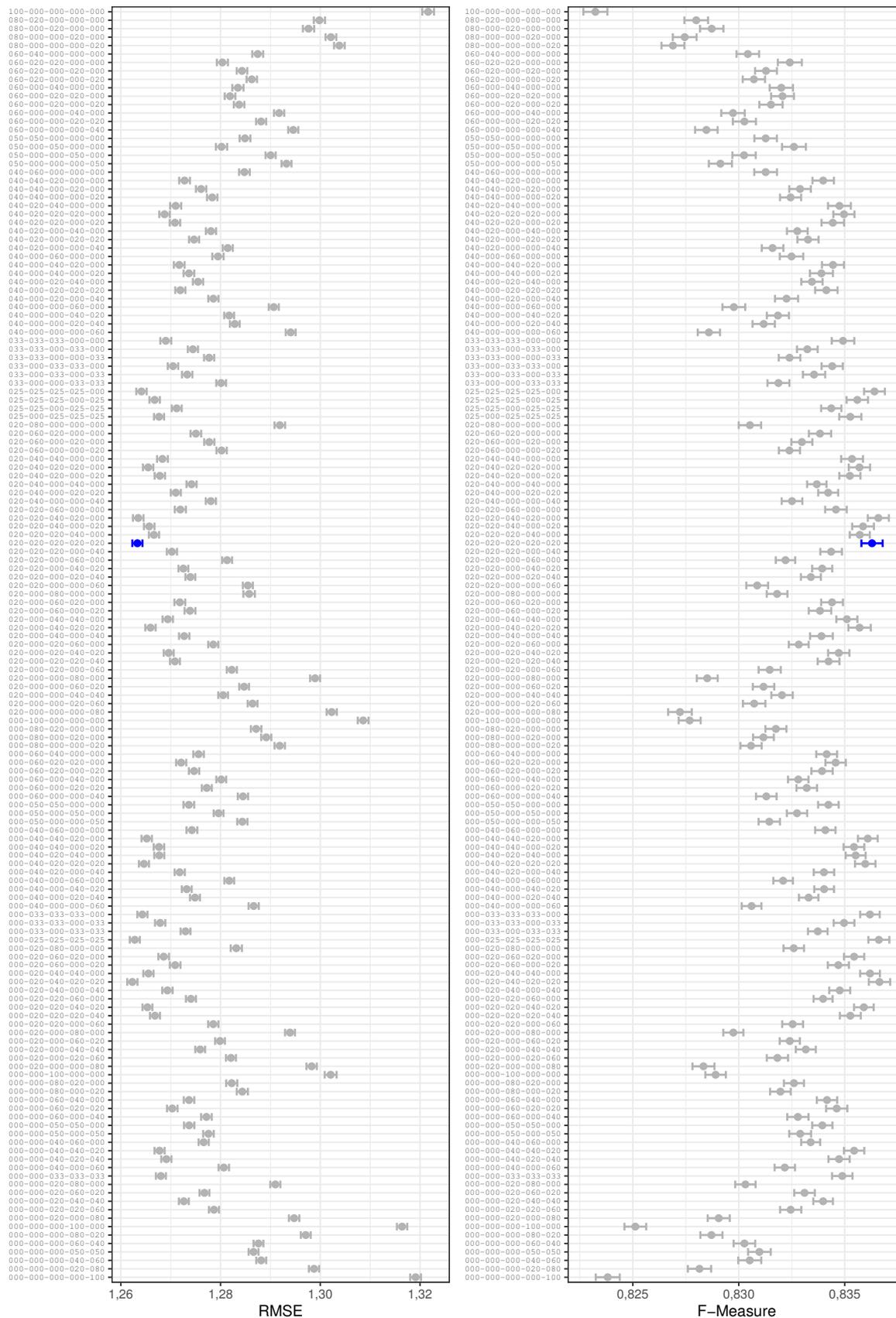
Fonte: autoria própria.

Figura F.8: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SVD++ e o conjunto de dados AmazonFoods.



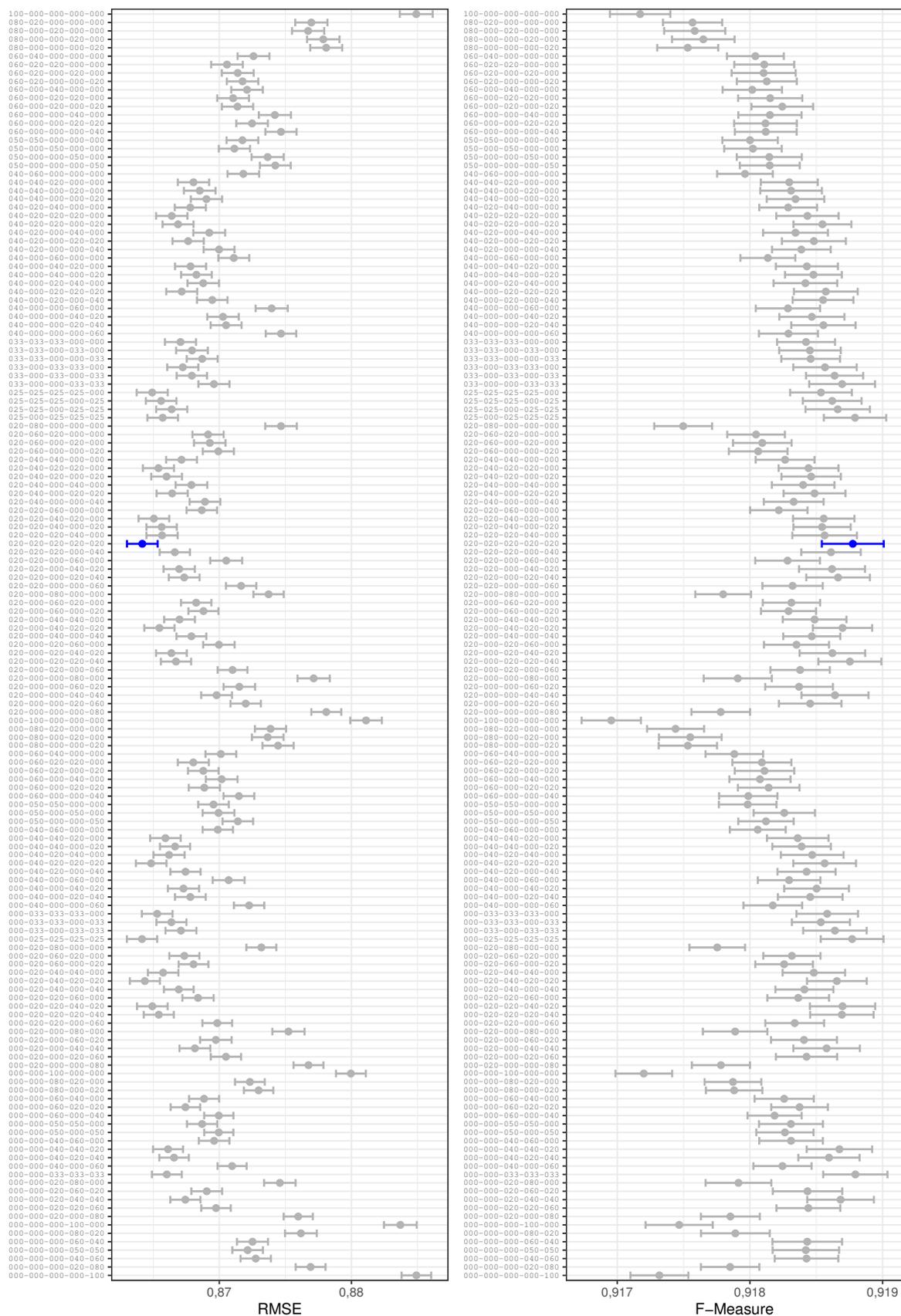
Fonte: autoria própria.

Figura F.9: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo UserKNN e o conjunto de dados AmazonFoodsCompleto.



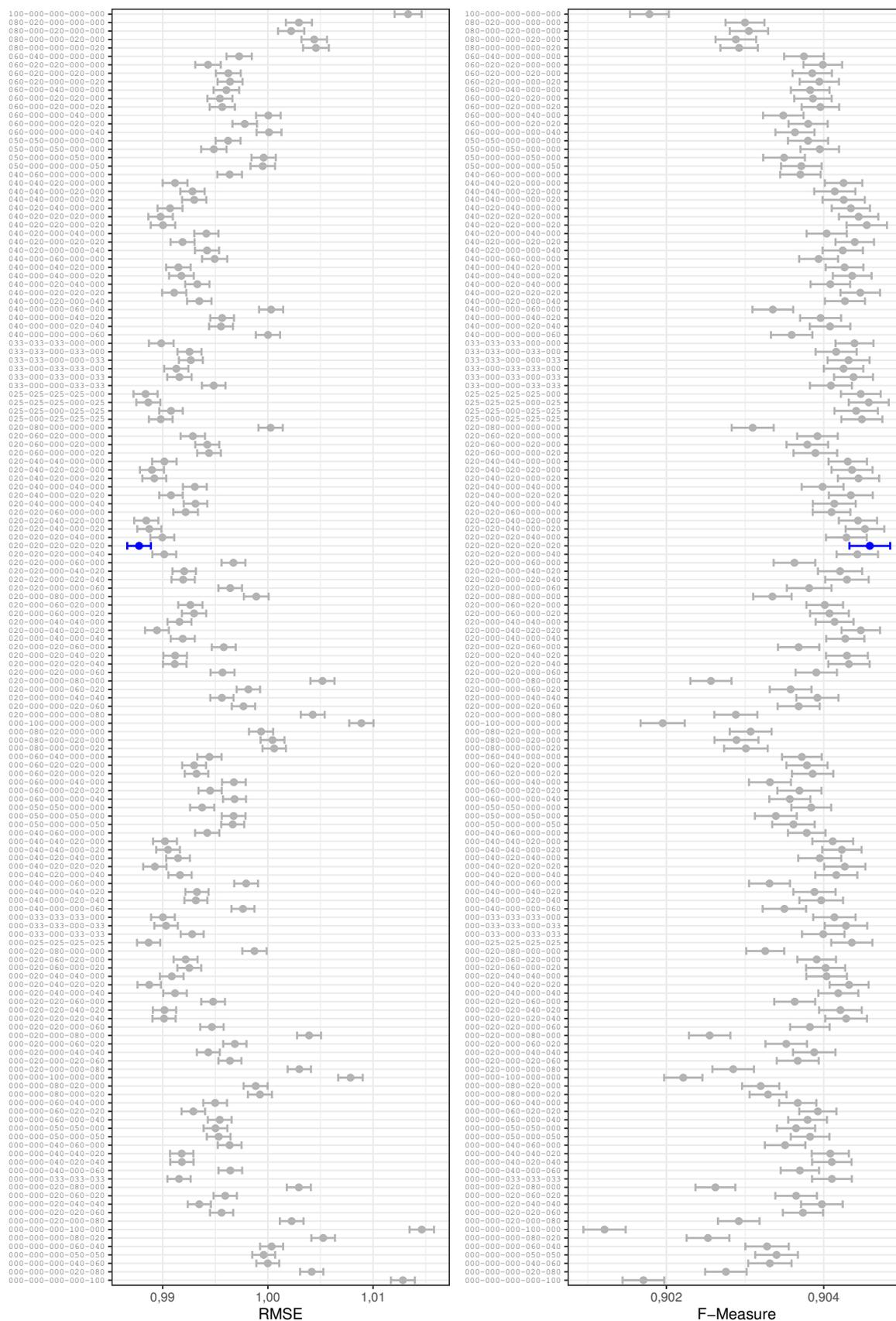
Fonte: autoria própria.

Figura F.10: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo ItemKNN e o conjunto de dados AmazonFoodsCompleto.



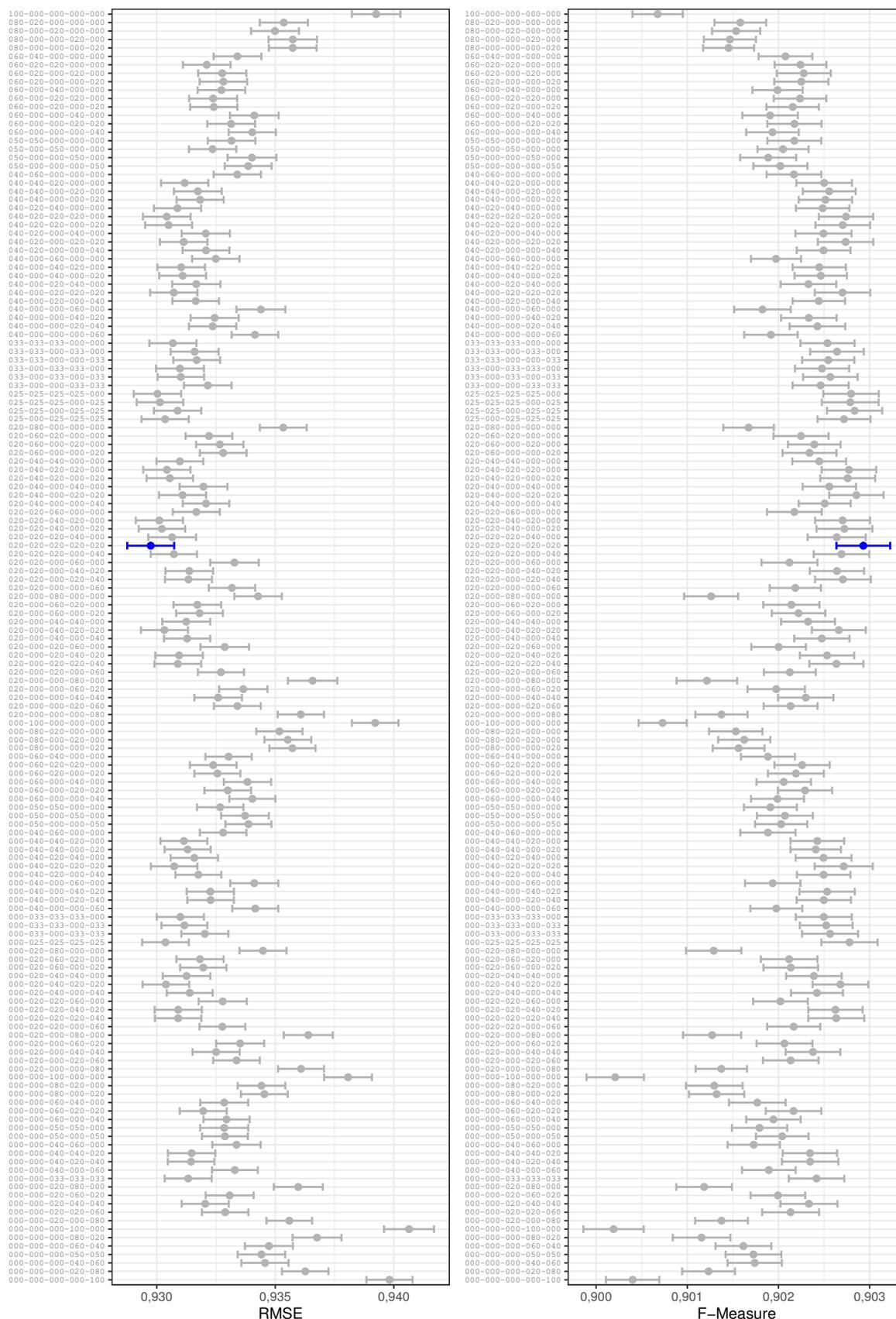
Fonte: autoria própria.

Figura F.11: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SMF e o conjunto de dados AmazonFoodsCompleto.



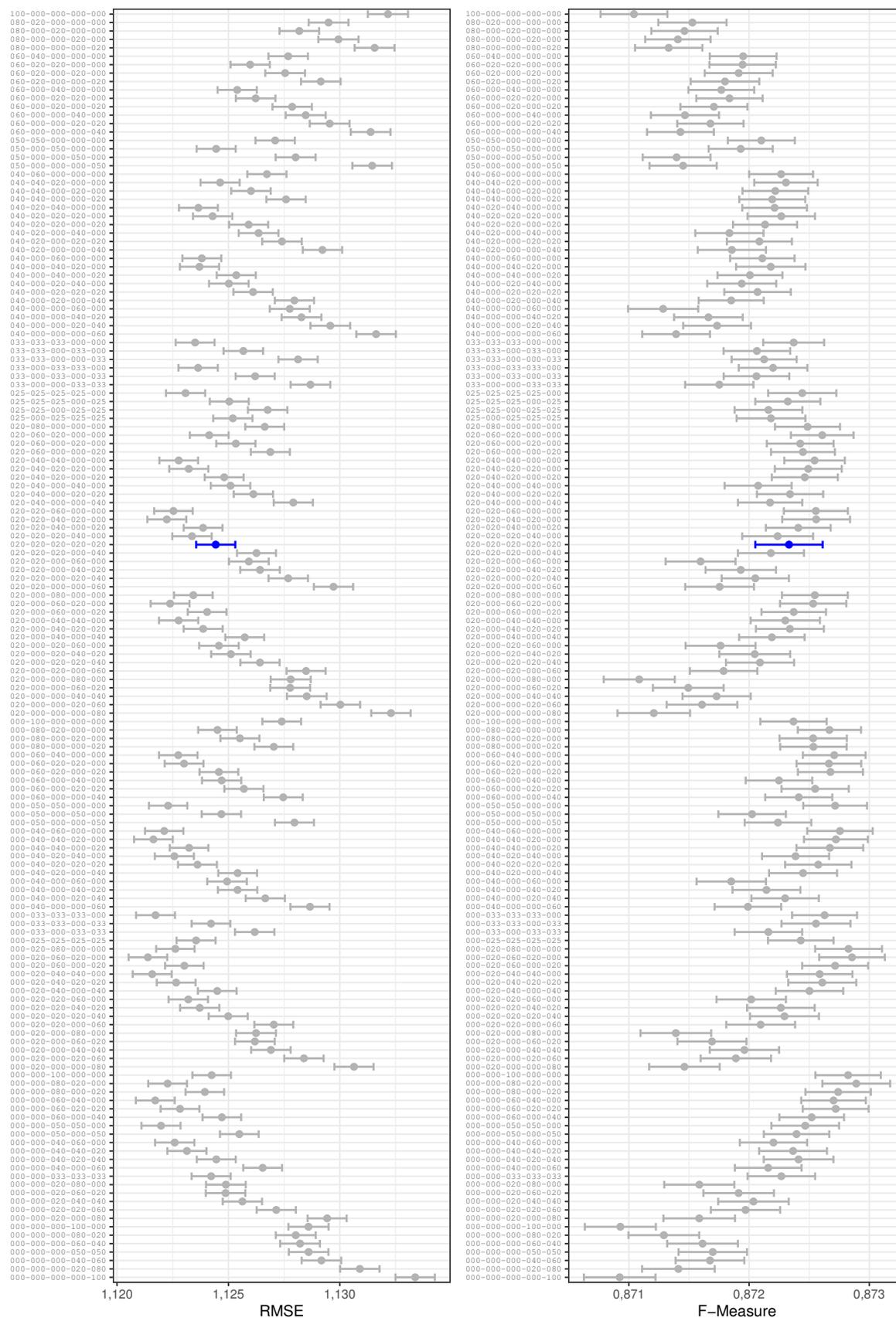
Fonte: autoria própria.

Figura F.12: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo BMF e o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Figura F.13: Intervalos de confiança referentes à QP3 (variação inicial) com o algoritmo SVD++ e o conjunto de dados AmazonFoodsCompleto.



Fonte: autoria própria.

Figura F.14: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo UserKNN e o conjunto de dados AmazonFoods.

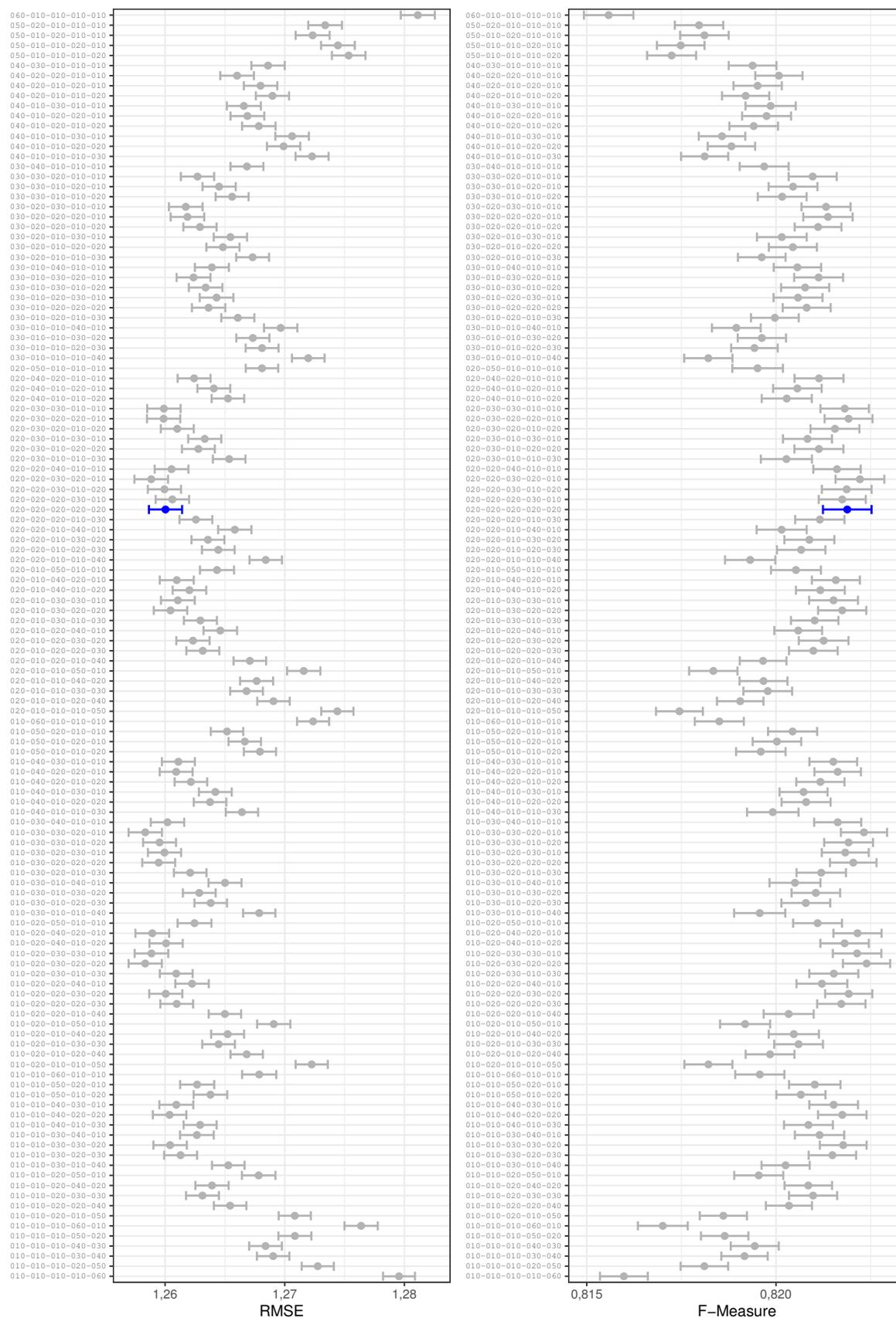
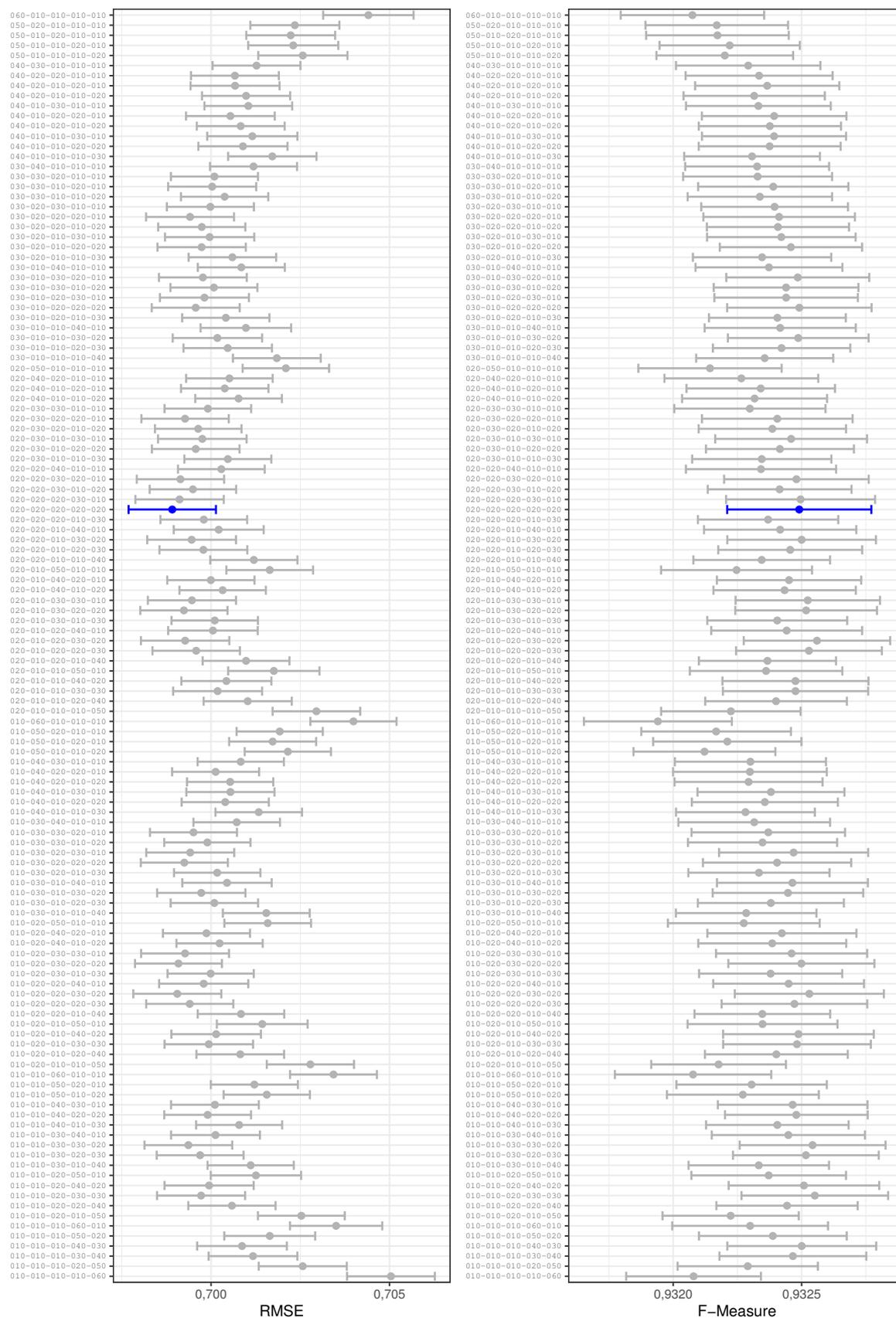
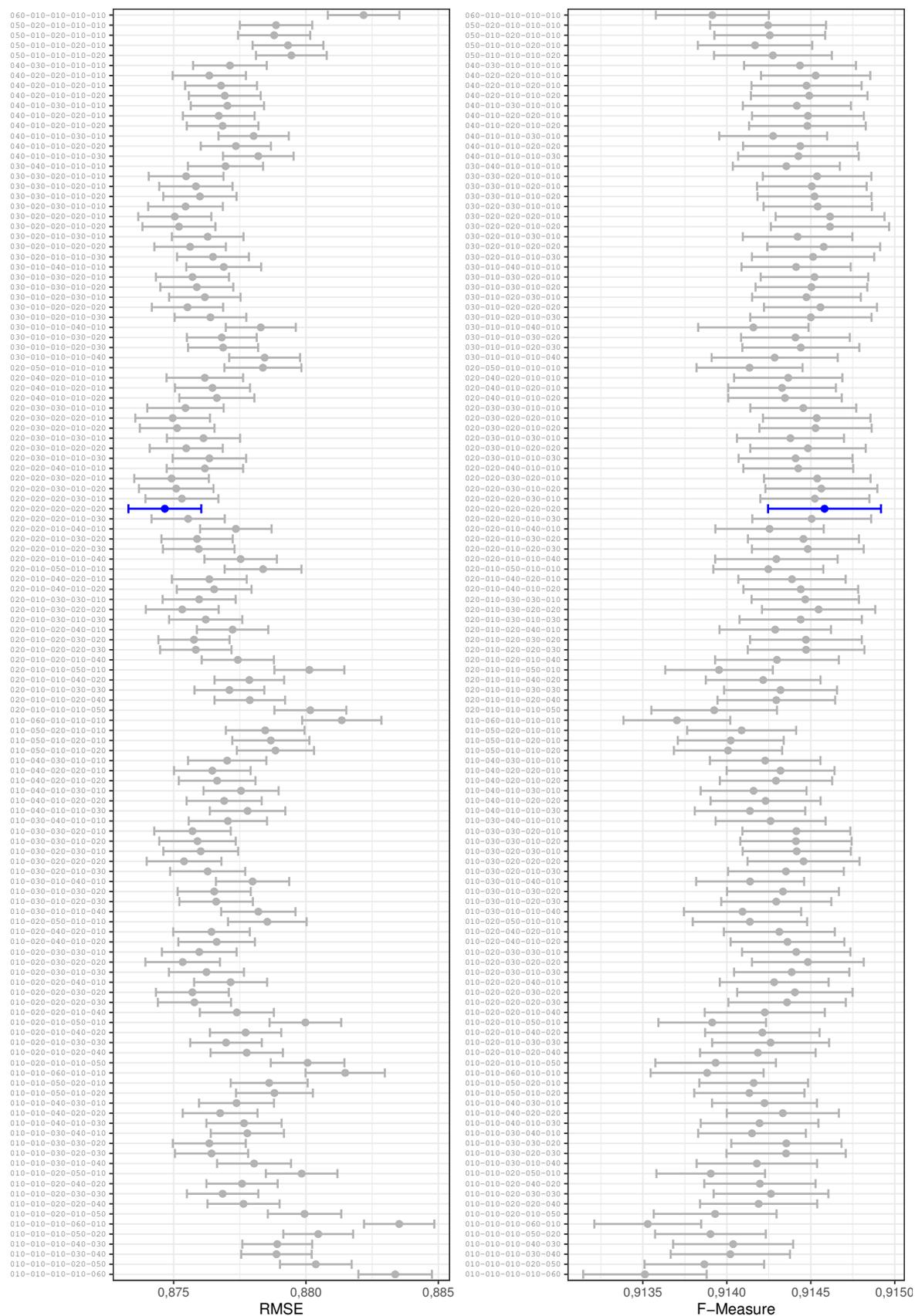


Figura F.15: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo ItemKNN e o conjunto de dados AmazonFoods.



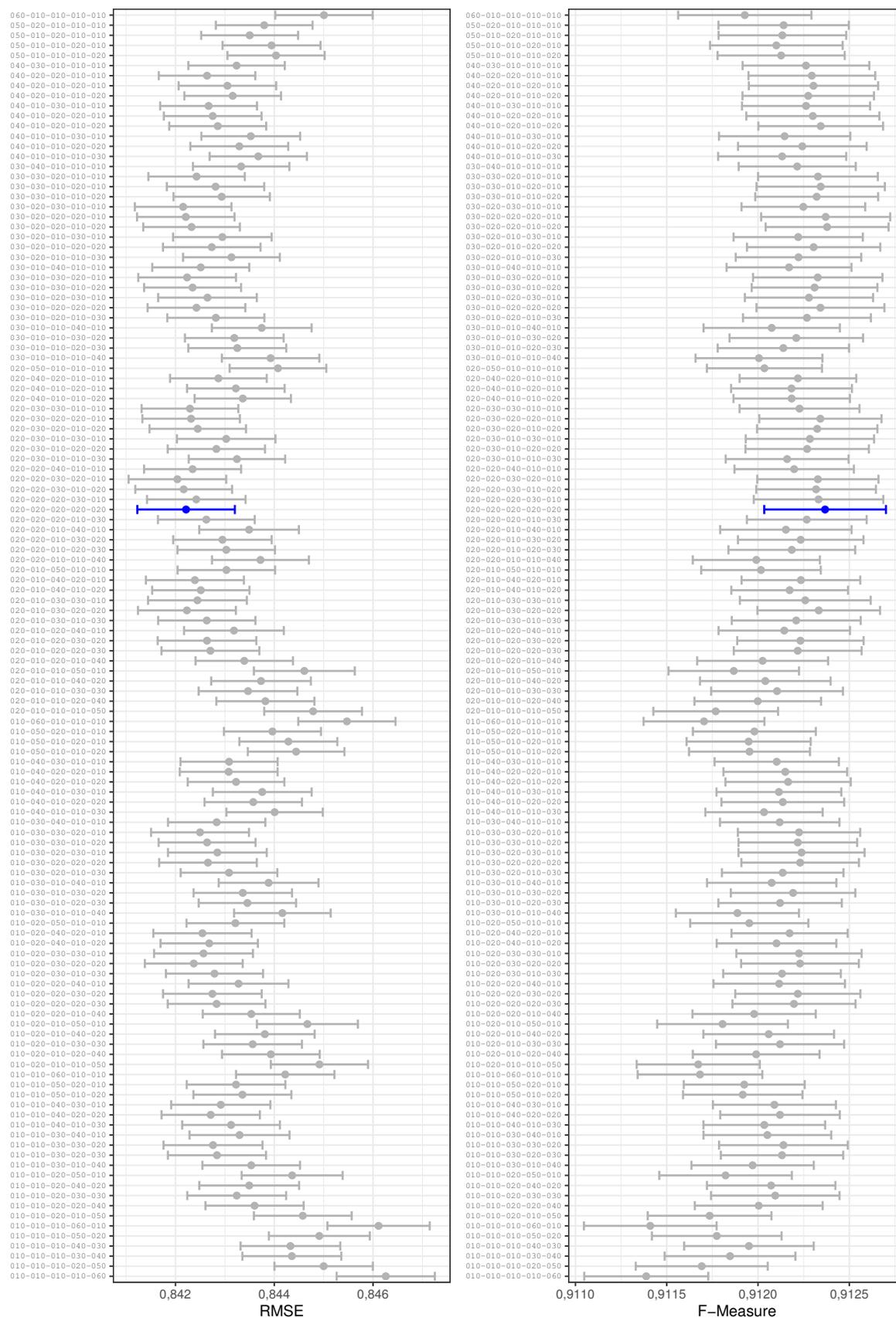
Fonte: autoria própria.

Figura F.16: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SMF e o conjunto de dados AmazonFoods.



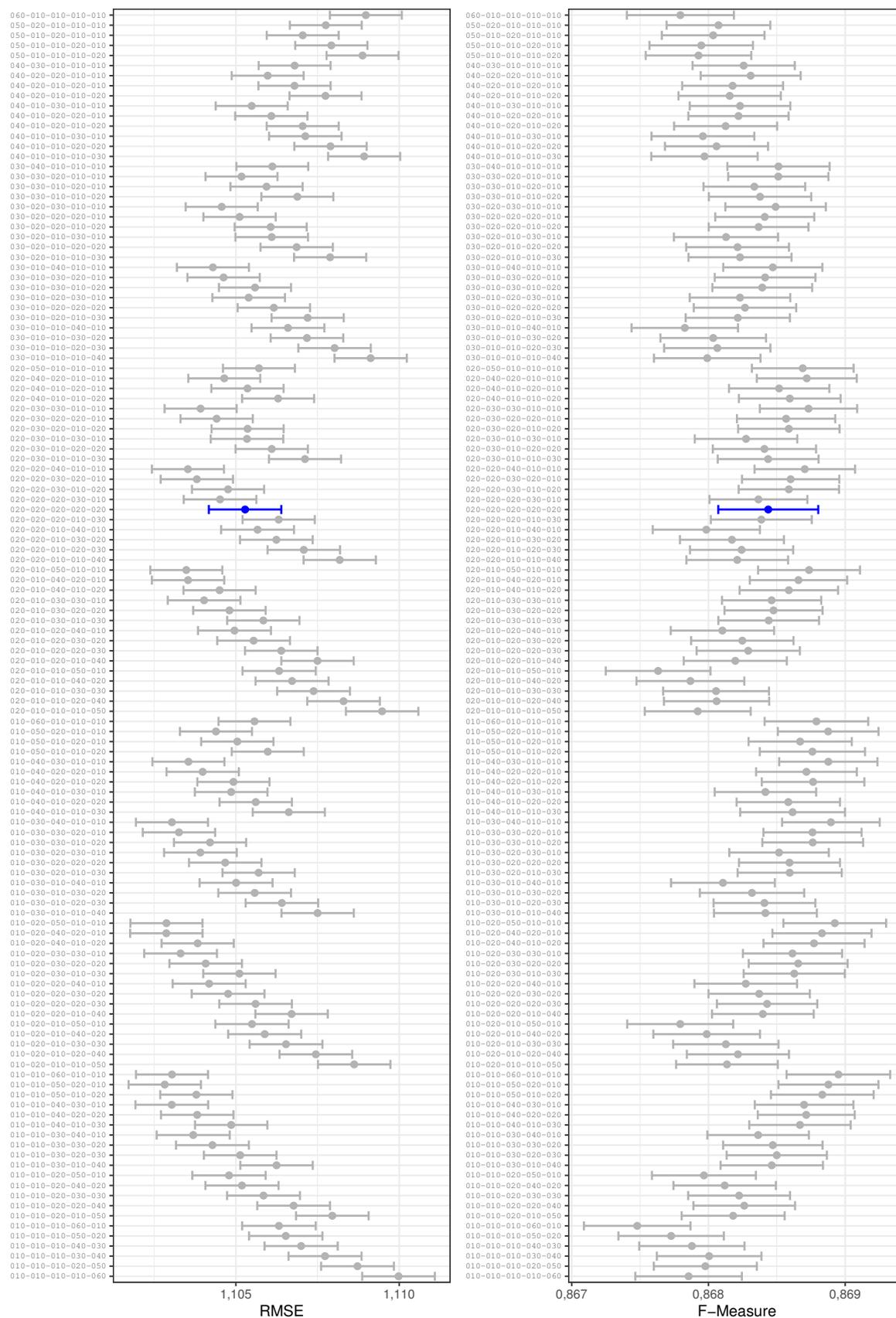
Fonte: autoria própria.

Figura F.17: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo BMF e o conjunto de dados AmazonFoods.



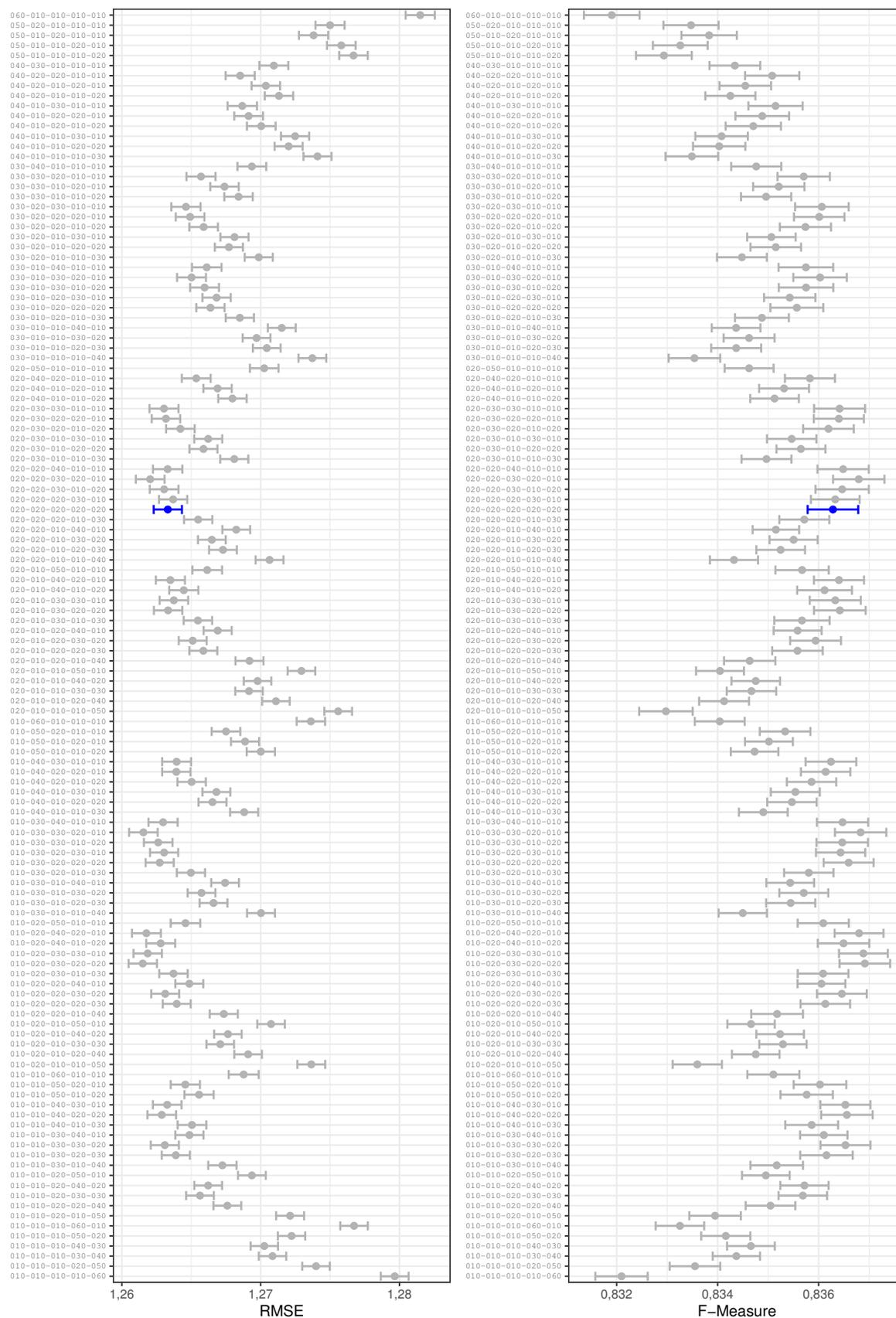
Fonte: autoria própria.

Figura F.18: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SVD++ e o conjunto de dados AmazonFoods.



Fonte: autoria própria.

Figura F.19: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo UserKNN e o conjunto de dados AmazonFoodsCompleto.

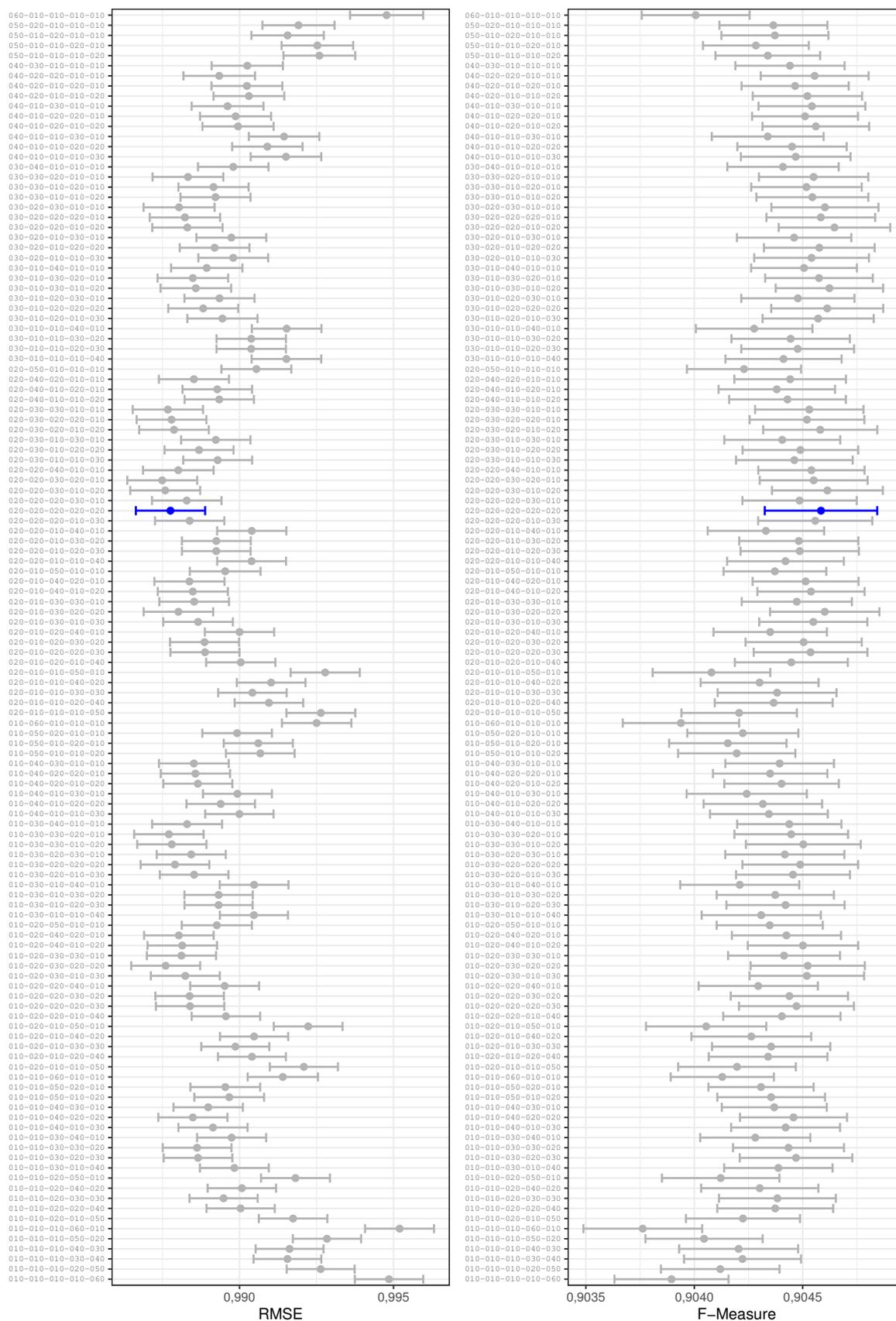


Fonte: autoria própria.

Figura F.20: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo ItemKNN e o conjunto de dados AmazonFoodsCompleto.

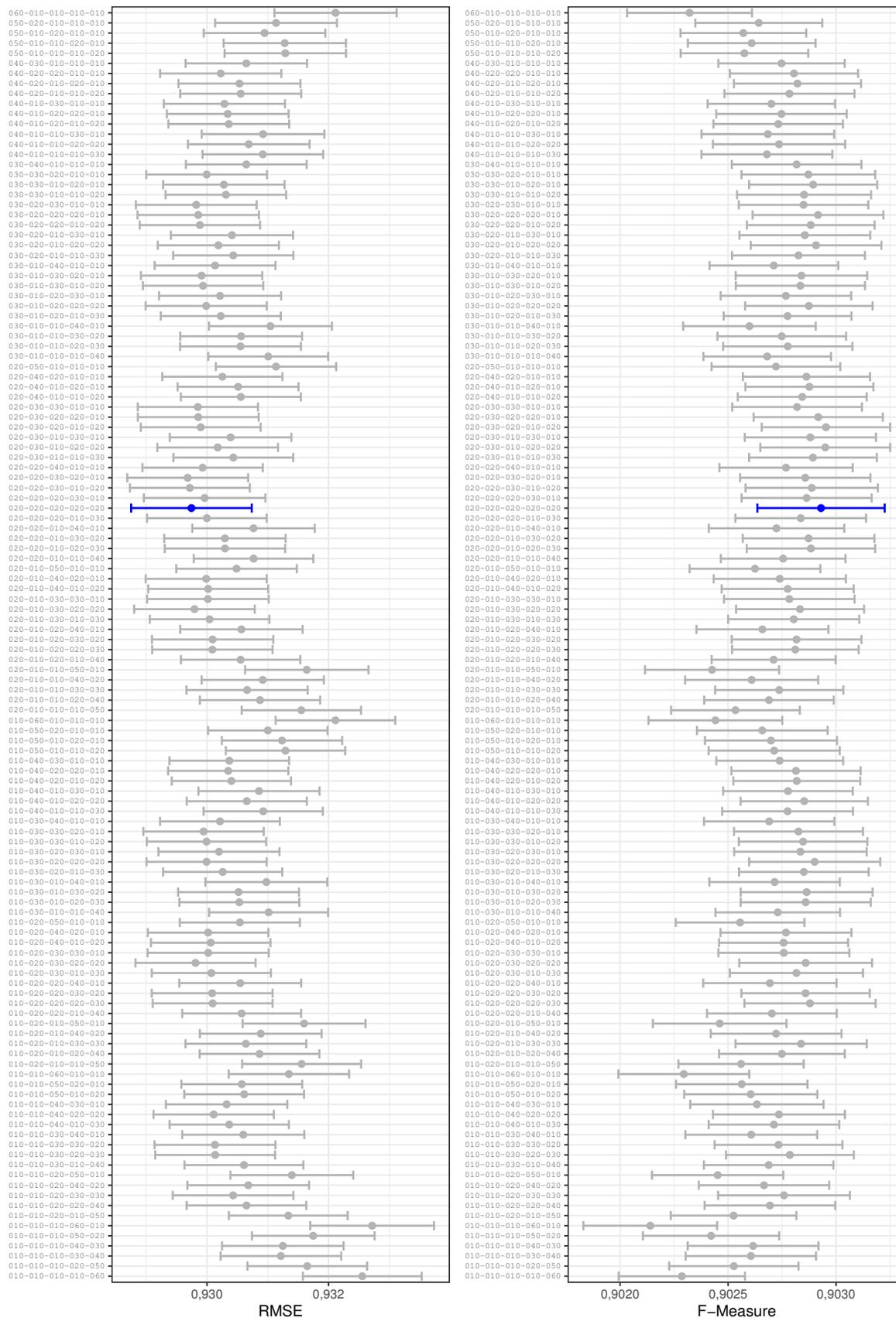


Figura F.21: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SMF e o conjunto de dados AmazonFoodsCompleto.



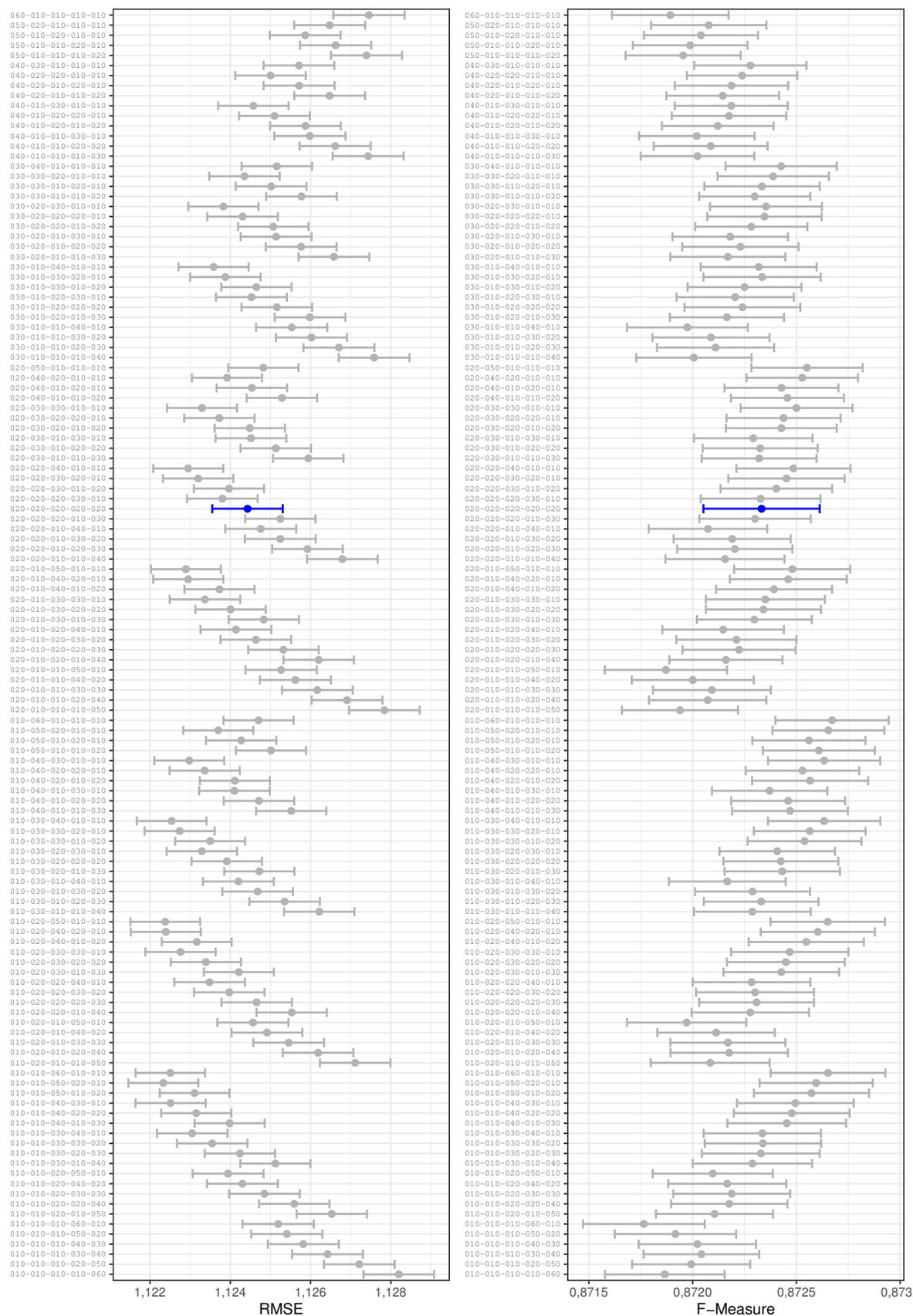
Fonte: autoria própria.

Figura F.22: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo BMF e o conjunto de dados AmazonFoodsCompleto.



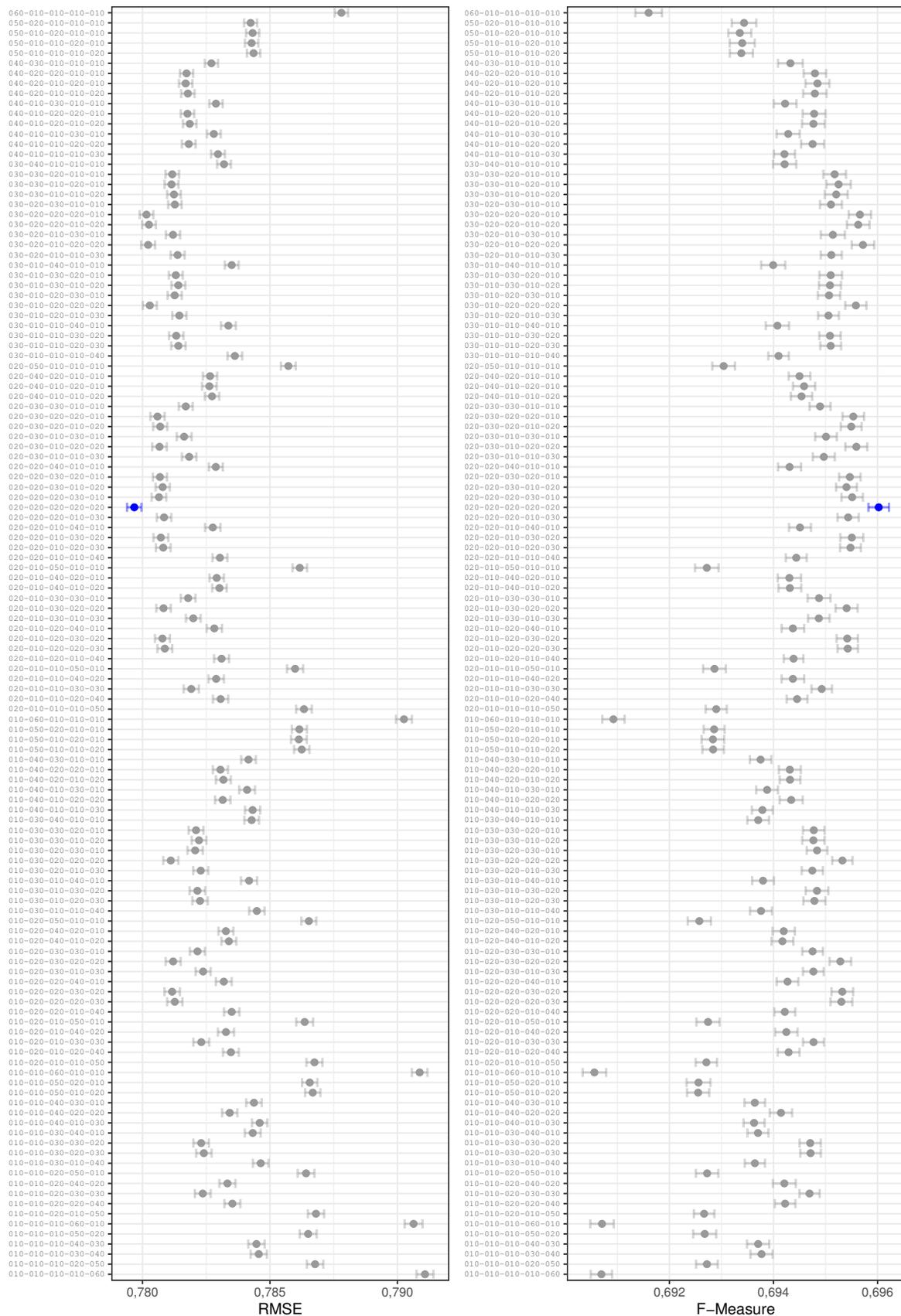
Fonte: autoria própria.

Figura F.23: Intervalos de confiança referentes à QP3 (variação complementar) com o algoritmo SVD++ e o conjunto de dados AmazonFoodsCompleto.



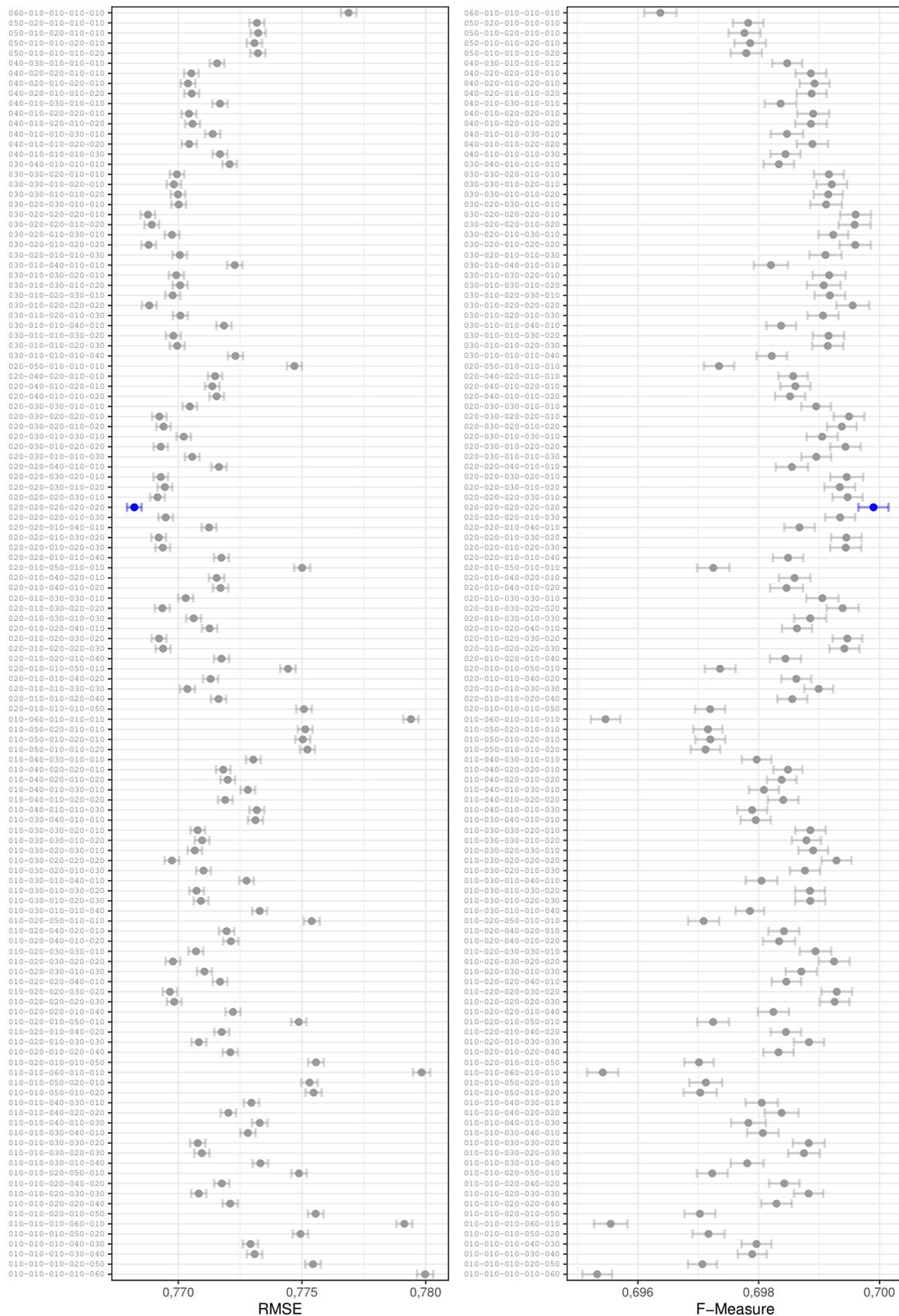
Fonte: autoria própria.

Figura F.24: Intervalos de confiança referentes à QP3 (análise complementar) com o algoritmo SMF e o conjunto de dados Personality2018.



Fonte: autoria própria.

Figura F.25: Intervalos de confiança referentes à QP3 (análise complementar) com o algoritmo BMF e o conjunto de dados Personality2018.



Fonte: autoria própria.