UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

LUCAS CORDEIRO BRASIL

COMPARAÇÃO ENTRE MODELOS COM DIFERENTES ABORDAGENS PARA CLASSIFICAÇÃO DE FAKE NEWS

CAMPINA GRANDE - PB

2021

LUCAS CORDEIRO BRASIL

COMPARAÇÃO ENTRE MODELOS COM DIFERENTES ABORDAGENS PARA CLASSIFICAÇÃO DE FAKE NEWS

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Professor Dr. Cláudio de Souza Baptista

CAMPINA GRANDE - PB

2021

B823c Brasil, Lucas Cordeiro.
Comparação entre modelos com diferentes abordagens para classificação de fake news. / Lucas Cordeiro Brasil. – 2021.

12 f.

Orientador: Prof. Dr. Cláudio de Souza Baptista.
Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Fake news. 2. Aprendizagem de máquina. 3. BERT. 4. Algoritmos. 5. Naive Bayes. 6. XGBoost. 7. Transformers. I. Baptista, Cláudio de Souza. II. Título.

CDU:004(045)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

LUCAS CORDEIRO BRASIL

COMPARAÇÃO ENTRE MODELOS COM DIFERENTES ABORDAGENS PARA CLASSIFICAÇÃO DE FAKE NEWS

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA:

Professor Dr. Cláudio de Souza Baptista

Orientador – UASC/CEEI/UFCG

Professor Dr. Rohit Gheyi

Examinador – UASC/CEEI/UFCG

Professor Tiago Lima Massoni

Professor da Disciplina TCC – UASC/CEEI/UFCG

Trabalho aprovado em: 20 de Outubro de 2021.

CAMPINA GRANDE - PB

ABSTRACT

Currently, false news is increasingly in evidence. Such news can be defined as intentionally propagated non-truthful information. With the large use of social networks as a source of information, it becomes necessary to have greater control and detection of Fake News, efficiently and quickly. Thus, this work seeks to use algorithms already consolidated in the machine learning area - Naive Bayes, XGBoost and BERT - to create false news detection models, comparing the results obtained in each model with works previously carried out in the area that have the best result until now.

Comparação entre Modelos com Diferentes Abordagens para Classificação de Fake News

Lucas Cordeiro Brasil
Unidade Acadêmica de Sistemas e Computação Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil
lucas.brasil@ccc.ufcg.edu.br

Cláudio de Souza Baptista
Unidade Acadêmica de Sistemas e Computação Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil
baptista@computacao.ufcg.edu.br

Anderson Almeida Firmino
Unidade Acadêmica de Sistemas e Computação Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil
anderson.ccomp@gmail.com

RESUMO

Atualmente, notícias falsas estão cada vez mais em evidência. Pode-se definir tais notícias como informações não verídicas propagadas intencionalmente. Com o grande uso de redes sociais como fonte de informação, torna-se necessário o maior controle e detecção de Fake News, de forma eficaz e rápida. Assim, este trabalho busca utilizar algoritmos já consolidados na área de aprendizagem de máquina - Naive Bayes, XGBoost e BERT - para criar modelos de detecção de notícias falsas, comparando os resultados obtidos em cada modelo com trabalhos anteriormente realizados na área que tenham os melhores resultados até o momento.

Palavras-Chave

Fake News, Aprendizagem de Máquina, BERT, Classificação.

1. INTRODUÇÃO

Fake News são informações não verídicas divulgadas de forma intencional. Atualmente, são comumente utilizadas e propagadas no meio online, seja por veículos de comunicação não confiáveis ou por usuários de redes sociais.

É importante ressaltar que a intenção das notícias possibilita eliminar ambiguidades entre fake news e outros conceitos similares, que não serão abordados neste trabalho, sendo eles: notícias satíricas (não há intenção de enganar os leitores), rumores, teorias conspiratórias, desinformações (criadas sem intenção) e trotes(voltadas para enganar pessoas individuais).

A identificação de notícias falsas, de maneira rápida e eficaz, torna-se fundamental, visando evitar um maior alcance delas e possíveis danos oriundos de sua disseminação.

Para exemplificar, uma mulher foi espancada e morta por dezenas de pessoas, a partir de um boato gerado por uma página em rede social, alegando que ela sequestrava crianças para magia negra [1]. No âmbito político, Fake News tiveram um papel importante nas eleições presidenciais dos Estados Unidos de 2016 [2].

Assim, a detecção automática de notícias falsas é uma possível solução para minimizar os impactos das mesmas. Desse modo, é possível utilizar o aprendizado de máquina para treinar modelos capazes de detectar, a partir das palavras e do contexto utilizado nas notícias, quanto a sua veracidade.

Para este trabalho, serão utilizados variados algoritmos de aprendizagem de máquina, já consolidados na área de classificação de textos, incluindo alguns que foram utilizados na obtenção dos melhores resultados, até o momento, no que diz respeito à detecção de Fake News. Tais algoritmos serão base para criação de modelos para detecção de notícias falsas em português e inglês, realizando um comparativo entre os resultados obtidos com cada um deles e trabalhos realizados anteriormente.

O restante deste artigo está organizado como segue. Na seção 2, há uma descrição sobre a arquitetura de Transformers, que é o estado da arte em processamento de linguagem natural. Na seção 3. são discutidos trabalhos relacionados na área de detecção de Fake News. Na seção 4, é apresentada a metodologia utilizada nesta pesquisa. Na seção 5, são discutidos os resultados obtidos. Por fim, na seção 6, apresentam-se as conclusões e direcionamentos para futuros trabalhos.

2. FUNDAMENTAÇÃO TEÓRICA

Foram utilizados neste trabalho algoritmos tradicionais de aprendizagem de máquina como o Naive Bayes e o XGBoost, como também técnicas de deep learning do estado da arte em processamento de linguagem natural, a saber a técnica de Transformers. Nesta seção, serão explicados alguns temas acerca da funcionalidade dos Transformers, assim como o conceito por trás do modelo BERT, uma das implementações mais usadas de Transformers.

2.1 Transformers

Transformer é uma arquitetura de rede neural criada por Vaswani et al [15] que, utilizando unicamente mecanismos de atenção, conseguiu simplificar os modelos arquiteturais base para a realização de tarefas de processamento de linguagem natural. Também foi possível, usando Transformers, conseguir resultados ainda melhores do que os alcançados anteriormente e com tempo de treinamento reduzido.

A seguir, é apresentado como o mecanismo de atenção dos Transformers atua e como são utilizados o Encoder e o Decoder dentro de sua estrutura.

2.1.1 Mecanismo de Atenção

Os mecanismos de atenção são parte fundamental do funcionamento dos Transformers, indicando quais palavras da sentença de entrada devem ser consideradas mais relevantes para o entendimento de uma palavra. Será apresentado a seguir como o Self-Attention é usado e sua expansão para o Multi-Head Attention.

2.1.1.1 Self-Attention

Ao receber uma sentença, para cada palavra são calculados três vetores - Query, Key e Value. Estes vetores são obtidos a partir do embedding da palavra multiplicado por três matrizes obtidas durante o treinamento, que são abstrações utilizadas pelo Transformer.

A partir dos vetores anteriores, é necessário calcular uma pontuação para cada palavra, realizado pelo produto escalar entre o vetor Query da palavra avaliada com o vetor Key de todas as palavras da sentença. Essa pontuação é fundamental para determinar o foco dado para outras partes da sentença ao processar uma palavra específica em sua posição determinada.

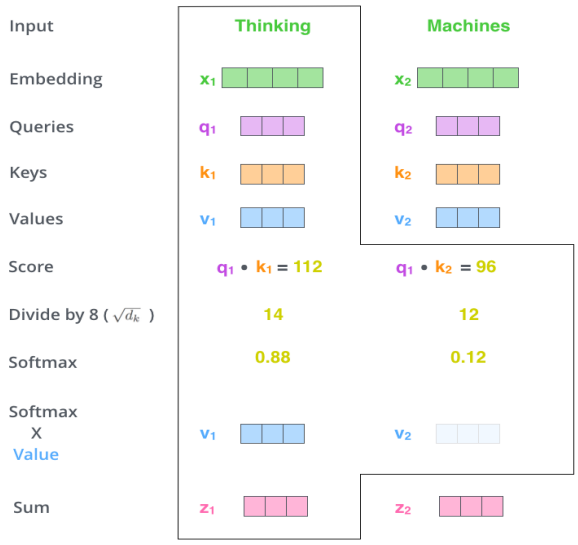
Em seguida, ocorre a divisão das pontuações pela raiz quadrada do tamanho do vetor Key utilizado e realizando uma função softmax sobre os valores obtidos.

Na etapa seguinte, multiplica-se o vetor Value pelos valores de saída da função softmax, buscando manter as medidas das palavras mais relevantes intactas e as medidas das palavras não relevantes serem altamente reduzidas.

Por fim, são somados os vetores ponderados de valores, produzindo então a saída do mecanismo de atenção para a palavra analisada.

Na figura 1, encontra-se um resumo de todos esses procedimentos realizados para a determinação de atenção de cada palavra.

Figura 1: Processo de cálculo de atenção nos Transformers



Fonte: Retirado de [16]

É importante mencionar que no modelo Transformer, o cálculo de self-attention é realizado de uma única vez para todas as palavras, de modo que cada vetor que representa as palavras é agrupado em uma matriz de embeddings, que ao ser multiplicada pelas matrizes Query, Key e Value geram representações para cada uma das palavras da sentença.

2.1.1.2 Multi-Head Attention

Dentro da execução dos Transformers, o conceito de Self-Attention é ampliado, sendo utilizado com múltiplas matrizes diferentes de Query, Key e Value para a mesma palavra, formando assim várias representações diferentes, de acordo com diferentes abordagens de atenção, surgindo assim o conceito de Multi-Head Attention.

Nessa abordagem, além de ter mais de uma representação para a palavra, com diferentes atenções, também propicia que o modelo foque em diferentes posições da sentença.

Por fim, cada matriz de atenção calculada é concatenada e seu resultado é multiplicado por uma matriz de pesos adicional, formando então uma matriz final com as informações das diferentes atenções para cada palavra.

2.1.2 Encoder/Decoder

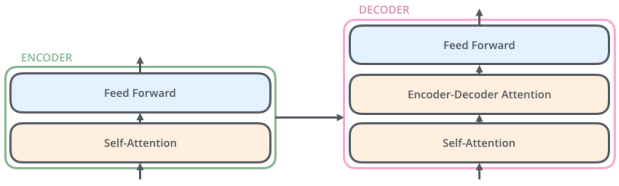
Em uma visão mais abrangente dos Transformers, eles são formados basicamente por um conjunto de igual número de Encoders e Decoders.

Os Encoders são formados por uma camada de atenção e por uma camada de rede neural Feed Forward. Cada Encoder recebe os embeddings que representam a sentença como entrada, calcula os vetores de atenção para cada palavra e então esses vetores resultantes são passados para a rede neural, que após o aprendizado, envia sua saída para o próximo Encoder.

Nos Decoders, há uma estrutura similar, porém há uma camada intermediária a mais, chamada de Encoder-Decoder Attention, que indica quais partes da sentença são mais relevantes e devem receber maior atenção.

Na figura 2 pode-se visualizar como estão estruturados os Encoders e Decoders, assim como suas camadas internas.

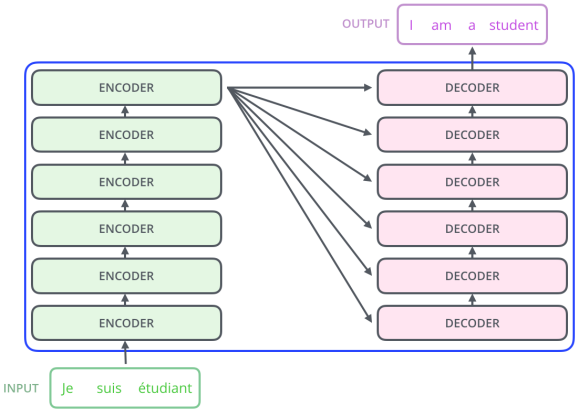
Figura 2: Estrutura interna dos Encoders e Decoders



Fonte: Retirado de [16]

Como funcionamento geral dos Transformers, pode-se observar que o primeiro Encoder recebe os embeddings da sentença e após a execução do mecanismo de atenção e da rede neural feed forward, o resultado é passado para o próximo Encoder, que realiza o mesmo processo até chegar ao último Encoder. A saída final é então passada para todos os Decoders, que também recebem como entrada a saída do Decoder anterior, assim como observa-se na figura 3.

Figura 3: Estrutura geral do Transformer



Fonte: Retirado de [16]

2.2 BERT

Uma das implementações mais populares de Transformers é o BERT - Bidirectional Encoder Representations from Transformers [17]. O BERT é um modelo de representação de linguagem proposto por Devlin et al. [17]. O BERT possui apenas a camada de Encoder e tem sido utilizado em várias aplicações de processamento de linguagem natural.

Durante o treinamento, destaca-se duas etapas principais utilizadas pelo BERT: Masked LM e Next Sentence Prediction. Além disso, é necessário realizar o fine-tuning do modelo para realização de tarefas específicas.

2.2.1 Masked LM

Nesta etapa, ao receber uma sequência de palavras, o modelo substitui 15% delas por um token especial [MASK]. Assim, o modelo tenta deduzir quais os valores originais de tais palavras mascaradas, de acordo com o contexto das outras palavras não mascaradas da sentença.

Pode-se citar alguns passos para a predição de tais palavras como uma sequência na qual é utilizada uma camada de classificação na saída do encoder, multiplica-se os vetores de saída pela matriz de embedding e calcula-se a probabilidade de cada palavra no vocabulário, utilizando uma função softmax.

2.2.2 Next Sentence Prediction

Durante o processo de treinamento, são recebidos pelo modelo pares de sentença e o modelo aprende a predizer se a segunda sentença é a sequência imediata da primeira. Em todo o treinamento, metade das entradas consiste de sentenças seguidas e a outra metade com a segunda sentença escolhida aleatoriamente. Para tal tarefa, o BERT utiliza tokens especiais, como o [CSL], indicando o início da primeira sentença e [SEP] que indica a separação entre as duas sentenças.

Na predição se as sentenças estão conectadas, a primeira sentença completa é processada pelo Transformer. A saída do token [CSL] é então transformada em um vetor de forma 2x1, com uma camada de classificação simples e por fim a probabilidade de ser a próxima sentença é calculada.

No treinamento, ambas as etapas são realizadas em conjunto, visando minimizar a função de perda combinada das duas estratégias.

2.2.3 Fine-tuning

O modelo BERT pode ser utilizado em diversas tarefas, necessitando apenas de alguns ajustes no modelo original para obtenção de bons resultados.

Problemas de classificação são realizados de forma similar ao processo de predição de próxima sentença, adicionando-se uma camada de classificação na saída do token [CSL] do Transformer.

Em problemas de Responder Perguntas, o modelo recebe uma pergunta a respeito da sentença e deve marcar a resposta na sentença. Para tal, o BERT pode ser treinado com o aprendizado de dois novos vetores que marcam o começo e o fim da sentença.

Para tarefas de Reconhecimento de Entidades Nomeadas, é recebida uma sentença textual e o modelo então deve dizer quais os variados tipos de entidade daquela sentença (Pessoa, Organização, Data, etc). Nesses casos, o treinamento pode ser feito alimentando o vetor de saída de cada token em uma camada de classificação que prediz a entidade.

3. TRABALHOS RELACIONADOS

A área de Aprendizado de Máquina, aplicada à Fake News, tem crescido bastante nos últimos anos dentre os pesquisadores. Dessa forma, novas técnicas e modelos vêm sendo desenvolvidos, dando um maior suporte para novas pesquisas acerca deste tema.

Em Monteiro et al. [4], é proposto o primeiro corpus em língua portuguesa para a detecção de Fake News, chamado de Fake.Br, contendo 7200 registros, com uma divisão balanceada de dados verdadeiros e falsos, coletados manualmente pelos autores. Além disso, foram realizados experimentos utilizando o algoritmo SVM para diferentes características, com o melhor resultado sendo obtido com a combinação de bag of words e emotividade. Também foram utilizadas outras técnicas de aprendizado de máquina, como Naive Bayes, Random Forest e Multilayer Perceptron, este último com os melhores resultados, obtendo uma acurácia de 90%.

Já Moura, Silva e Cardoso [9] realizaram experimentos com Regressão Logística e Random Forest, com 87% e 91% de f1-score, respectivamente, utilizando extração de features. Além disso, eles utilizaram word embeddings do Multilingual BERT e do BERT, aplicando-os nos algoritmos anteriores, com resultados de f1-score de 86% e 90%, respectivamente.

No trabalho de Feijó e Moreira [10], foi realizado o pré-treinamento do BERT e do Albert, com uma grande massa de dados em português (992 milhões de tokens). Após o pré-treinamento, em testes realizados com os modelos, obteve-se 98% de f1-score para ambos os modelos na detecção de fake news.

Em Moraes et al. [5], os autores realizaram uma análise de notícias falsas brasileiras, identificando padrões de escrita nas mesmas. Para seus experimentos, foi utilizada uma base de dados que combinou os dados do corpus Fake.Br com alguns dados do corpus Fakepedia. Os dados foram classificados de acordo com as classes gramaticais das palavras, bem como o sentimento dos textos. Foram utilizados os algoritmos AdaBoost, Naive Bayes e SVM, com melhores resultados de acurácia de 93%, 84% e 94%, respectivamente. Menciona-se aqui a acurácia dos experimentos, visto que essa foi a única métrica exibida no trabalho.

Em outro trabalho realizado por Faustini e Covões [21], os autores utilizaram algoritmos clássicos de aprendizado de máquina, com os melhores resultados de 91% de f1-score para o SVM, 85% de f1-score para Naive Bayes e 88% de f1-score para Random Forest, estes com utilização de Bag of Words para extração de features dos dados.

Ainda pode-se citar o trabalho de Freire, Silva e Goldschmidt [22] que propõe uma abordagem chamada de Hybrid Crowd Signals (HCS) e, realiza experimentos com o corpus Fake.Br, obtendo resultados de f1-score de 99,9% com um dos métodos que aplicam a abordagem, o HCS-F.

Para experimentos com dados em inglês, pode-se citar o trabalho de Wu e Liu [6], que propõe um novo modelo, o TraceMiner. Assim, o algoritmo proposto pelos autores obteve f1-score de 91,24%, enquanto que o algoritmo SVM obteve 75,81% de f1-score e o XGBoost obteve 82,26% de f1-score.

Em Janze e Rizius [11], os autores também utilizaram XGBoost e SVM, com f1-score de 79,64% para o primeiro e 82,18% para o segundo.

Já em Kaliyar, Goswami e Narang [12], é implementada uma adaptação do BERT, chamada de FakeBert, com 98,9% de acurácia nos testes realizados. A acurácia é aqui mostrada, já que esta foi a única métrica utilizada pelos autores.

Gundapu e Mamidi [13] em seu trabalho, propuseram um novo modelo, Ensemble Model, que combina três modelos de transformers: BERT, XLNet e Albert, conseguindo 98,5% de f1-score, superando os demais modelos quando utilizados separadamente.

Por fim, pode-se citar o trabalho de Khan et al. [14], que utilizando três diferentes bases de dados em inglês , tanto com modelos de aprendizagem tradicional como modelos de pré-treinamento avançado, obtendo os melhores resultados com Naive Bayes (93% de f1-score), BERT (95% de f1-score) e RoBERTa (98% de f1-score).

Neste trabalho, serão utilizados alguns dos algoritmos que foram base para os resultados apresentados nesta seção (Naive Bayes, XGBoost e BERT), buscando comparar os resultados obtidos neste trabalho com os obtidos em trabalhos anteriores.

4. METODOLOGIA

Nesta seção, será apresentada a metodologia utilizada para o desenvolvimento dos experimentos realizados neste trabalho. Inicialmente, houve o processo de busca e obtenção de bases de dados já existentes na área de Fake News, em seguida a preparação dos dados utilizados, a modelagem realizada e a validação dos modelos treinados.

4.1 Bases de Dados

Em experimentos de aprendizado de máquina, é fundamental a escolha e uso de bases de dados de qualidade para a obtenção de resultados confiáveis. Devido ao grande interesse de pesquisadores sobre o tema de Fake News nos últimos anos, é possível encontrar bases de dados confiáveis para realização dos experimentos, tanto em língua portuguesa como em língua inglesa.

Inicialmente, foi encontrada a base de dados FACTCK.BR[7] e planejava-se utilizá-la para a realização dos experimentos. Entretanto, ao analisar o corpus, foi observado que ele não era balanceado, o que poderia comprometer os resultados [20]. Assim, foi buscada outra base que pudesse ser utilizada, e foi encontrado o corpus Fake.Br, publicado por Monteiro et al. [4]

Neste corpus existem 7.200 notícias, com exatamente 3.600 falsas e 3.600 verdadeiras. Desse modo, trata-se de uma base de dados balanceada, com informações do texto da notícia e a classificação: fake news ou não. A quantidade de dados presentes nele também é atrativa, podendo dar maior confiabilidade ao treinamento dos modelos. É importante ressaltar que na criação do corpus, para cada notícia falsa coletada, suas respectivas notícias verdadeiras foram capturadas de maneira semi-automática, como pode ser visto no próprio exemplo utilizado pelos autores, mostrado na Tabela 1.

Tabela 1. Exemplo de notícias falsas e verdadeiras alinhadas.

Falsa	Verdadeira
Michel Temer propõe fim do carnaval por 20 anos, “PEC dos gastos”. Michel Temer afirmou que não deve haver gastos com aparatos supérfluos sem pensar primeiramente na educação do Brasil. A medida pretende calcelar o carnaval de 2018.	Michel Temer não quer o fim do Carnaval por 20 anos. Notícias falsas misturam proximidade dos festejos, crise econômica e medidas impopulares do governo do peemedebista.
Acabou a mordomia ! Ingresso mais barato pra mulher ´e ilegal. Baladas que davam meia entrada para mulher, ou até mesmo gratuidade, esto na ilegalidade agora. Acabou o preconceito com os homens nas casas de show de todo o Brasil.	Ingresso feminino barato como marketing 'não inferioriza mulher’, diz juíza do DF. Afirmação consta em decisão sobre preços diferentes para homens e mulheres em festa no Lago Paranoá. 'Prática permite que mulher possa optar por participar de tais eventos sociais’, diz texto.

Fonte: adaptado de Monteiro et al. [4]

Para os experimentos em língua inglesa, algumas possíveis bases de dados foram encontradas, já que este é o idioma mais utilizado em pesquisas da área. Dentre elas, a base de dados Fake News [8] foi selecionada, já que a mesma também contém dados balanceados (10.387 falsos e 10.413 verdadeiros). Neste corpus, encontram-se dados sobre o autor, título, texto e rótulo (1 para verdadeira e 0 para falsa) da notícia, como mostra a Tabela 2. Para este corpus, os textos não foram disponibilizados pré-processados.

Tabela 2: Exemplos de dados do corpus Fake News

Título	Autor	Texto	Rótulo
House Dem Aide: We Didn't Even See Comey's Letter Until Jason Chaffetz Tweeted It	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Letter Until Jason Chaffetz Tweeted It By Darrell Lucus o...	1
BBC Comedy Sketch "Real Housewives of ISIS" Causes Outrage	Chris Tomlinson	The BBC produced spoof on the "Real Housewives" TV programmes, which has a comedic Islamic State twi...	0

Fonte: autoria própria.

4.2 Preparação dos Dados

No corpus Fake.Br, além dos dados puros, também é disponibilizada uma versão pré-processada, que foi a utilizada neste trabalho. Os textos das notícias foram colocados totalmente em letra minúscula, bem como foram retiradas todas as *stopwords*, acentos e diacríticos.

Já para a base de dados Fake News, os dados foram encontrados de modo bruto, contendo os textos integrais assim como foram recuperados em sua criação. Para este trabalho, foi escolhido que não fosse realizado nenhum pré-processamento desses dados, visando observar se isso afetaria demais os resultados obtidos.

Para os experimentos utilizando BERT, foi utilizada a tokenização dos textos disponibilizada pelo próprio modelo, o qual automaticamente realiza um pré-processamento dos dados. Para os demais algoritmos utilizados, não foi realizado tratamento dos dados, sendo realizada a tokenização utilizando o CountVectorizer e transformando em uma matriz de frequência normalizada com o TfidfVectorizer, ambos da biblioteca scikit-learn [19].

4.3 Modelagem

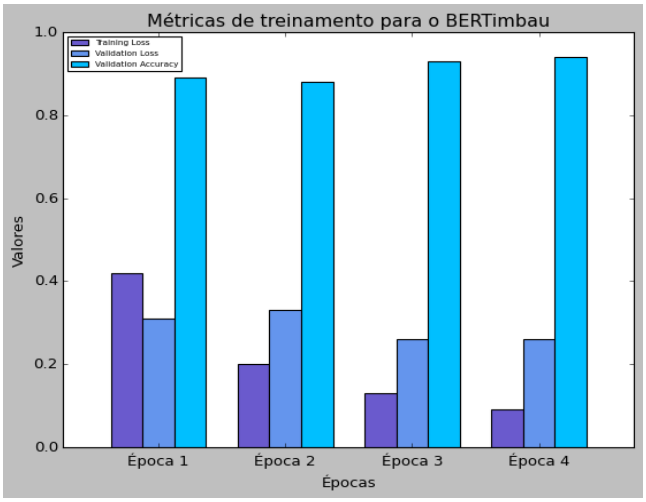
Para a implementação de modelos de classificação de Fake News, foram utilizados o BERTimbau [18] large e o BERT-large, para respectivamente os dados em português e inglês. Também foram executados algoritmos baseados em aprendizagem supervisionada - Naive Bayes Multinomial e XGBoost, para os dois corpus selecionados para os experimentos.

Em todos os modelos, foram utilizados 80% dos dados para o treinamento e 20% dos dados para testes. Especificamente para os modelos baseados no BERT, os 80% de dados de treinamento foram divididos de forma que 90% fosse destinado para o treino e os outros 10% para validação.

Além disso, para o BERTimbau e BERT em inglês, o treinamento foi realizado por 4 épocas (O paper seminal do BERT sugere 1 a 5 épocas para fine-tuning [17]), avaliando após cada uma delas o desempenho do treino realizado, calculando medidas de Training Loss, Validation Loss e Validation Accuracy, como mostram as figuras 4 e 5.

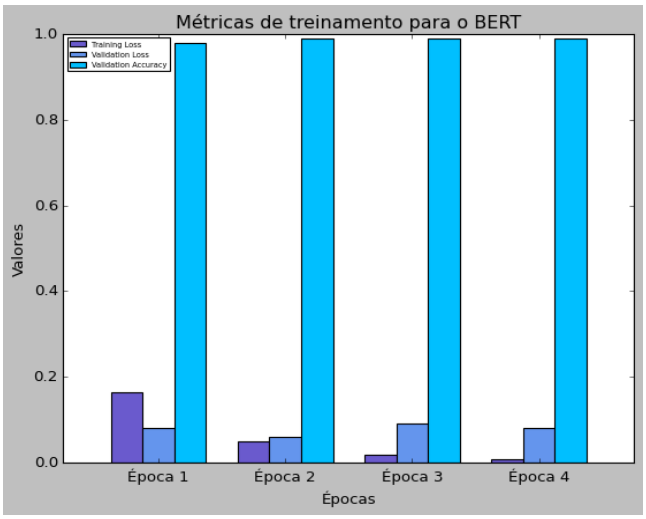
Como hiperparâmetros para o treinamento também foram utilizados tamanho de lote de 16 e taxa de aprendizado (Adam) de 5e-6, estando de acordo com os valores propostos por Devlin et al. [17],

Figura 4: Representação gráfica de métricas de treinamento do BERTimbau



Fonte: autoria própria

Figura 5: Representação gráfica de métricas de treinamento do BERT



Fonte: autoria própria

Na avaliação de cada modelo, foram calculadas as métricas de acurácia, recall, precision e f1-score, para todos os modelos dos experimentos realizados. Para efeitos de comparação com trabalhos realizados anteriormente, a f1-score foi a métrica escolhida como principal, por ser a mais representativa, além de ser bastante utilizada nos trabalhos de classificação.

Para o algoritmo Naive Bayes, não foram realizados ajustes, devido a ele não possuir hiperparâmetros que possibilitem tais mudanças.

No XGBoost, mesmo que ele contenha hiperparâmetros passíveis de mudanças, tais ajustes não foram realizados, já que utilizando seus valores padrão foram obtidos resultados bastante satisfatórios.

4.4 Validação

A fim de validar os modelos criados anteriormente, foram realizados testes de predição da classificação dos textos, com 20% dos dados em português e inglês que foram separados inicialmente para a etapa de testes.

Do mesmo modo que no treinamento, para realização da fase de testes foi realizada a tokenização dos dados, sendo esta realizada inerentemente pelos modelos BERT e explicitamente para os demais modelos, novamente foram utilizados CountVectorizer e TfidfVectorizer.

Os tokens gerados foram então passados para os modelos, visando a obtenção da classificação como Fake News ou notícia verdadeira. Como nas duas bases de dados eram anotadas, foi possível então comparar as predições realizadas pelos modelos com os valores reais que cada notícia possui, calculando então quão precisos são os modelos.

5. RESULTADOS

Os resultados obtidos para os modelos, tanto em português como em inglês, podem ser vistos nas tabelas 5 e 6.

Tabela 5: Métricas dos modelos para dados em português

Modelo/ Métrica	Acurácia	Precisão	Recall	F1-score
BERT	0.92	0.93	0.92	0.92
Naive Bayes	0.59	0.77	0.59	0.51
XGBoost	0.96	0.96	0.96	0.96

Fonte: autoria própria

Tabela 6: Métricas dos modelos para dados em inglês

Modelo/ Métrica	Acurácia	Precisão	Recall	F1-score
BERT	0.98	0.98	0.98	0.98
Naive Bayes	0.81	0.86	0.82	0.81
XGBoost	0.93	0.93	0.93	0.93

Fonte: autoria própria

A seguir, a tabela 7 realiza um comparativo entre os melhores resultados obtidos neste trabalho com a base Fake.Br e trabalhos anteriores utilizando esta mesma base.

Tabela 7: Comparação de resultados com experimentos anteriores

Trabalho	Modelo/Algoritmo	F1-score
Faustini e Covões [21]	Naive Bayes	85%
Faustini e Covões [21]	SVM	91%
Faustini e Covões [21]	Random Forest	88%
Freire, Silva e Goldschmidt [22]	HCS-F	99,9%
Autoria pŕopria	Naive Bayes	51%
Autoria própria	XGBoost	96%
Autoria própria	BERTimbau	92%

Fonte: autoria própria

5.1 Avaliação dos Modelos em Português

A partir dos resultados das métricas, é possível retirar algumas observações acerca deles. Para o modelo Naive Bayes em português, por exemplo, percebe-se que os valores de classificação dele não foram bons, obtendo apenas 0,51 de f1-score e acurácia de 0,59, bem abaixo dos resultados obtidos por Monteiro et al. e Moraes et al., que tiveram respectivamente 0,89 e 0,82 de acurácia.

Entretanto, se compararmos os modelos BERT e XGBoost para língua portuguesa com os melhores resultados obtidos, observa-se que os modelos deste trabalho superam vários dos trabalhos anteriores, não conseguindo superar apenas o de Feijo e Moreira, que teve 98% de f1-score e o de Freire, Silva e Goldschmidt, que obteve 99,9% de f1-score.

Para o de Feijó e Moreira, isso pode em parte ser explicado pelo extenso pré-treinamento que os autores fizeram com textos em português, com 4.8GB de dados nesta etapa, criando um modelo mais adaptado aos treinamentos de tarefas específicas em português.

Já no trabalho de Freire, Silva e Goldschmidt, os resultados são impressionantemente altos e indicam uma influência bastante positiva do uso de Crowd Signals para classificação de Fake News.

Ainda assim, o modelo XGBoost destaca-se, conseguindo resultados bem expressivos de 0.96 de f1-score e acurácia, com valores bem próximos dos melhores até o momento na detecção de Fake News, sem necessitar de vastos pré-treinamentos, assim como o próprio treinamento simplificado em comparação com o BERT, necessitando de menor tempo de treinamento para obtenção de resultados satisfatórios.

Ressalta-se que em termos de velocidade do treinamento, os modelos utilizando XGBoost e Naive Bayes

demandam bem menos tempo do que o necessário para treinar o modelo BERT.

5.2 Avaliação dos Modelos em Inglês

Para os modelos com dados em língua inglesa, percebe-se que o menos preciso foi novamente o baseado no algoritmo Naive Bayes, que obteve f1-score de 0.81. Em uma comparação direta, temos que o trabalho de Khan et al. obtiveram melhores resultados utilizando este mesmo algoritmo, com f1-score de 93%.

Os resultados mais baixos para os modelos Naive Bayes, para ambas as línguas, podem ser explicados tanto pela falta de pré-processamento dos dados, como também pelas poucas opções de features a serem selecionadas para a realização dos treinamentos, de modo que para o corpus em português, só havia a possibilidade de utilização dos textos da notícia em si, enquanto que em inglês, além dos textos, informações de título e autor também estavam disponíveis, mesmo não sendo utilizadas.

Com o modelo XGBoost, resultados muito bons foram alcançados, com f1-score e acurácia de 0.93, conseguindo se aproximar dos modelos estado da arte na detecção de notícias falsas e sendo um modelo mais simples e fácil de ser implementado e treinado, com tempo de treinamento bem menor quando comparado a modelos como o BERT.

Já no modelo BERT, foram atingidos resultados excelentes, com 0.98 de f1-score, este conseguindo ficar no mesmo patamar de trabalhos como os de Kaliyar, Goswami e Narang, de Gundapu e Mamidi, e de Khan et al, já que todos eles obtiveram resultados por volta de 0.98 de f1-score também.

Ainda assim, ressalta-se que todos estes trabalhos citados anteriormente conseguiram seus resultados com adaptações do modelo BERT ou com outra variação do modelo, o RoBERTa. Portanto, para experimentos com a utilização do BERT em sua implementação original, os resultados alcançados neste trabalho foram os melhores até o momento.

Para os modelos em inglês, também é possível perceber a diferença significativa do tempo de treinamento do modelo BERT para os modelos utilizando Naive Bayes e XGBoost, estes últimos sendo bem mais rápidos para conclusão da etapa de treino.

6. CONSIDERAÇÕES FINAIS

Este trabalho possui como principais contribuições experimentos que mostram que modelos como o BERT e similares(RoBERTa, AlBERTa, entre outros) são de fato uma excelente opção para classificação de notícias como falsas ou não, com o experimento em língua inglesa confirmando os trabalhos realizados anteriormente na área, que também conseguiram ótimas soluções com este modelo.

Para o português, também foi possível mostrar que o BERTimbau oferece uma boa base para realização da classificação de Fake News, já que ele foi pré-treinado extensivamente com dados em língua portuguesa e neste trabalho obteve resultados bastante satisfatórios.

Também foi possível conseguir ótimos resultados com a utilização de modelos baseados no algoritmo XGBoost, para ambas as línguas das bases de dados, destacando-se por serem modelos que não necessitam de grandes pré-treinamentos para alcançar métricas dos testes apenas um pouco inferiores àquelas obtidas com modelos de pré-treinamento.

Desse modo, utilizar o algoritmo XGBoost é uma solução viável para a detecção de Fake News, principalmente quando não é possível realizar um treinamento extensivo e que demande muito tempo, como no caso do BERT. Como visto nos experimentos realizados, tal modelo apresenta resultados bem próximos dos melhores resultados obtidos até o momento. Além disso, não realizar o pré-processamento dos dados para o modelo XGBoost não indicou uma alteração negativa para os resultados obtidos.

A maior limitação encontrada foi que os dados dos corpus utilizados não são muito favoráveis à utilização do algoritmo Naive Bayes. Por não possuírem tantas colunas de informações relevantes para cada notícia, a seleção de features para o treinamento foi comprometida, o que contribuiu para que os resultados utilizando este algoritmo fossem os piores dentre todos os utilizados. Também vale ressaltar que para este modelo, a não realização de pré-processamento dos dados pode ter contribuído para o resultado abaixo do esperado nos testes.

Em trabalhos posteriores, pode-se utilizar outras bases de dados, que possuam mais características sobre as notícias, de modo que o algoritmo Naive Bayes possa ser melhor utilizado nos experimentos.

Um importante trabalho a ser realizado é o uso de Cross Language Learning entre dados de língua inglesa (língua fonte) e dados de língua portuguesa (língua alvo), a fim de testar se haveria uma melhora nos resultados obtidos para o português, bem como aplicar esta mesma ideia para outras línguas alvo, que possuam menos dados disponíveis para experimentos na área de Fake News.

REFERÊNCIAS

[1] Mulher espancada após boatos em rede social morre em Guarujá, SP - http://g1.globo.com/sp/santos-regiao/noticia/2014/05/mulher-espancada-apos-boatos-em-rede-social-morre-em-guaruja-sp.html.

[2] Notícias falsas sobre eleição nos EUA têm mais alcance que notícias reais - http://g1.globo.com/mundo/eleicoes-nos-eua/2016/noticia/2016/11/noticias-falsas-sobre-eleicoes-nos-eua-superam-noticias-reais.htmlConger., S., and Loch, K.D. (eds.). Ethics and computer use. Commun. ACM 38, 12 (entire issue).

[3] Kai Shu, Amy Silva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. SIGKDD Explor. Newsl. 19, 1(June 2017) 22-36. DOI: https://doi.org/10.1145/3137597.3137600

[4] Monteiro, Rafael & Santos, Roney & Pardo, Thiago & Almeida, Tiago & Ruiz, Evandro & Vale, Oto. (2018). Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings. 10.1007/978-3-319-99722-3_33.

[5] Marcos Paulo Moraes, Jonice de Oliveira Sampaio, Anderson Cordeiro Charles. 2019. Data mining applied in

fake news classification through textual patterns. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (WebMedia '19). Association for Computing Machinery, New York, NY, USA, 321–324. DOI:https://doi.org/10.1145/3323503.3360648

[6] Liang Wu and Huan Liu. 2018. Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18). Association for Computing Machinery, New York, NY, USA, 637–645. DOI:https://doi.org/10.1145/3159652.3159677

[7] FACTCK.BR: A dataset to study Fake News in Portuguese. https://github.com/jghm-f/FACTCK.BR

[8] Fake News: Build a system to identify unreliable news articles. https://www.kaggle.com/c/fake-news/data?select=test.csv

[9] Moura R., Sousa-Silva R., Lopes Cardoso H. (2021) Automated Fake News Detection Using Computational Forensic Linguistics. In: Marreiros G., Melo F.S., Lau N., Lopes Cardoso H., Reis L.P. (eds) Progress in Artificial Intelligence. EPIA 2021. Lecture Notes in Computer Science, vol 12981. Springer, Cham. https://doi.org/10.1007/978-3-030-86230-5_62

[10] Feijó, D. and V. Moreira. Mono vs Multilingual Transformer-based Models: a Comparison across Several Language Tasks. *ArXiv* abs/2007.09757 (2020)

[11] Christian Janze, Marten Risius. Automatic Detection of Fake News on Social Media Platforms (2017). https://www.researchgate.net/publication/326405790_Automatic_Detection_of_Fake_News_on_Social_Media_Platforms

[12] Kaliyar, R.K., Goswami, A. & Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl* 80, 11765–11788 (2021). https://doi.org/10.1007/s11042-020-10183-2

[13] Gundapu, Sunil & Mamidi, Radhika. (2021). Transformer based automatic COVID-19 fake news detection system. https://www.researchgate.net/publication/348214408_Transformer_based_automatic_COVID-19_fake_news_detection_system

[14] Junaed Younus Khan, Md. Tawkat Islam Khondaker, Sadia Afroz, Gias Uddin, Anindya Iqbal. A benchmark study of machine learning models for online fake news detection, Machine Learning with Applications, Volume 4 (2021), 100032, ISSN 2666-8270. DOI: https://doi.org/10.1016/j.mlwa.2021.100032.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[16] Jay Alammar. The Illustrated Transformer. http://jalammar.github.io/illustrated-transformer/

[17] Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL* (2019). https://arxiv.org/abs/1810.04805

[18] Souza F., Nogueira R., Lotufo R. (2020) BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In: Cerri R., Prati R.C. (eds) Intelligent Systems. BRACIS 2020. Lecture Notes in Computer Science, vol 12319. Springer, Cham. https://doi.org/10.1007/978-3-030-61377-8_28

[19] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, null (2/1/2011), 2825–2830.

[20] Paterno, Lohann & Ferreira, Coutinho & Dosciatti, Mariza & Paraiso, Emerson. (2014). Estudo do Impacto de um Corpus Desbalanceado na Identificação de Emoções em Textos.

[21] Pedro Henrique Arruda Faustini, Thiago Ferreira Covões, Fake news detection in multiple platforms and languages, Expert Systems with Applications, Volume 158, 2020, 113503, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.113503.

[22] Paulo Márcio Souza Freire, Flávio Roberto Matias da Silva, Ronaldo Ribeiro Goldschmidt. Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. Expert Systems with Applications, Volume 183. 2021. 115414. ISSN 0957-4174, DOI:https://doi.org/10.1016/j.eswa.2021.115414.