

**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE TECNOLOGIA E RECURSOS NATURAIS
UNIDADE ACADÊMICA DE ENGENHARIA CIVIL
GRADUAÇÃO EM ENGENHARIA CIVIL**

MARIANA ALVES OLIVEIRA

**APLICAÇÃO DE REDES NEURAS ARTIFICIAIS NA
PREVISÃO DE ACIDENTES RODOVIÁRIOS EM PISTA SIMPLES**

Estudo do trecho da BR-230 pertencente ao estado da Paraíba

Campina Grande - Paraíba

2021

MARIANA ALVES OLIVEIRA

**APLICAÇÃO DE REDES NEURAS ARTIFICIAIS NA PREVISÃO DE
ACIDENTES RODOVIÁRIOS EM PISTA SIMPLES**

Estudo do trecho da BR-230 pertencente ao estado da Paraíba

Trabalho de Conclusão de Curso, apresentado à
Universidade Federal de Campina Grande –
UFCG, para encerramento do componente
curricular e conclusão da graduação em
Engenharia Civil.

Campina Grande – Paraíba

2021

MARIANA ALVES OLIVEIRA

**APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS NA PREVISÃO DE ACIDENTES
RODOVIÁRIOS EM PISTA SIMPLES**

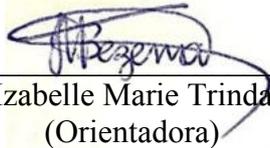
Estudo do trecho da BR-230 pertencente ao estado da Paraíba

Trabalho de Conclusão de Curso apresentado e julgado apto em:

21/05/2021

NOTA: 9,9

Banca examinadora:



Prof.^a Dr.^a Isabelle Marie Trindade Bezerra
(Orientadora)



Eng. Hélio da Silva Queiroz Júnior
(Coorientador)

Prof. Dr. Walter Santa Cruz
(Avaliador – membro interno)

Prof.^a Dr.^a Marília Marcy Cabral de Araújo
(Avaliadora – membro interno)

Itiel Alexandre Rodrigues Alves
(Avaliador – membro externo)

RESUMO

No Brasil, os acidentes viários constituem uma das grandes problemáticas sociais e da saúde pública no país. Na Paraíba, em específico, não é diferente, o caminho que une a cidade de Campina Grande ao sertão do estado traz números expressivos. Uma das formas de mitigação que tem sido estudada é a aplicação de redes neurais para a previsão de acidentes, sendo o objetivo do presente projeto desenvolver modelo de rede neural artificial para prever a frequência de acidentes em rodovias de pista simples da BR-230 com trecho do município de Campina Grande ao município de Cachoeira dos Índios, no estado da Paraíba. Para fins comparativos, foram utilizadas três diferentes formas de análise dos dados a partir de uma rede neural escolhida com base nos resultados obtidos em testes, as quais correspondem a análise com margens de erros com faixas variando de 5 a 30 km, a análise dividindo a rodovia em trechos de 5 a 30 km e a terceira análise consistiu em dividir a rodovia em trechos entre municípios. Esse procedimento de análise por diferentes métodos permitiu concluir que a melhor forma de avaliar trechos onde é necessária intervenção com medidas de segurança é dividindo a rodovia entre seus municípios, resultando em extensões maiores com maior taxa de acerto para rede neural. Baseado nos resultados foi possível observar que o trecho onde mais ocorrem acidentes fica compreendido entre o município de Sousa e Cachoeira dos Índios, passando por Cajazeiras, onde há uma maior concentração de casos levando em conta a proporcionalidade da extensão do trecho.

Palavras-chave: redes neurais; modelo de previsão de acidentes; rodovias federais; pista simples.

ABSTRACT

In Brazil, road accidents are one of the major social and public health problems in the country. In Paraíba, in particular, it is no different, the path that connects the city of Campina Grande to the state's backlands brings expressive numbers. One of the forms of mitigation that has been studied is the application of neural networks for the prediction of accidents, the objective of this project being to develop an artificial neural network model to predict the frequency of accidents on single-lane highways on the BR-230 with stretch from the municipality of Campina Grande to the municipality of Cachoeira dos Índios, in the state of Paraíba. For comparative purposes, three different forms of data analysis were used from a neural network chosen based on the results obtained in tests, which correspond to the analysis with error margins with ranges ranging from 5 to 30 km, the analysis dividing the highway in stretches of 5 to 30 km and the third analysis consisted of dividing the highway into stretches between municipalities. This analysis procedure by different methods allowed us to conclude that the best way to assess stretches where intervention with safety measures is necessary is to divide the highway between its municipalities, resulting in larger extensions with a higher rate of correctness for the neural network. Based on the results it was possible to observe that the stretch where accidents occur most is between the municipality of Sousa and Cachoeira dos Índios, passing through Cajazeiras, where there is a greater concentration of cases taking into account the proportionality of the stretch.

Key-Words: neural networks; accident prediction model; federal highways; simple clue.

LISTA DE FIGURAS

Figura 1: Neurônio simples. (Fonte: Deep Learning Book, 2019).....	16
Figura 2: Esquema do aprendizado supervisionado (fonte: reprodução autor)	19
Figura 3: Padrão de dados linearmente separáveis e não linearmente separáveis (fonte: Deep Learning Book, 2020).....	22
Figura 4: Rede Neural Perceptron esquematizada.....	22
Figura 5: Diagrama de Perceptron Multicamadas (fonte: Moreira, 2018).....	23
Figura 6: Fluxograma do método adotado na pesquisa	25
Figura 7: Trecho Campina Grande - Soledade (fonte: Google Maps)	33
Figura 8: Trecho Soledade - São Mamede (fonte: Google Maps).....	33
Figura 9: Trecho São Mamede - Patos (fonte: Google Maps).....	33
Figura 10: Trecho Patos - Pombal (fonte: Google Maps).....	34
Figura 11: Trecho Pombal - Sousa (fonte: Google Maps).....	34
Figura 12: Trecho Sousa - Cajazeiras (fonte: Google Maps)	35
Figura 13: Trecho Cajazeiras - Cachoeira dos Índios (fonte: Google Maps).....	35
Figura 14: Arquitetura da rede neural 6 (fonte: autora).....	40
Figura 15: Arquitetura da rede neural 14 (fonte: autora).....	40
Figura 16: Arquitetura da rede neural 21 (fonte: autora).....	40
Figura 17: Arquitetura rede neural 6 (fonte: autora).....	41
Figura 18: Porcentagem de acertos por faixa (média da rede neural 6)	42
Figura 19: Trecho de maior ocorrência de acidentes na BR - 230 entre Campina Grande e Cachoeira dos Índios (fonte: CNT).....	44
Figura 20: Trecho Campina Grande – Soledade. Total de 1287 acidentes de 2017 a 2019 (fonte: CNT).....	45
Figura 21: Trecho Soledade – Patos. Total de 904 acidentes de 2017 a 2019. (Fonte: CNT)	45
Figura 22: Trecho Patos – Pombal. Total de 531 acidentes de 2017 a 2019. (Fonte: CNT)	46
Figura 23: Trecho Pombal – Sousa. Total de 508 acidentes de 2017 a 2019. (Fonte: CNT)	46
Figura 24: Trecho Sousa – Cajazeiras. Total de 745 acidentes de 2017 a 2019. (Fonte: CNT)	47
Figura 25: Gráfico comparativo entre métodos 1, 2 e 3	49

LISTA DE TABELAS

Tabela 1: Parâmetros adotados na tabela da acidentes da PRF.....	26
Tabela 2: Atribuição de pesos para cada parâmetro	27
Tabela 3: Subdivisão do trecho com base nos municípios e suas extensões.....	32
Tabela 4: População dos municípios paraibanos.....	32
Tabela 5: Arquiteturas das redes neurais	36
Tabela 6: Médias para R treino e percentual de acerto das redes neurais.....	37
Tabela 7: Cinco melhores resultados das redes neurais por faixa de quilometragem..	38
Tabela 8: Três arquiteturas que apareceram com maior frequência	39
Tabela 9: Resultado médio da rede neural 6.....	41
Tabela 10: Número de acidentes por trecho	42
Tabela 11: Porcentagem de acerto com subdivisão em municípios	43
Tabela 12: Porcentagem de acertos por trecho com nova subdivisão.....	44
Tabela 13: Média de resultados da rede neural 6 para primeiro método de análise....	47
Tabela 14: Resultados do segundo método de análise da rede neural	48
Tabela 15: Resultados da terceira forma de análise da rede.....	48
Tabela 16: Classificação do tipo de veículo	54
Tabela 17: Média e desvio padrão dos dados de faixa etária	55
Tabela 18: Normalização da faixa etária.....	55
Tabela 19: Média e desvio padrão do quilômetro	57
Tabela 20: Normalização dos dados de quilômetros.....	57
Tabela 21: Valores alvos e valores obtidos pela rede neural 6	63

Sumário

1.	INTRODUÇÃO	9
2.	JUSTIFICATIVA	11
3.	OBJETIVOS	12
3.1.	Geral	12
3.2.	Específicos	12
4.	REFERENCIAL TEÓRICO	13
4.1.	ACIDENTES DE TRÂNSITO	13
4.3.	INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA	15
4.4.	REDES NEURAIS ARTIFICIAIS (RNA's)	15
4.5.	PERCEPTRON	21
4.6.	MULTI-LAYER PERCEPTRON	23
5.	METODOLOGIA	25
6.	RESULTADOS E DISCUSSÕES	36
6.1.	ESCOLHA DA REDE	36
6.2.	ANÁLISE DO TRECHO	42
6.3.	COMPARAÇÃO ENTRE OS MÉTODOS	47
7.	CONCLUSÕES	50
8.	REFERÊNCIAS	51

1. INTRODUÇÃO

Dentre os problemas significativos que envolvem o trânsito e até mesmo a saúde pública no Brasil, tem-se a incidência de acidentes nas rodovias federais como uma dessas problemáticas. Segundo a Confederação Nacional de Transporte (CNT), o número de acidentes ocorridos em rodovias federais, em 2019, foi 67.427 que, apesar de representar diminuição de 2,6% em relação ao ano anterior, ainda é um número significativo.

Ainda de acordo com a CNT (2020), mesmo que o número de acidentes tenha diminuído, os acidentes com vítimas (mortos e feridos), tiveram elevação de 3,3% entre os anos, subindo de 53.963, em 2018, para 55.756, em 2019. Foram 2.526 feridos a mais em 2019.

Analisando o estado da Paraíba, objeto de estudo do presente trabalho, VITORINO et al. (2020) diz:

“No que se refere aos dados da Secretaria de Estado da Segurança e Defesa Social (SEDS) em 2018, na Paraíba, foram contabilizadas 753 vítimas de acidentes letais de trânsito (ALT). O número de acidentes na Paraíba caiu de 8,5% em relação ao ano de 2017, após serem pagos pelo Seguro de Danos Pessoais Causados por Veículos Automotores de Vias Terrestres (DPvat) 1.017 indenizações por morte.”

Tendo em vista os números apresentados, é possível perceber a importância de atentar para o quadro de acidentes rodoviários no Brasil. De acordo com SINAY & TAMAYO (2005), a segurança dos usuários do sistema de trânsito (condutores, passageiros e pedestres), cada vez mais, se constitui um dos objetivos principais do planejamento e operação desta modalidade de transporte.

O uso de tecnologias de inteligência artificial vem trazendo e ganhando espaço na área de pesquisas e aplicação. Ao tratar-se do trânsito nas rodovias brasileiras, a inteligência artificial já vem sendo utilizada, como por exemplo na captura de imagens, monitoramento do trânsito, controle de velocidade, detecção de problemas na rodovia, dentre outros aspectos.

Uma das vertentes da inteligência artificial em rodovias que vem sendo estudada é a de previsão de acidentes, a fim de identificar trechos de maior ocorrência de acidentes para a mitigação deles. Estudos como os de Silva (2019), Vitorino (2020) e Barros (2003) já trazem percepções das rodovias brasileiras nesse quesito.

Diante do exposto, uma tentativa de ter um controle de quais fatores contribuem diretamente para a ocorrência deles fazendo uso da modelagem de redes neurais para previsão

de acidentes, sendo este o método utilizado no presente trabalho para o trecho de pista simples da BR-230 situado entre Campina Grande e Cachoeira dos Índios, na Paraíba.

2. JUSTIFICATIVA

Segundo a Organização Mundial de Saúde, o Brasil está em quinto colocado no ranking dos países com maior índice de morte no trânsito (OMS, 2019), sendo os acidentes rodoviários responsáveis por deixar 400 mil pessoas com algum tipo de seqüela (ONSV, 2019).

Ainda segundo o Observatório Nacional de Segurança Viária (ONSV, 2019), cerca de 60% dos leitos hospitalares do Sistema Único de Saúde (SUS) são preenchidos por acidentados, resultando em custos anuais de R\$ 52 bilhões com acidentes no trânsito.

Levando em consideração os dados expostos acima, é possível perceber que os acidentes de trânsito fazem parte de uma problemática presente no cotidiano do cidadão brasileiro. Diversas vezes, ao ligar a televisão no noticiário ou abrir um portal de notícias, encontra-se manchetes com acidentes que levaram vidas, que muitas vezes não tinham culpa da fatalidade e morreram por irresponsabilidade de outro condutor ou por conta das diversas condições envolvidas na rodovia em determinado dia e horário.

Ao ver uma porcentagem de 60% dos leitos hospitalares do SUS ocupados por vítimas dos acidentes de trânsito, os leitos passam a significar pessoas que têm família, trabalho, sonhos e uma vida pela frente que, por imprudência, acabam.

Este estudo baseia-se em trazer, sobretudo, um auxílio para que as autoridades competentes possam mitigar os acidentes viários, preservando vidas. Para isso, foi adotado o uso das redes neurais, uma vez que é um dos campos de estudo que vem se desenvolvendo e gerando tecnologias mais modernas a serem utilizadas a favor do homem.

3. OBJETIVOS

3.1. Geral

Desenvolver modelo de rede neural artificial para prever a frequência de acidentes em rodovias de pista simples da BR-230 com trecho do município de Campina Grande ao município de Cachoeira dos Índios, no estado da Paraíba.

3.2. Específicos

- Coletar dados sobre acidentes na BR-230, trecho de pista simples;
- Definir os parâmetros que influenciam na ocorrência de acidentes, para desenvolver a proposta do trabalho;
 - Tratar os dados e realizar normalização;
 - Propor arquiteturas e realizar avaliação de acordo com o percentual de acertos.
 - Avaliar a aplicabilidade do sistema para previsão de acidentes para o objeto de estudo em questão.

4. REFERENCIAL TEÓRICO

4.1. ACIDENTES DE TRÂNSITO

Os acidentes de trânsito não ocorrem por acaso, mas são decorrentes de deficiências das vias, dos veículos e, principalmente, das falhas humanas (Marín & Queiroz, 2000).

Segundo dados da Pesquisa Nacional de Saúde (PNS), feita em 2013 no Brasil, 3,1% dos 60.200 entrevistados haviam se envolvido em acidente de trânsito com vítimas em um intervalo de pelo menos um ano anterior à pesquisa (IBGE, 2015).

Ao fazer uma análise do crescimento na frota de automóveis no Brasil, o Instituto de Pesquisa Econômica Aplicada (Ipea) traz dados do Datasus (Departamento de Informática do Sistema Único de Saúde) que apontam que, em 2015, os acidentes de transportes terrestres mataram cerca de 43 mil pessoas por ano no país, representando uma das principais causas de morte. Dentre as mortes ocasionadas no trânsito, 20% delas ocorreram em rodovias federais, gerando ainda um saldo de cerca de 26 mil feridos graves por ano.

Além de trazer traumas para as vítimas dos acidentes e seus familiares, os acidentes também geram custo para o governo. Segundo a Empresa Brasil de Comunicação (EBC) entre 2009 e 2018 os acidentes de trânsito resultaram em mais de 1,6 milhões de feridos e um custo de R\$2,9 bilhões para o SUS (Sistema Único de Saúde), trazendo prejuízos significativos para a saúde pública.

Aproximadamente, um quarto dos atendimentos por acidentes em serviços sentinela de urgência e emergência do país é por acidentes de transporte (Ministério da Saúde, 2017). Considerando-se apenas as causas externas, os acidentes de trânsito terrestres (ATT) respondem pela segunda maior taxa de DALY (*disability-adjusted life years*), indicador utilizado no estudo da carga de doença, atrás apenas de homicídios e violências, sendo o uso de álcool apontado como importante fator de risco associado a mortes e incapacidades por essa causa, sobretudo na população masculina (Malta, 2017).

Andrade e Antunes (2019), em estudo sobre a tendência do número de vítimas em acidentes de trânsito nas rodovias federais brasileiras, observaram em seus resultados que a região nordeste apresenta o maior percentual de óbitos em acidentes com vítima (10%) e verificaram na mesma região o maior número de acidentes com lesões graves (32%).

VITORINO et al. (2020) em seus estudos avaliaram as características de tendência, sazonalidade e aleatoriedade dos acidentes. Analisaram 9 rodovias federais que cruzam o estado

da Paraíba, tendo como resultado uma maior incidência de acidentes de trânsito na BR-101, BR-104 e BR-230, quando comparadas as demais rodovias estudadas, sendo esta última o objeto do presente estudo.

4.2. MODELO DE PREVISÃO DE ACIDENTES

Diante da preocupação com os números de acidentes de trânsito, tem-se buscado, cada vez mais, alternativas que auxiliem na previsão e, conseqüentemente, atenuem esses números.

Nesse sentido, os modelos de previsão de acidentes (MPAs) são amplamente utilizados para estimar a frequência de acidentes para uma determinada unidade espacial durante um determinado período.

Uma das mais importantes aplicações práticas de MPAs é o ranking de sites que visa identificar locais perigosos com problemas de segurança subjacentes (WANG et al., 2011), sendo essa uma das análises feitas no presente estudo para o trecho em questão. Desta forma, as técnicas de modelagem são o meio de obtenção de modelos que reproduzam de forma mais fidedigna a realidade, sendo bastante importante o emprego de técnicas adequadas à natureza dos dados a serem tratados (COSTA, 2013).

Savolainen *et al.* (2011) destacam que o desenvolvimento de contramedidas eficazes requer uma compreensão completa dos fatores que afetam a probabilidade de um acidente ocorrer ou, dado que um acidente ocorreu, as características que podem mitigar ou exacerbar o grau de lesão sofrido pelos envolvidos no acidente na estrada.

Em seu estudo, Silva (2019) aborda duas análises em torno do estudo das rodovias através das redes neurais: análise dos acidentes quanto a severidade e previsão de frequência. Nos modelos de previsão da frequência de acidentes, tem-se buscado estabelecer a relação entre o número de acidentes e os fatores de risco, basicamente os fatores viário-ambientais. A variável-resposta desses modelos é o número de acidentes por segmento ou o número de acidentes por segmento por ano. Um modelo de previsão de frequência dos acidentes permite avaliar os fatores contribuintes para a ocorrência do número total de acidentes.

Lord e Mannering (2010) apresentam a revisão e avaliação de alternativas metodológicas, trazendo como principais modelos para previsão de acidentes a regressão de Poisson, regressão binomial, regressão binomial negativa, regressão Poisson lognormal, regressão gama, regressão inflacionada de zero, equações de estimação generalizadas, modelo

multinomial negativo, modelo de efeitos aleatórios, modelo de parâmetros aleatórios e modelo multivariado.

Para o presente estudo, foi utilizado o modelo multivariado, que será melhor demonstrado a seguir, juntamente com a abordagem sobre redes neurais.

4.3. INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA

A inteligência artificial é uma vertente da ciência computacional que procura desenvolver sistemas que possuam características associadas ao comportamento humano, trazendo abordagens cognitivas onde o aprendizado é feito a partir de observações.

O Aprendizado de Máquina (AM), uma subárea da Inteligência Artificial, é bastante utilizado e se coloca como uma poderosa ferramenta para resolver problemas em vários domínios de aplicação (SILVA, 2019).

Quando se fala em aproximar o comportamento da máquina ao do ser humano, tem-se a abordagem do teste de Turing, onde ele antecipou algumas questões da inteligência artificial. Esse teste visa mensurar a capacidade da máquina demonstrar um comportamento inteligente que seja equivalente ao do ser humano.

O teste é feito com uma pessoa que entra em uma conversa em que obterá respostas vindas de outro ser humano e de uma máquina projetada para responder de forma indistinguível do homem, de modo que os 3 estejam isolados um do outro. Se o primeiro homem não conseguir distinguir a diferença entre as respostas, diz-se que a máquina foi aprovada no teste de Turing.

Um programa tem seu aprendizado pautado a partir de uma classe de tarefas, de uma medida de desempenho e da experiência, onde o desempenho em determinada classe melhora com a experiência.

4.4. REDES NEURAS ARTIFICIAIS (RNA's)

A Teoria das Redes Neurais Artificiais, também chamadas aqui de RNA, vem se consolidando, mundialmente, como uma nova e eficiente ferramenta para lidar com a ampla classe dos problemas complexos (KOVÁCS, 2006). As redes neurais tornaram-se um artifício para contextos em que há uma extensa gama de dados que precisam ser modelados e analisados em aspectos multidisciplinares.

As RNAs foram criadas com o intuito de se aproximar do sistema neurológico humano, na tentativa de aproximar esse modelo computacional quanto à funcionalidade e potencialidade do cérebro humano. Criados os neurônios artificiais, ao interligá-los, fazendo conexões, cria-se uma rede neural capaz de desempenhar atividades criativas e cognitivas. Dessa forma, pode-se definir uma rede neural como um processador maciçamente distribuído constituído de grande número de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso (KOPILER et al., 2019).

As principais características da rede neural são que elas correspondem a modelos adaptáveis e treináveis, podem representar domínios complexos, são capazes de generalizar uma informação incompleta, são robustas, são capazes de fazer um armazenamento associativo de informações, processam informações espaço-temporais e possuem grande paralelismo, garantindo rapidez de processamento. Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples, denominadas de “neurônios” ou unidades de processamento (HAYKIN, 2001), que são interligados por conexões, que representam as sinapses, onde cada uma delas possui um peso associado, e são compostas pelas camadas de entrada, intermediária e de saída, representadas na Figura 1.

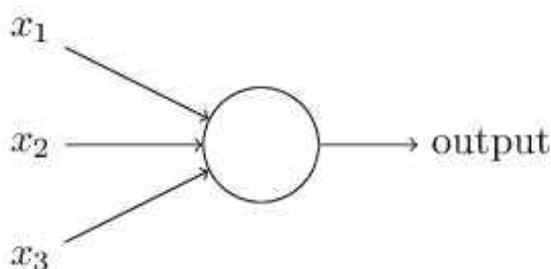


Figura 1: Neurônio simples. (Fonte: Deep Learning Book, 2019)

Na entrada são inseridos os padrões que são submetidos à rede; nas camadas intermediárias – também chamadas de *hidden* (escondidas) – é onde ocorre boa parte do processamento e, por fim, a saída é aquela que apresenta o resultado.

A estrutura da RNA, também chamada de arquitetura, apresenta o número de camadas por redes e a quantidade de neurônios por camada. Uma das principais etapas do emprego de RNAs na resolução de uma problemática é a escolha da melhor arquitetura, pois este processo é experimental e por isso, demanda tempo de execução. Na prática o processo deve ser aplicado

com o intuito de testar os vários métodos de aprendizado e as diferentes configurações que uma rede possa ter para a resolução do problema em questão (MIRANDA et al., 2009).

As entradas são propagadas a partir da topologia da RNA, sendo transformadas pelos pesos sinápticos e pela função de ativação (AF) dos neurônios (MACHADO E FONSECA JÚNIOR, 2003). Recebendo entradas de n neurônios (y_i), o neurônio k calcula a sua saída através de:

$$y_k = AF \left\{ \sum_{i=1}^n (y_i w_{ki}) + b_k \right\}$$

Onde y_i é a saída calculada pelo neurônio i , w_{ki} representa o peso sináptico entre o neurônio i e o neurônio k e b_k é o peso entre um valor constante e diferente de zero ao neurônio k , conhecido como bias. Se o neurônio estiver ligado às entradas, o termo y_i é substituído pela entrada correspondente.

O neurônio matemático recebe um ou mais sinais de entrada e devolve um único sinal de saída, que pode ser distribuído como sinal de saída da rede, ou como sinal de entrada para um ou vários outros neurônios da camada posterior (que formam a rede neural artificial). Os dendritos e axônios são representados matematicamente apenas pelas sinapses, e a intensidade da ligação é representada por uma grandeza denominada peso sináptico, simbolizada pela letra w . Quando as entradas, x são apresentadas ao neurônio, elas são multiplicadas pelos pesos sinápticos correspondentes, gerando as entradas ponderadas (DEEP LEARNING BOOK, 2019), como mostrado a seguir:

$$output = \begin{cases} 0, & \text{se } \sum_j w_j x_j \leq threshold \\ 1, & \text{se } \sum_j w_j x_j > threshold \end{cases}$$

O aprendizado de uma RNA é realizado por meio de processos iterativos de ajustes aplicados aos pesos sinápticos, o chamado treinamento. O aprendizado só ocorre quando a rede

neural atinge uma solução generalizada para um determinado problema. Em síntese, treinar uma rede é ajustar a sua matriz de pesos sinápticos de forma que o vetor de saída coincida com um valor desejado para cada vetor de entrada (MIRANDA et al., 2009).

A utilização de uma RNA na solução de uma tarefa passa inicialmente por essa fase de aprendizagem, quando a rede extrai informações relevantes de padrões de informação apresentados para ela, criando assim uma representação própria para o problema. A etapa de aprendizagem consiste em um processo iterativo de ajuste de parâmetros da rede, os pesos das conexões entre as unidades de processamento, que guardam, ao final do processo, o conhecimento que a rede adquiriu do ambiente em que está operando. (BRAGA *et al.*, 2000).

Após o treinamento da RNA, espera-se que a rede projetada seja capaz de generalizar. Diz-se que uma rede generaliza bem quando o mapeamento de entrada-saída computado pela rede for correto (ou aproximadamente correto) para o conjunto amostral não utilizado no processo de treinamento da rede. O processo de aprendizagem pode ser visto como um problema de “ajuste de curva”. A própria rede pode ser considerada como um mapeamento não-linear de entrada saída. (MIRANDA *et al.*, 2009).

Li *et al.* (2012) frisam em seus estudos que é requerida uma suposição sobre a distribuição de dados para a modelagem estatística e que é estabelecida uma forma funcional entre as variáveis dependentes e explicativas. Diversas vezes essas premissas podem não ser verdadeiras, e em sendo violadas, conduzem a estimativas equivocadas e produção incorreta de inferências.

Chang (2005) e Li *et al.* (2012) reiteram que não é necessário um vínculo pré-existente entre as variáveis da rede neural artificial (RNA), mas tem-se o intuito de construir um modelo a partir de dados reais supervisionados que resulta em uma arquitetura de pesos e função de ativação não linear.

Existem, porém, dois principais tipos de aprendizado: o supervisionado e o não supervisionado. Braga (2000) traz em sua publicação a diferença entre eles: no aprendizado supervisionado a entrada e saída desejadas para a rede são fornecidas externamente. O objetivo é ajustar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. A rede tem sua saída corrente (calculada) comparada com a saída desejada, recebendo informações do supervisor sobre o erro da resposta atual. A cada padrão de entrada submetido à rede compara-se a resposta desejada (que representa uma ação ótima para ser

realizada pela rede) com a resposta calculada, ajustando-se os pesos das conexões para minimizar o erro.

No aprendizado supervisionado, como explicado anteriormente, assume-se uma série de dados para treinar o sistema com instâncias rotuladas, ou seja, os dados já estão pareados com suas respectivas respostas.

Nesse tipo de aprendizado, pretende-se prever uma variável y que depende de outras variáveis x , alimenta-se o sistema com as variáveis x e y conhecidas e ele reproduz. A expectativa é de que, após apresentar esses dados pareados, a máquina consiga prever a variável y de observações que nunca viu, tendo apenas a variável x delas.

A Figura 2 apresenta o esquema de funcionamento do aprendizado supervisionado:

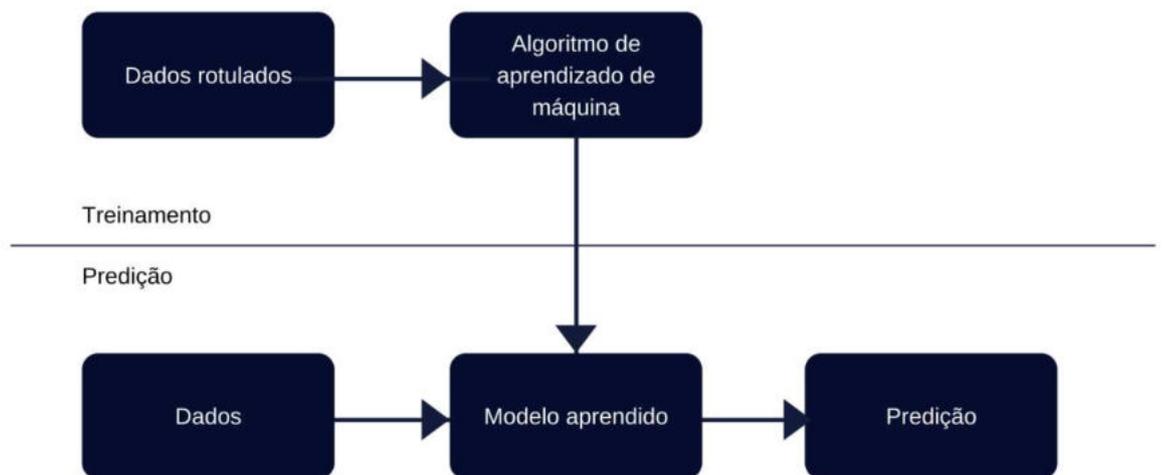


Figura 2: Esquema do aprendizado supervisionado (fonte: reprodução autor)

Há dois tipos de problemas que podem ser resolvidos utilizando o aprendizado supervisionado: a classificação e a regressão. Os problemas de classificação são aqueles onde pretende-se prever um valor discreto, que é classificar um dado segundo uma categoria. Já os problemas de regressão são aqueles que visam prever um valor contínuo.

Já o não-supervisionado, somente os padrões de entrada estão disponíveis para a rede, ao contrário do aprendizado supervisionado, cujo conjunto de treinamento possui pares de entrada e saída. A partir do momento em que a rede estabelece uma harmonia com as

regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características da entrada e criar novas classes ou grupos automaticamente (BRAGA *et al.*, 2000).

O não supervisionado é aplicado para os casos em que as amostras não têm saídas conhecidas ou saídas esperadas (SILVA, 2019), ao contrário do supervisionado, é quando se possui apenas a variável independente x , sem seus pares y , visando que a máquina consiga gerar esses valores.

Existe ainda outros dois tipos de aprendizado: o aprendizado por reforço e a capacidade de generalização.

No aprendizado por reforço, dado um certo ambiente, visa-se mapear que ações devem ser executadas, ou seja, ao invés de estimar uma função que mapeia x e y , passa-se a aprender um método que mapeia S (estado) em A (ações). Cada par (S, A) está associado a determinada recompensa de forma que a máquina compreende qual o processo que maximiza essa recompensa ao longo das interações.

O aprendizado por reforço tem influência da psicologia e da economia, principalmente no campo da teoria utilizada nos sistemas de jogos e nas estratégias de trocas no mercado de finanças.

Quanto a capacidade de generalização, ela corresponde a duas características para os modelos de aprendizado de máquina que são a força para aprender as regularidades nos dados fornecidos para o treinamento e a generalização desses dados para gerar os novos. Em contrapartida, essas duas características se opõem, onde o aumento de uma resulta na relativa perda de outra.

Quando as RNAs são utilizadas na modelagem de processos, é usual utilizar-se o aprendizado supervisionado. A implementação deste método, pode ser feita off-line ou on-line. No modo off-line, os dados do conjunto de treinamento não mudam, e caso seja necessário à inclusão de novos dados, um novo banco de dados destinado ao treinamento deve ser elaborado, contendo os dados antigos e novos. No modo on-line, a rede opera em constante processo de adaptação, sendo atualizada constantemente (EYNG, 2008).

4.5. PERCEPTRON

O perceptron, introduzido por Rosenblatt, em 1958, é uma forma simples de RNA cuja principal aplicação se dá nos problemas de classificação de padrões. O perceptron de camada única só é capaz de classificar padrões linearmente separáveis. Na prática, o problema a ser trabalhado não admite uma separação linear exata, tornando-se necessário o uso de um perceptron multicamadas (AMBRÓSIO, 2002).

Um Perceptron segue o modelo feedforward, o que significa que as entradas são enviadas para o neurônio, processadas e resultam em uma saída (DEEP LEARNING BOOK, 2019). A rede neural feedforward é arquitetada de forma que as conexões não formam ciclos recorrentes, mas são sempre alimentadas para frente, encaminhando sempre os valores de entrada em direção à camada de saída.

Um elemento importante da RNA é a função de ativação. Ela está relacionada à capacidade representativa das redes neurais e introduz uma componente não linear. Transforma não linearmente o sinal de entrada, e a saída transformada é enviada para a próxima camada de neurônios como entrada. Assim, relaciona-se à decisão de ativar ou não o neurônio com base na informação recebida (SILVA, 2019). As principais funções de ativação identificadas nos trabalhos são a tangente hiperbólica (tansig), sigmóide (logsig) e linear (purelin).

Tendo em mãos esses conceitos, pode-se definir a backpropagation do perceptron, denominada também de retroprogramação, que corresponde a um algoritmo de treinamento dividido em duas etapas: a de feedforward dos valores e a etapa de back, onde calcula-se o erro e propaga de volta para camadas antecessoras. Dessa forma, para que o processo de backpropagation seja mais preciso, é necessária a etapa de feedforward.

As redes de uma só camada resolvem apenas problemas linearmente separáveis, a solução de problemas não linearmente separáveis passa pelo uso de redes com uma ou mais camadas intermediárias ou escondidas (BRAGA et al., 2000), como exemplificado na Figura 3, adotando o modelo de neurônio de McCulloch & Pitts.

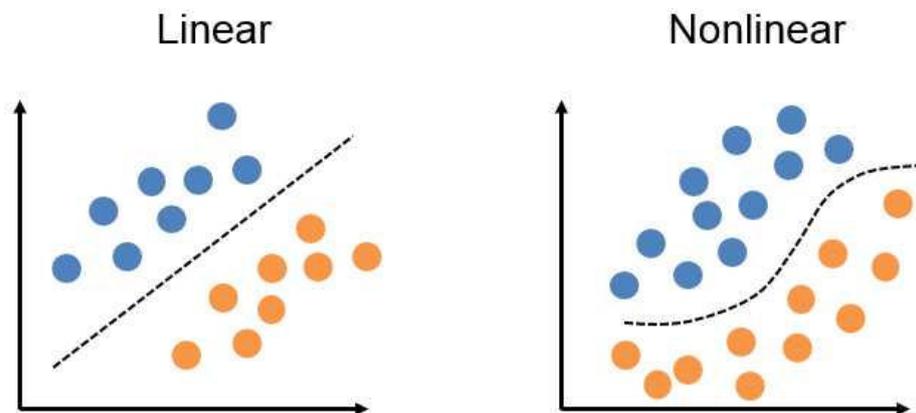


Figura 3: Padrão de dados linearmente separáveis e não linearmente separáveis (fonte: Deep Learning Book, 2020)

Na Figura 4 pode-se observar o esquema de uma rede neural do tipo perceptron.

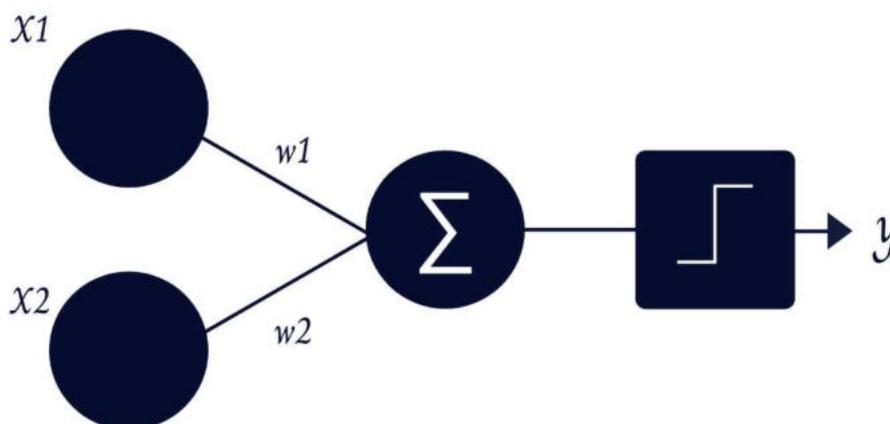


Figura 4: Rede Neural Perceptron esquematizada

Onde x_1 e x_2 são os inputs, w_1 e w_2 são os pesos de cada input e o somatório é a combinação linear desses inputs ponderada pelos pesos.

$$\vartheta = \sum_{i=0}^n x_i w_i$$

E o output é:

$$f(\vartheta) = \begin{cases} 1, & \text{se } \vartheta \geq 0 \\ 0, & \text{se } \vartheta < 0 \end{cases}$$

4.6. MULTI-LAYER PERCEPTRON

Um Multilayer Perceptron (MLP) é uma rede neural artificial composta por mais de um Perceptron. Eles são compostos por uma camada de entrada para receber o sinal, uma camada de saída que toma uma decisão ou previsão sobre a entrada, e entre esses dois, um número arbitrário de camadas ocultas que são o verdadeiro mecanismo computacional do MLP (DEEP LEARNING BOOK, 2019).

A Data Science Academy, no Deep Learning Book (2019), aborda que os MLPs são frequentemente aplicados a problemas de aprendizagem supervisionados: treinam em um conjunto de pares entrada-saída e aprendem a modelar a correlação (ou dependências) entre essas entradas e saídas. O treinamento envolve o ajuste dos parâmetros ou pesos do modelo para minimizar o erro. Um exemplo de MLP pode ser verificado na Figura 5:

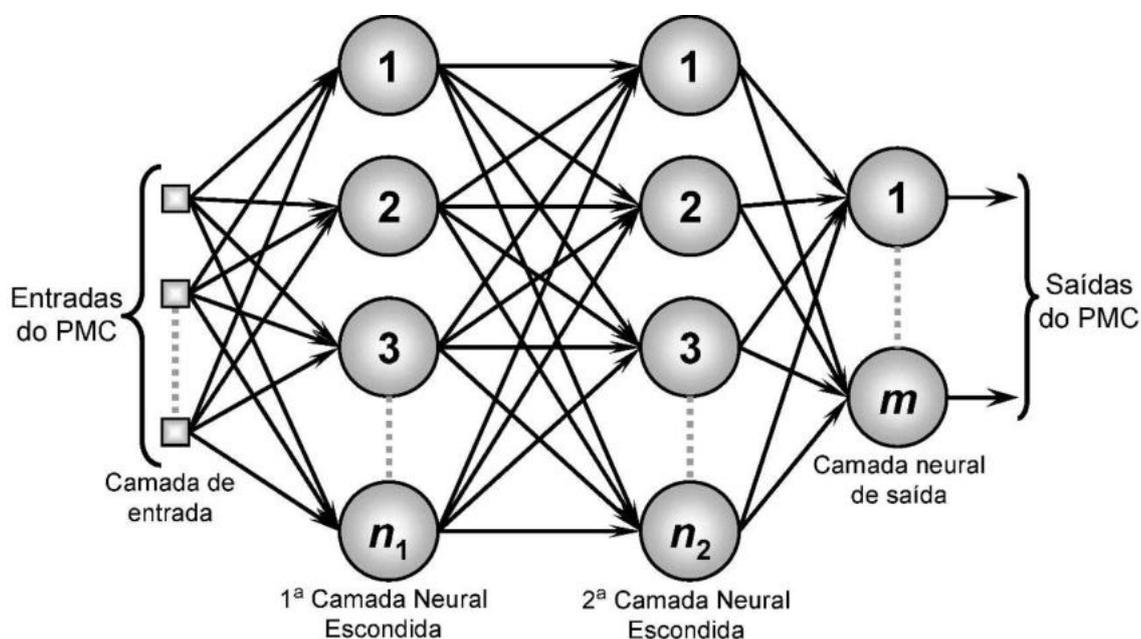


Figura 5: Diagrama de Perceptron Multicamadas (fonte: Moreira, 2018)

O treinamento ocorre em duas fases, cada fase percorre a rede em um sentido. Estas duas fases são chamadas de fase forward e fase backward. A fase forward é utilizada para definir a saída da rede para um dado padrão de entrada. A fase backward utiliza a saída desejada e a saída fornecida pela rede para um dado padrão de entrada (BRAGA et al., 2000).

Quando se trata do backpropagation, há vantagens como a simplicidade de implementação e a boa capacidade de generalizar. Em contrapartida, há desvantagens como a dificuldade de justificar respostas, tem alto custo computacional e baixa velocidade de aprendizado.

5. METODOLOGIA

Na Figura 6 está representada a sequência das etapas desenvolvidas:

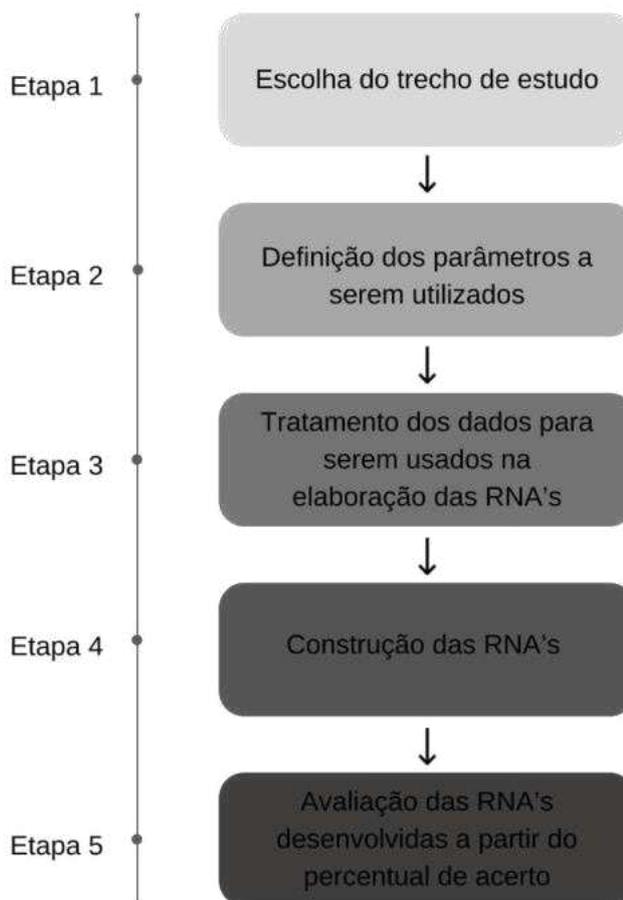


Figura 6: Fluxograma do método adotado na pesquisa

Como apresentado no fluxograma, o desenvolvimento da proposta deste trabalho seguiu 5 etapas, apresentadas a seguir.

Etapa 1: escolha do trecho de estudo

Nessa etapa, inicialmente foi definido que seria trabalhado o trecho de pista simples da BR-230 pertencente ao estado da Paraíba. Posteriormente, foi levado em consideração características do tráfego, como volume de tráfego, que tornavam o trecho inicial da rodovia, de Cabedelo a Campina Grande, divergente do trecho que leva da cidade de Campina Grande para o sertão do estado.

Visando ter homogeneidade nos dados, o trecho em estudo ficou restrito a Campina Grande até Cachoeira dos Índios, correspondente ao trecho do km 142 ao km 518,2.

Etapa 2: definição dos parâmetros a serem utilizados

A escolha dos parâmetros utilizados para o desenvolvimento da proposta, foi realizada tomando como base as classificações apresentadas nas tabelas de dados de acidentes da Polícia Rodoviária Federal (PRF), encontradas nos dados abertos do portal do departamento.

Foram adotados como parâmetros o traçado da rodovia, as condições meteorológicas, a fase do dia, o tipo de veículo envolvido no acidente e a idade do condutor, como apresentado na Tabela 1, visando ter na camada de saída a quilometragem propícia a acidentes de acordo com a variação dos parâmetros adotados.

Tabela 1: Parâmetros adotados na tabela da acidentes da PRF

Traçado da Rodovia	Condições meteorológicas	Tipo de veículo	Fase do dia	Idade do condutor
Curva	Céu claro	Carros e veículos pequenos	Amanhecer	Variando de 15 a 90 anos
Desvio temporário	Chuva	Motos	Anoitecer	
Interseção de vias	Garoa/chuvisco	Caminhões leves e Ônibus	Plena noite	
Ponte	Granizo	Caminhões pesados e Especiais	Pleno dia	
Reta	Neve			
Retorno Regulamentado	Nevoeiro/Neblina			
Rotatória	Nublado			
Túnel	Sol			
Viaduto	Vento			

Traçado da rodovia: esse parâmetro foi escolhido devido a influência direta que o traçado da rodovia tem com desempenho do veículo e visibilidade do condutor.

Condições meteorológicas: esse parâmetro foi escolhido uma vez que as condições climáticas podem alterar características da pista, como a aderência pneu/asfalto, como também afetar a visibilidade do condutor.

Fase do dia: esse parâmetro foi considerado importante por trazer a informação do momento do dia em que os acidentes ocorrem, podendo trazer comprometimento de visibilidade em determinados trechos no período de plena noite.

Tipo de veículo: classificado em 22 tipos diferentes, foram agrupados por categorias do DNIT que são carros e veículos pequenos, motos, caminhões leves e ônibus e caminhões pesados e especiais, esse parâmetro foi adotado por tratar diretamente do objeto envolvido no acidente.

Idade do condutor: este parâmetro varia de 15 a 90 anos e foi escolhido levando em consideração que menores de idade, conseqüentemente sem habilitação, ou idosos que possuam algum tipo de limitação podem ocasionar acidentes por imprudência.

Etapa 3: tratamento dos dados para serem usados na elaboração das RNA's

A tabela da PRF apresenta informações de acidentes do ano de 2007 até 2020. Para esse estudo foram utilizados os dados dos anos 2017 a 2019, por serem dados mais recentes disponíveis para consulta.

Com isso, deu-se início ao tratamento dos dados correspondentes aos parâmetros selecionados como entrada para a rede neural artificial. Dentre os parâmetros escolhidos, havia dados quantitativos e qualitativos, sendo ambos tratados das seguintes formas:

Dados qualitativos

Para os parâmetros que tinham uma característica ou atributo associado, sendo no presente estudo o traçado da rodovia, as condições meteorológicas, a fase do dia e o tipo de veículo, foram atribuídos valores numéricos (pesos) com base no quanto eles aumentam a probabilidade de um acidente ocorrer. Os pesos e suas justificativas de atribuição foram apresentados na Tabela 2 a seguir:

Tabela 2: Atribuição de pesos para cada parâmetro

Atribuição de pesos para cada parâmetro (de 0 a 1)		
Parâmetro	Peso	Justificativa

Traçado da rodovia		
Túnel	0,1	Menor incidência
Viaduto	0,2	Pouca incidência
Ponte	0,3	Pouca incidência
Retorno Regulamentado	0,4	Pouca incidência
Rotatória	0,5	Exige mais atenção do que em túneis, retornos, viadutos e pontes
Desvio temporário	0,6	Desvios não previstos pelo condutor
Interseção de vias	0,7	Cruzamentos tendem a ser mais perigosos
Reta	0,8	Maior incidência
Curva	0,9	Comprometimento da visibilidade
Condições meteorológicas		
Neve	0,1	Improvável de acontecer na região
Granizo	0,2	Improvável de acontecer na região
Vento	0,3	Pouco frequente na região
Nevoeiro/Neblina	0,4	Comprometimento da visibilidade
Sol	0,5	Comprometimento da visibilidade e maior incidência
Garoa/chuvisco	0,6	Comprometimento da visibilidade e pista escorregadia
Nublado	0,7	Maior incidência
Céu claro	0,8	Maior incidência
Chuva	0,9	Comprometimento da visibilidade e pista escorregadia
Fase do dia		
Anoitecer	0,3	Comprometimento da visibilidade
Amanhecer	0,5	Maior razão de casos com óbito em relação ao anoitecer
Plena noite	0,7	Comprometimento da visibilidade
Pleno dia	0,9	Pelo movimento ser maior durante o dia
Tipo de veículo		
Caminhões pesados e Especiais	0,2	Menor incidência
Caminhões leves e Ônibus	0,5	Grande fluxo em estradas (BR)
Carros e veículos pequenos	0,7	Maior incidência
Motos	0,9	Maior vulnerabilidade do condutor

Um parâmetro específico, o tipo de veículo, tem seus valores divididos em grupos maiores de acordo com a divisão do DNIT. Essa divisão está apresentada na Tabela 16 do anexo.

Dados quantitativos

Os parâmetros com característica quantitativa (aqueles que tem um valor numérico associado) foram idade do condutor e o quilômetro correspondente, usado como variável resposta (saída), sendo tratados pelo processo de normalização para que fosse evitado problemas devido às unidades e faixas de valores diferentes entre os parâmetros, tornando assim o trabalho computacional mais rápido.

O processo de normalização dos dados consistiu em reduzir os limites dos dados ao intervalo numérico entre 0 e 1. Dessa forma, os valores que antes variavam de 15 a 90 para a idade e valores de 142 km a 518,2 km para quilometragem, passaram a variar de 0 a 1.

Para isso, foi utilizada a normalização de dados com a função de distribuição normal, que depende da média e do desvio padrão dos dados, dada pela equação:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Onde:

x é a variável do parâmetro;

μ é a média dos valores;

σ é desvio padrão dos valores.

Para os valores obtidos foi considerada precisão de quatro casas decimais para que não conflitassem os valores entre as diferentes idades.

A média, o desvio padrão e os valores normalizados dos dados de faixa etária e da quilometragem são apresentados nas Tabelas 17, 18, 19 e 20 do anexo.

Etapa 4: Construção das RNA's

Após tratamento e normalização dos dados, iniciou-se a utilização do software de simulação de matrizes com pacote para redes neurais artificiais, para a elaboração, treino e testes da rede neural. Nele, primeiramente, importou-se as planilhas com os dados já tratados referentes aos valores normalizados dos parâmetros de entrada, a planilha com os valores normalizados de saída e a planilha com os valores normalizados a serem utilizados como teste após o treino, que foram retirados aleatoriamente do banco de dados.

Para a etapa de treino, os dados inseridos foram os referentes aos parâmetros de entrada como dados de entrada e os quilômetros como dados de saída. Em seguida, deu-se início às tentativas de diferentes arquiteturas de redes neurais, testando cada uma delas dez vezes uma vez que se trata de dados aleatórios e o resultado da rede pode ter grandes variações em decorrência disso.

Para todas as tentativas, o tipo de rede neural utilizada foi a multilayer perceptron (MLP), a função de treinamento definida foi a TRAINLM, que é o algoritmo de Levenberg-Marquardt, função padrão do software de simulação de matrizes com pacote para redes neurais artificiais aplicado, utilizado para resolver problemas de mínimos quadrados não-lineares, tendendo a ser rápido no treinamento de redes neurais (BARROS, 2018) a função de adaptação de aprendizado foi a LEARNNGDM e a função de performance MSE, que é um estimador que mede o quão perto a convergência da rede neural conseguiu chegar do objetivo (BARROS, 2018). Definidas as funções, deu-se início à montagem das arquiteturas a fim de chegar a um modelo com maior assertividade na previsão de dados. Para isso, existem três variáveis envolvidas na construção das redes: o número de camadas, a quantidade de neurônios em cada camada e a função de ativação.

Quanto ao número de camadas, variou-se de 2 a 4 visando resposta mais rápida do computador; para o número de neurônios, houve uma variação de 10 a 40 neurônios por camada, sempre prezando por combinação entre essas duas variantes que não exigissem tanto da máquina, uma vez que essas restrições foram adotadas levando em consideração o tempo que o computador leva para processar redes maiores e o quanto é exigido do processador da máquina.

Já a função de ativação pode ser TANSIG, LOGSIG ou PURELIN. Após realização de análise inicial de 5 arquiteturas diferentes, observou-se que as de melhor desempenho, ou seja, maior taxa de regressão, apresentavam apenas a TANSIG em sua configuração.

Após a modelagem de 30 arquiteturas de rede neural, analisou-se os valores e os gráficos de regressão e performance delas, onde cada uma delas é treinada e testada pelo menos 10 vezes por se tratar de dados aleatórios, obtendo-se os dados e os gráficos de regressão.

Foi realizado, em seguida, o teste com os valores retirados aleatoriamente da planilha inicial de parâmetros já normalizados, denominados no presente estudo como valores alvo, para cada uma das tentativas de cada arquitetura das redes neurais, comparando os resultados com

os alvos esperados, ou seja, observou-se se aquela determinada arquitetura proporcionou um alto índice de acertos sobre o que aconteceu nas situações reais.

Para isso, calculou-se a assertividade de cada rede para variações na faixa de 5km, 10km, 15km, 20km, 25km e 30km, tomando como base o módulo da diferença entre o valor alvo e o valor obtido nos testes, como mostrado na relação abaixo:

$$E = |Valor\ alvo - Valor\ teste|$$

Para cada arquitetura, foi determinada a média dos 10 testes realizados para cada uma das faixas consideradas. Em seguida, foram analisados os percentuais de acertos para cada uma das faixas, escolhendo aquele que tivesse maior frequência e representatividade na análise.

Para que a análise fosse feita por faixa de quilômetros, utilizou-se a função inversa da normalização, convertendo os valores de quilômetro para os valores reais correspondentes.

Etapa 5: Avaliação das RNA's desenvolvidas a partir do percentual de acerto

Tendo desenvolvido 30 arquiteturas de RNA's, separou-se a média de regressão de treino e de porcentagem de acertos dos 10 testes realizados para cada uma das faixas consideradas. Em seguida, foram selecionados os cinco melhores resultados para cada faixa de quilometragem, levando em consideração também a taxa de regressão do treino como critério de avaliação em casos que a porcentagem de acerto entre duas ou mais arquiteturas eram muito próximas.

Para escolha da arquitetura a ser utilizada, avaliou-se quais apareciam com maior incidência com melhores resultados nas faixas consideradas, ou seja, as redes que mais se aproximaram em termos de percentual de acerto dentro do trecho analisado.

Para a análise de dados e da rede, foram adotados três métodos:

O primeiro consistiu em definir diversas margens de erros com faixas de quilômetros variando de 5 a 30 km, ou seja, a rede neural computa como certo se o resultado obtido no teste estiver situado a uma distância máxima (para frente ou para trás) de 5 km, 10 km, 15 km, 20 km, 25 km ou 30 km, dependendo da faixa em análise, do valor do alvo.

O segundo método dividiu a rodovia em trechos de 5 km, 10km, 15km, 20km, 25km e 30km e a rede neural computa como certo se o resultado obtido no teste estiver exatamente dentro do mesmo trecho que o valor do alvo.

E por último, o terceiro método consistiu em dividir a rodovia em trechos entre municípios, sem tomar como parâmetro a extensão dos segmentos.

O trecho então foi subdividido em Campina Grande – Soledade, Soledade – São Mamede, São Mamede – Patos, Patos – Pombal, Pombal – Sousa, Sousa – Cajazeiras e Cajazeiras – Cachoeira dos Índios. A Tabela 3 mostra o quilômetro em que cada uma está situada na rodovia e a distância entre elas, também demonstradas nos mapas das Figuras 7 a 13.

Tabela 3: Subdivisão do trecho com base nos municípios e suas extensões

Município	Localização na BR-230 (km)	Distância até o próximo município (km)
Campina Grande	142	59,6
Soledade	201,6	102,4
São Mamede	304	28
Patos	332	63,1
Pombal	395,1	54,9
Sousa	450	39,8
Cajazeiras	489,8	22,2
Cachoeira dos índios	512	-

A escolha dos municípios se deu através dos números de acidentes que ocorreram nos trechos e a influência da cidade para a região, levando em consideração o tamanho populacional delas, apresentados na Tabela 4, e a distância entre os municípios.:

Tabela 4: População dos municípios paraibanos

Município	População (hab)
Campina Grande	412000
Patos	108192
Sousa	68723
Cajazeiras	62289
Pombal	32802
Soledade	15102
Cachoeira dos índios	10305
São Mamede	7702

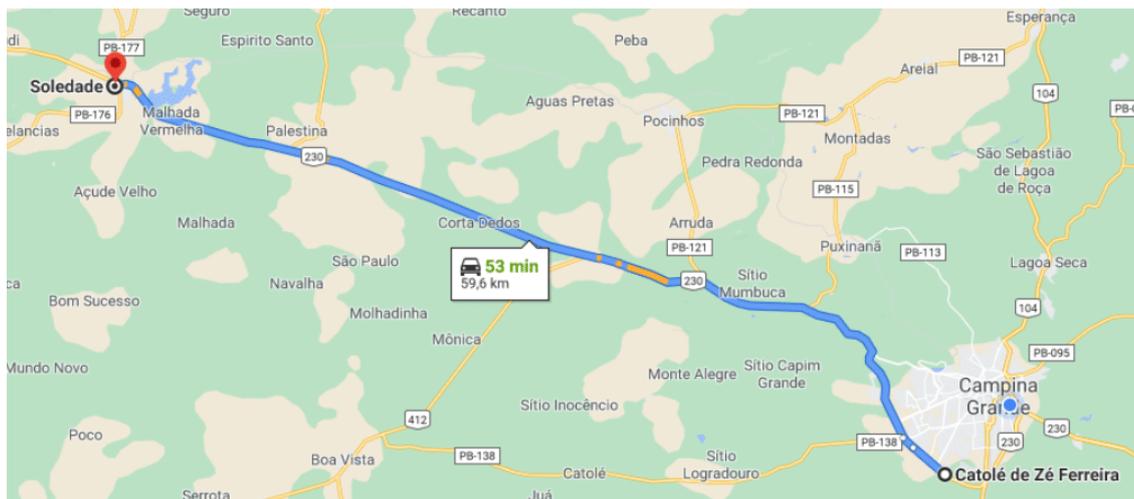


Figura 7: Trecho Campina Grande - Soledade (fonte: Google Maps)

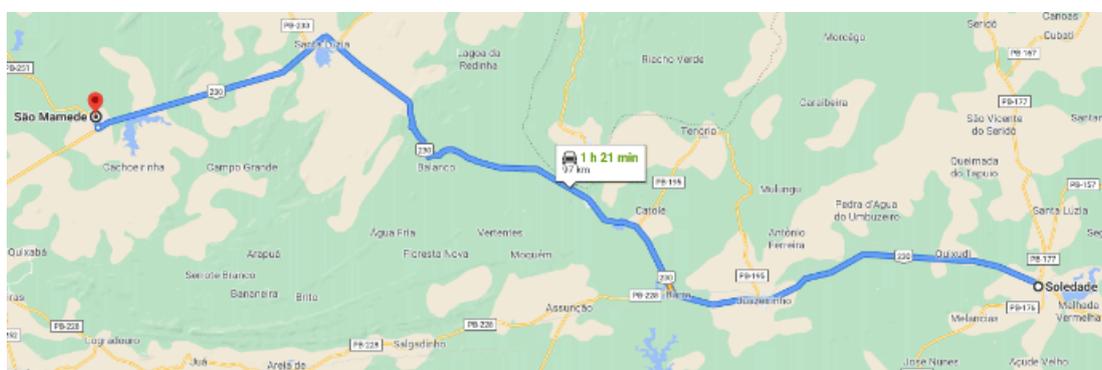


Figura 8: Trecho Soledade - São Mamede (fonte: Google Maps)

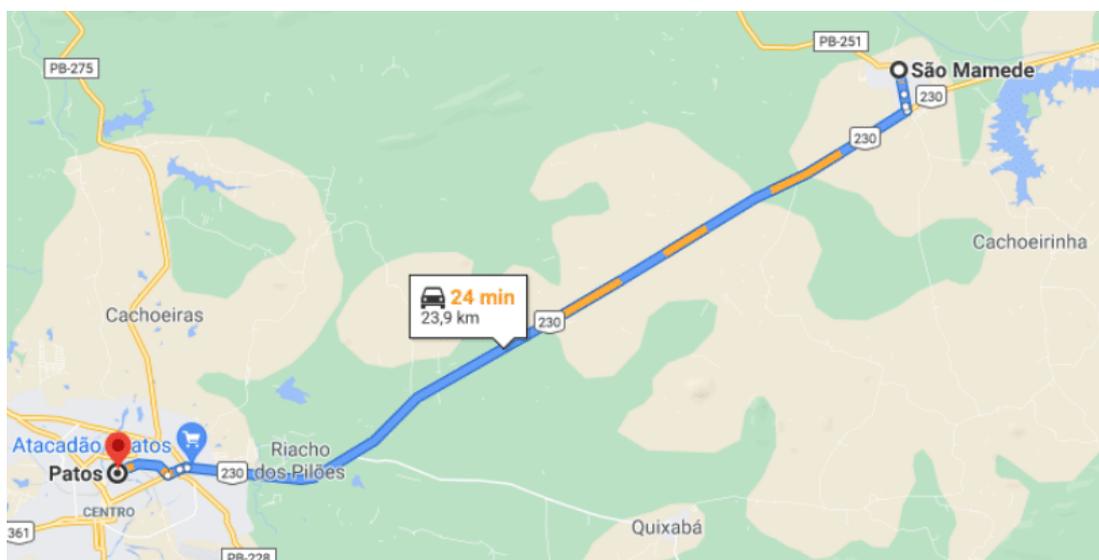


Figura 9: Trecho São Mamede - Patos (fonte: Google Maps)

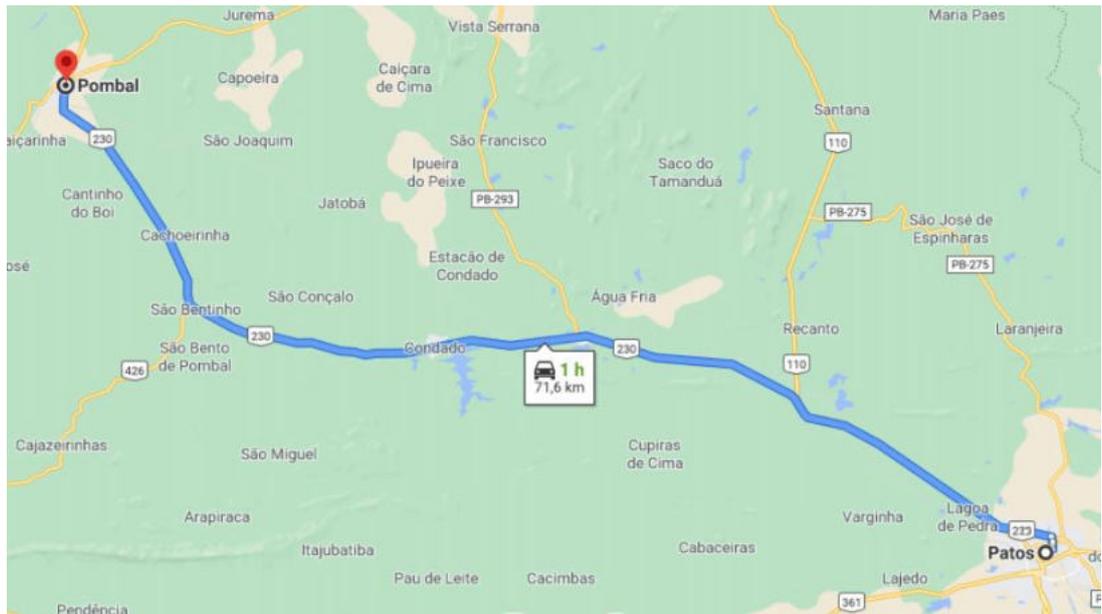


Figura 10: Trecho Patos - Pombal (fonte: Google Maps)



Figura 11: Trecho Pombal - Sousa (fonte: Google Maps)

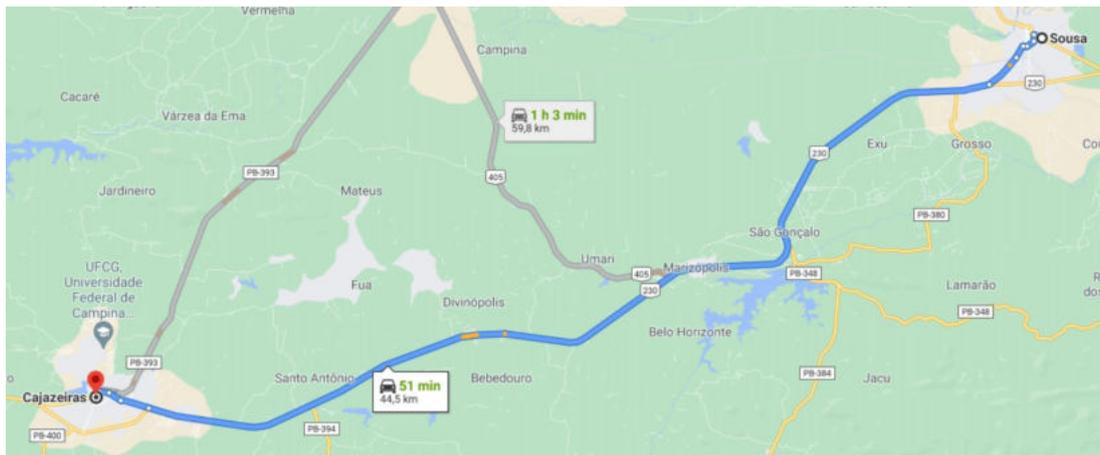


Figura 12: Trecho Sousa - Cajazeiras (fonte: Google Maps)

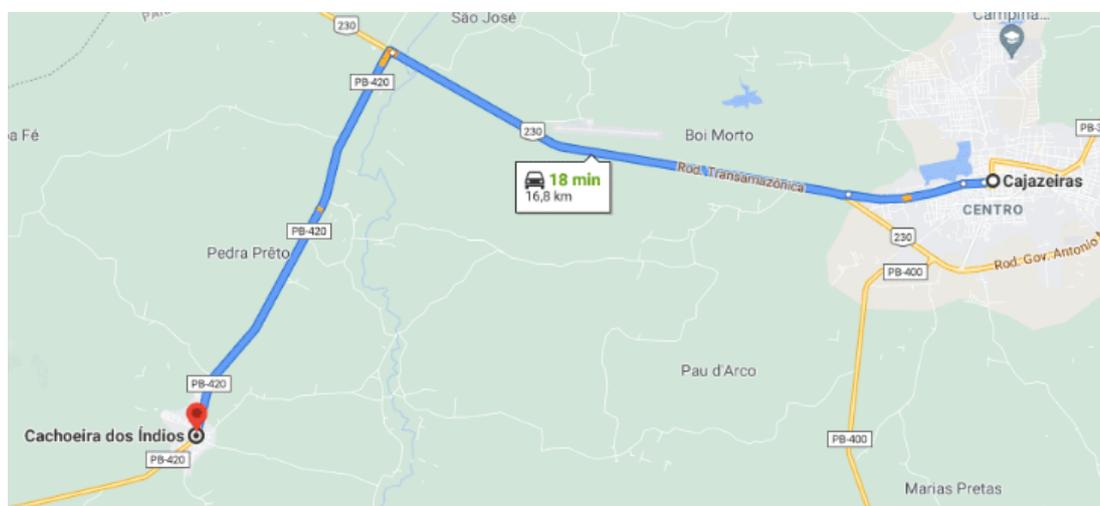


Figura 13: Trecho Cajazeiras - Cachoeira dos Índios (fonte: Google Maps)

6. RESULTADOS E DISCUSSÕES

6.1. ESCOLHA DA REDE

Para a análise de cada um dos métodos, foi utilizada a mesma rede neural, sendo necessário fazer testes para encontrar a que possuía melhor resultado em porcentagem de acertos.

As redes neurais treinadas e testadas, de forma aleatória, estão apresentadas na Tabela 5 com seus respectivos números de camadas, número de neurônios por camada e função de ativação:

Tabela 5: Arquiteturas das redes neurais

Rede	Número de camadas	Número de neurônios				Função de ativação			
		1ª camada	2ª camada	3ª camada	4ª camada	1ª camada	2ª camada	3ª camada	4ª camada
1	4	25	20	15	1	TANSIG	TANSIG	TANSIG	TANSIG
2	4	25	20	20	1	TANSIG	TANSIG	TANSIG	TANSIG
3	3	30	15	1	-	TANSIG	TANSIG	TANSIG	-
4	4	25	20	30	1	TANSIG	TANSIG	TANSIG	TANSIG
5	4	25	25	30	1	TANSIG	TANSIG	TANSIG	TANSIG
6	4	20	20	25	1	TANSIG	TANSIG	TANSIG	TANSIG
7	3	30	30	1	-	TANSIG	TANSIG	TANSIG	-
8	4	15	20	25	1	TANSIG	TANSIG	TANSIG	TANSIG
9	4	30	25	20	1	TANSIG	TANSIG	TANSIG	TANSIG
10	3	30	25	1	-	TANSIG	TANSIG	TANSIG	-
11	2	30	1	-	-	TANSIG	TANSIG	-	-
12	2	25	1	-	-	TANSIG	TANSIG	-	-
13	4	25	20	25	1	TANSIG	TANSIG	TANSIG	TANSIG
14	4	25	20	10	1	TANSIG	TANSIG	TANSIG	TANSIG
15	4	30	30	30	1	TANSIG	TANSIG	TANSIG	TANSIG
16	3	30	25	1	-	TANSIG	TANSIG	TANSIG	TANSIG
17	4	25	25	20	1	TANSIG	TANSIG	TANSIG	TANSIG
18	4	15	15	10	1	TANSIG	TANSIG	TANSIG	TANSIG
19	2	40	1	-	-	TANSIG	TANSIG	-	-
20	4	30	25	25	1	TANSIG	TANSIG	TANSIG	TANSIG
21	4	20	20	10	1	TANSIG	TANSIG	TANSIG	TANSIG
22	3	30	20	1	-	TANSIG	TANSIG	TANSIG	-

23	3	25	25	1	-	TANSIG	TANSIG	TANSIG	-
24	4	25	25	25	1	TANSIG	TANSIG	TANSIG	TANSIG
25	4	30	20	15	1	TANSIG	TANSIG	TANSIG	TANSIG
26	4	25	25	20	1	TANSIG	TANSIG	TANSIG	TANSIG
27	3	30	10	1	-	TANSIG	TANSIG	TANSIG	-
28	2	35	1	-	-	TANSIG	TANSIG	-	-
29	3	25	10	1	-	TANSIG	TANSIG	TANSIG	-
30	2	20	1	-	-	TANSIG	TANSIG	-	-

Na Tabela 5 é possível perceber que a maior parte das arquiteturas são constituídas de 4 camadas. Apesar de haver uma demora maior para que a rede seja treinada, elas apresentaram uma regressão maior em relação as de 3 ou 2 camadas.

Redes como a 10, que apresentaram uma regressão próxima às de 4 camadas, tiveram grande variação entre seus testes, apresentando dificuldade de generalização maior que as redes de 4 camadas.

Ao fazer as médias da regressão de treino e a porcentagem de acerto das faixas de quilometro (km) de cada uma das redes neurais, obteve-se o resultado apresentado na Tabela 6:

Tabela 6: Médias para R treino e percentual de acerto das redes neurais

Rede neural	R treino	Percentual de acerto					
		5 km	10 km	15 km	20 km	25 km	30 km
1	0,698981	3,1	6,2	10,2	12,7	15,3	16,6
2	0,700049	2,6	5,1	9	11,4	12,9	15,4
3	0,607427	2,5	5,2	9,3	11,5	14,1	16,4
4	0,732862	2,8	5,7	11,3	12,7	14,5	16,6
5	0,684758	3,1	5,7	9,8	12,1	15	17,8
6	0,66442	3	7,2	11,7	14,1	16,6	18,3
7	0,724334	1,6	4,1	9,4	11,8	13,7	16,5
8	0,690708	2	4,1	8,9	11,3	13,6	16,5
9	0,693086	1,5	4,9	9,2	11,1	12,9	15,1
10	0,710567	2,7	5	10,2	13,2	14,5	16,4
11	0,572611	2,9	6,2	9,1	12,5	14,3	16,4
12	0,553155	2,4	5,6	9,1	11,9	13,9	15,5
13	0,69026	1,8	4,5	9,4	11,7	13,5	15,7
14	0,711184	2,8	5,8	10,9	14,2	16	18,3
15	0,706726	2,9	5,3	9,3	11,9	14	16

16	0,713877	2,3	4,5	9,9	12,4	14,1	16,8
17	0,728905	2,7	5,1	10,3	12,7	14,6	17,2
18	0,671777	2,6	5,8	9,5	12,3	14,3	16,9
19	0,612871	2,3	4,8	8,6	11,5	13,2	15,1
20	0,687917	2	4,2	6,7	9,3	11,6	14,8
21	0,701691	3	6,4	10,9	13	14,7	17,4
22	0,69613	2,5	4,8	10,2	11,9	14	16,9
23	0,710578	1,8	4,5	9,3	11,8	13,6	16,4
24	0,726901	2,5	5,5	11,4	13,2	15,3	17,9
25	0,69702	3,2	5,6	10	12,4	14,7	16
26	0,72719	2	4,8	9,8	12,7	14,7	16,9
27	0,641702	2,1	4,2	8,6	11,4	13,4	15,6
28	0,58936	1,8	5,6	9,6	12	14,1	16,6
29	0,630737	2,4	4,6	8,9	11,5	13,6	15,6
30	0,518505	3,3	5,1	9	10,9	13,2	15

A Tabela 6 trouxe as médias dos 10 treinos e testes feitos com cada uma delas que, apesar de apresentarem resultados acima de 0,60 em sua maioria, apresentavam também regressões de treino abaixo de 0,50, provocando médias abaixo do esperado, que era próximo de 0,70.

Após análise das 30 redes neurais foram escolhidas as cinco consideradas melhores por apresentarem bons resultados para cada uma das faixas de quilometragem. As demais redes, por mais que tenham apresentado, em alguns casos, a regressão de treino mais elevada, não apresentaram uma porcentagem de acerto tão alta, como mostrado a seguir na Tabela 7.

Tabela 7: Cinco melhores resultados das redes neurais por faixa de quilometragem

Para 5 km			
REDE NEURAL	30	% de acerto	3,3
REDE NEURAL	25	% de acerto	3,2
REDE NEURAL	1	% de acerto	3,1
REDE NEURAL	5	% de acerto	3,1
REDE NEURAL	21	% de acerto	3
Para 10 km			
REDE NEURAL	6	% de acerto	7,2
REDE NEURAL	21	% de acerto	6,4
REDE NEURAL	1	% de acerto	6,2
REDE NEURAL	11	% de acerto	6,2

REDE NEURAL	14	% de acerto	5,8
Para 15 km			
REDE NEURAL	6	% de acerto	11,7
REDE NEURAL	24	% de acerto	11,4
REDE NEURAL	4	% de acerto	11,3
REDE NEURAL	14	% de acerto	10,9
REDE NEURAL	21	% de acerto	10,9
Para 20 km			
REDE NEURAL	14	% de acerto	14,2
REDE NEURAL	6	% de acerto	14,1
REDE NEURAL	24	% de acerto	13,1
REDE NEURAL	10	% de acerto	13,1
REDE NEURAL	21	% de acerto	13
Para 25 km			
REDE NEURAL	6	% de acerto	16,6
REDE NEURAL	14	% de acerto	16
REDE NEURAL	24	% de acerto	15,3
REDE NEURAL	1	% de acerto	15,3
REDE NEURAL	5	% de acerto	15
Para 30 km			
REDE NEURAL	14	% de acerto	18,3
REDE NEURAL	6	% de acerto	18,3
REDE NEURAL	24	% de acerto	17,9
REDE NEURAL	5	% de acerto	17,8
REDE NEURAL	21	% de acerto	17,4

Analisando a Tabela 7, observa-se que apenas três redes apresentaram maior incidência de melhores resultados para as faixas de quilômetros analisadas. Dessa forma, é possível notar que 3 arquiteturas aparecem 5 vezes: a rede 6, a rede 14 e a rede 21, conforme destacado na Tabela 8 a seguir:

Tabela 8: Três arquiteturas que apareceram com maior frequência

Redes que mais aparecem	R TREINO	Frequência
6	0,66442	5
14	0,711184	5
21	0,701691	5

As arquiteturas das redes 6, 14 e 21 estão demonstradas abaixo nas Figuras 14, 15 e 16, respectivamente:

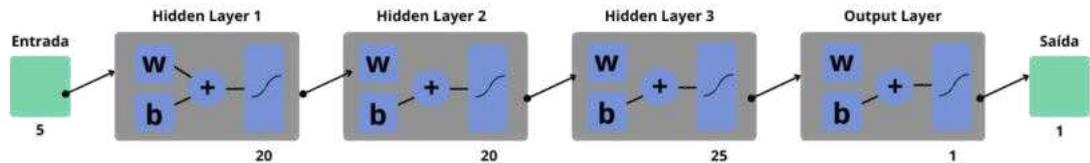


Figura 14: Arquitetura da rede neural 6 (fonte: autora)

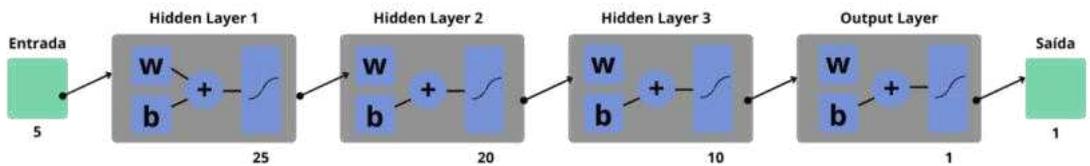


Figura 15: Arquitetura da rede neural 14 (fonte: autora)

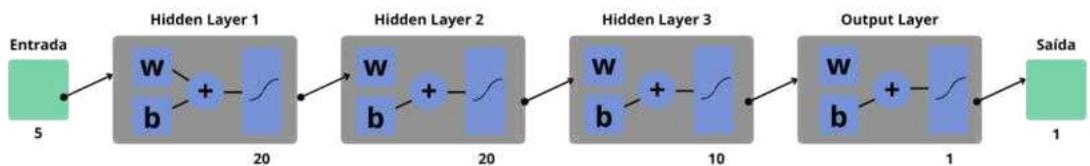


Figura 16: Arquitetura da rede neural 21 (fonte: autora)

Para a escolha de qual seria mais assertiva, foi levado em consideração que a rede 6 aparece 3 vezes como melhor resultado em porcentagem de acerto, correspondendo a 7,2% para 10 km, 11,7% para 15 km e 16,6% para 25 km, enquanto a rede 14 aparece duas vezes e a rede 21 não aparece como melhor resultado nenhuma vez. Além disso, nos casos em que a rede 14 obteve melhor resultado, é possível observar que a diferença para a rede 6 era de até 0,1, nos casos de 20 km e 30 km, como apresentado na Tabela 7.

Dessa forma, a arquitetura da rede 6 foi escolhida como a que melhor atinge os resultados, mesmo tendo a regressão de treino inferior ao da rede 14.

6.1. ANÁLISE DA REDE NEURAL

A rede neural 6 possui a seguinte arquitetura, mostrada na Figura 17:

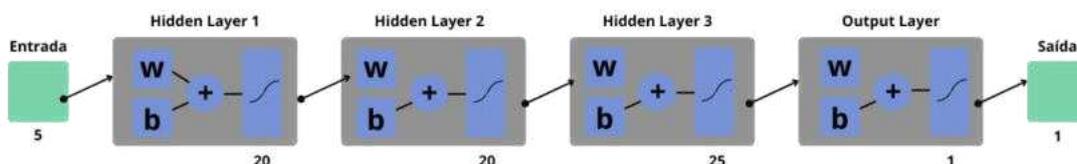


Figura 17: Arquitetura rede neural 6 (fonte: autora)

Composta por 4 camadas com 20, 20, 25 e 1 neurônios, respectivamente, e com função de ativação TANSIG para todas as camadas, a rede neural 6 apresentou os melhores resultados quando em comparação com as demais, como mostrado na Tabela 7 e justificado no tópico anterior. A comparação entre os dados alvos e os dados obtidos pela rede neural 6 estão na Tabela 21 do anexo.

Para garantir a homogeneidade dos dados, foram utilizadas as médias para comparação entre as redes. As médias da rede 6, obtidas nos treinos e testes para cada faixa de quilometragem, estão apresentadas na Tabela 9 e no gráfico da Figura 11:

Tabela 9: Resultado médio da rede neural 6

Resultado médio	
REDE NEURAL	6
R TREINO	0,66442
Acertos no teste (%)	
5 km	3
10 km	7,2
15 km	11,7
20 km	14,1
25 km	16,6
30 km	18,3

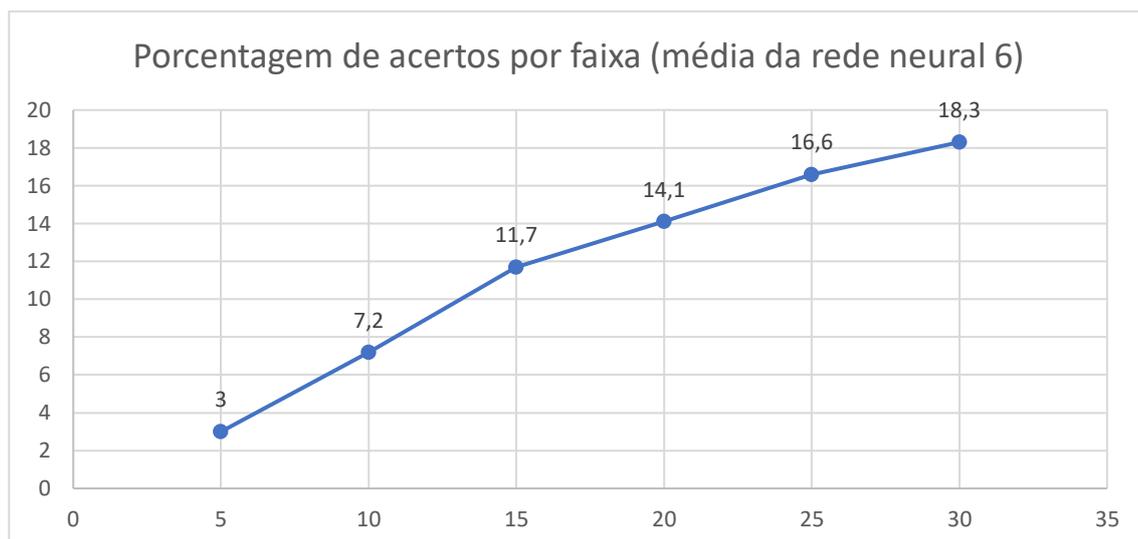


Figura 18: Porcentagem de acertos por faixa (média da rede neural 6)

Com o gráfico da Figura 18, é possível perceber que quanto maior a extensão do trecho analisado, maior é a porcentagem de acerto da rede neural.

6.2. ANÁLISE DO TRECHO

Ao analisar o trecho compreendido entre Campina Grande e Cachoeira dos Índios, nota-se a possibilidade de subdivisões dentro dele para uma melhor resposta de quais quilômetros da rodovia estão mais propícios à ocorrência de acidentes.

Levando em consideração a subdivisão do trecho nota-se, pelos dados da tabela de acidentes da PRF, que o trecho Campina Grande - Soledade possui maior número absoluto de acidentes, como observado na Tabela 10:

Tabela 10: Número de acidentes por trecho

Trecho	Número de acidentes			
	2017	2018	2019	Total
Campina Grande - Soledade	493	312	482	1287
Soledade - São Mamede	206	285	264	755

São Mamede - Patos	38	69	42	149
Patos - Pombal	154	216	161	531
Pombal - Sousa	200	165	143	508
Sousa - Cajazeiras	144	331	270	745
Cajazeiras - Cachoeira dos Índios	135	224	129	488

Ao considerar essas faixas no teste da rede neural 6, a porcentagem de acerto corresponde a 21%, como pode ser observado na Tabela 11. Essa oscilação, quando comparada ao acerto de 14,1% para faixas de 20 km e de 18,3% para faixas de 30 km no resultado médio da rede 6, é coerente observando que há variação de 22,2 km entre os municípios até 102,4 km, bem acima das faixas consideradas nas análises iniciais.

Tabela 11: Porcentagem de acerto com subdivisão em municípios

Trecho	Acertos por trecho	Comprimento do trecho (km)
Campina Grande - Soledade	8	59,6
Soledade - São Mamede	5	102,4
São Mamede - Patos	0	28
Patos - Pombal	4	63,1
Pombal - Sousa	1	54,9
Sousa - Cajazeiras	1	39,8
Cajazeiras - Cachoeira dos Índios	2	22,2
Acertos totais (%)	21	

No entanto, fazendo uma análise levando em consideração o tamanho dos trechos, visando retirar trechos muito curtos quando comparado aos demais, desconsiderou-se a

subdivisão em São Mamede e em Cajazeiras que, apesar de ser um município com mais influência que Cachoeira dos Índios, este último representa o ponto final de análise.

Dessa forma, na Tabela 12 estão os números de acerto por trecho e a porcentagem total de acerto para a nova subdivisão:

Tabela 12: Porcentagem de acertos por trecho com nova subdivisão

Trecho	Acertos por trecho	Comprimento do trecho (km)
Campina Grande - Soledade	8	59,6
Soledade - Patos	9	130,4
Patos - Pombal	4	63,1
Pombal - Sousa	1	54,9
Sousa - Cachoeira dos Índios	4	62
Acertos totais (%)	26	

Esse aumento na porcentagem para 26% se deu por apresentar segmentos maiores do que as faixas de 5 a 30 km utilizadas para a análise inicial, tornando a probabilidade de acerto da rede neural maior.

A Figura 19 mostra a extensão da BR-230 em análise, sinalizando os trechos com maior ocorrência de acidentes no período de 2017 a 2019.



Figura 19: Trecho de maior ocorrência de acidentes na BR - 230 entre Campina Grane e Cachoeira dos Índios (fonte: CNT)

Trazendo a análise para o ano de 2020, a partir dos dados fornecidos pelo painel de acidentes da Confederação Nacional do Transporte (CNT), é possível notar que esses trechos

continuaram tendo a maior incidência de acidentes quando comparados aos demais, sendo possível ver os pontos nas Figuras 20 a 24 a seguir:

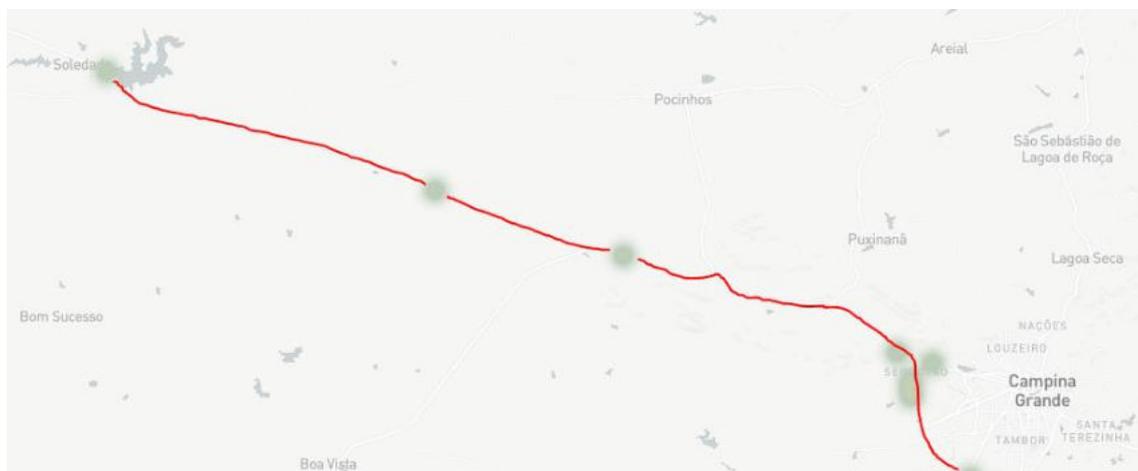


Figura 20: Trecho Campina Grande – Soledade. Total de 1287 acidentes de 2017 a 2019 (fonte: CNT)



Figura 21: Trecho Soledade – Patos. Total de 904 acidentes de 2017 a 2019. (Fonte: CNT)



Figura 22: Trecho Patos – Pombal. Total de 531 acidentes de 2017 a 2019. (Fonte: CNT)

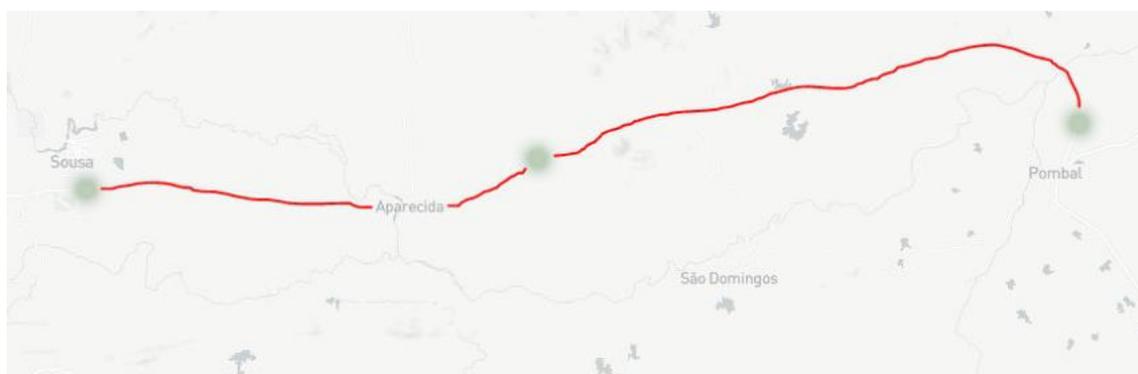


Figura 23: Trecho Pombal – Sousa. Total de 508 acidentes de 2017 a 2019. (Fonte: CNT)

O próximo trecho, correspondente ao trajeto de Sousa a Cajazeiras, chama atenção por ter a terceira maior incidência em número absoluto em comparação com os demais trechos, apresentando 745 acidentes de 2017 a 2018, mesmo sendo um trecho curto de 39,8 km. A Figura 24 mostra o trajeto com as incidências também de 2020, que levou em consideração o tipo e a causa dos acidentes:

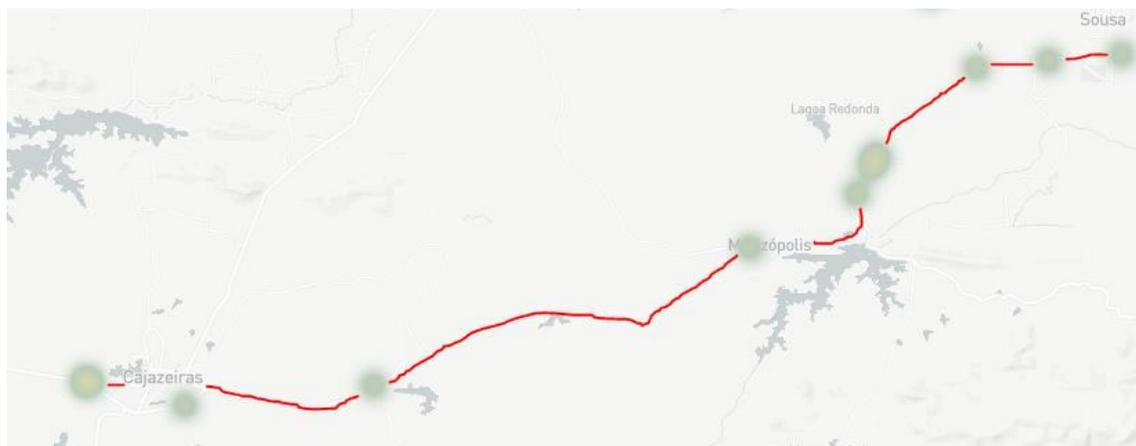


Figura 24: Trecho Sousa – Cajazeiras. Total de 745 acidentes de 2017 a 2019. (Fonte: CNT)

6.3. COMPARAÇÃO ENTRE OS MÉTODOS

A rede neural foi analisada de três formas: levando em consideração margens de erro com variações de 5 km a 30 km de distância, considerando trechos fixos subdivididos de 5 km a 30 km e, por fim, subdividida em trechos entre os principais municípios, como exposto inicialmente.

Para o primeiro método, a média do resultado da arquitetura 6, o acerto, ou seja, a previsão correta do km de ocorrência do acidente foi de 18,3%, como é possível verificar na Tabela 13.

Tabela 13: Média de resultados da rede neural 6 para primeiro método de análise

Melhor resultado da rede	
R TREINO	0,66442
Acertos no teste (%)	
5 km	3
10 km	7,2
15 km	11,7
20 km	14,1
25 km	16,6
30 km	18,3

O segundo método obteve os seguintes resultados para as diferentes extensões dos trechos, apresentados na Tabela 14:

Tabela 14: Resultados do segundo método de análise da rede neural

Acertos por extensão do trecho	
Extensão do trecho	Acerto (%)
5 km	0
10 km	1
15 km	6
20 km	12
25 km	12
30 km	15

Já no terceiro método, levando em conta as subdivisões entre os municípios, o resultado obtido está apresentado na Tabela 15:

Tabela 15: Resultados da terceira forma de análise da rede

Acertos por subdivisão do trecho	
Trecho	Acerto
CG - Soledade	8
Soledade - Patos	9
Patos - Pombal	4
Pombal - Sousa	1
Sousa - Cachoeira dos Índios	4
Total (%)	26

O gráfico comparando os três métodos encontra-se na Figura 25 a seguir:

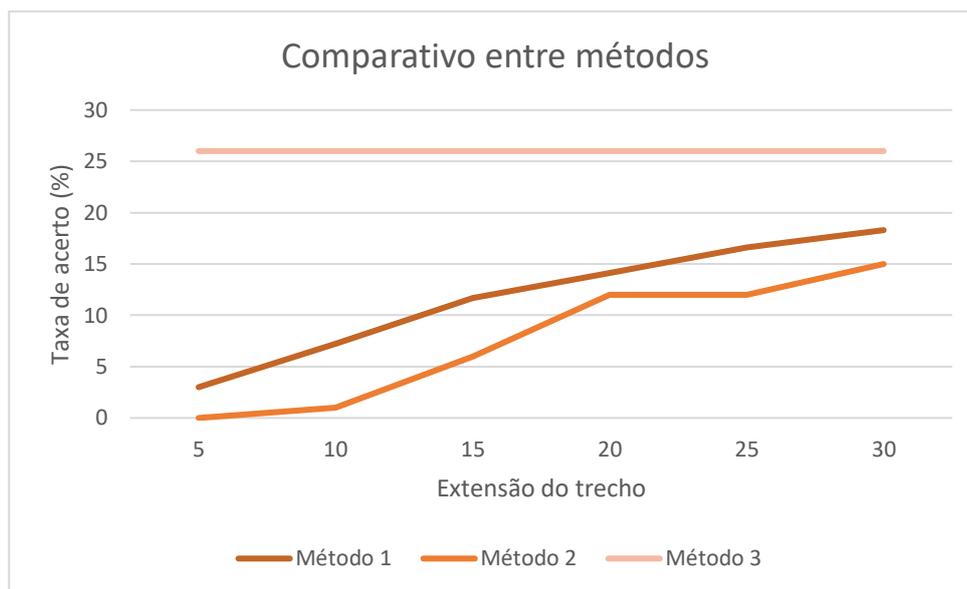


Figura 25: Gráfico comparativo entre métodos 1, 2 e 3

Vale ressaltar, para melhor interpretação do gráfico, que o método 3 tem seus trechos variando em quilometragens distintas das que foram definidas para os métodos 1 e 2, uma vez que eles correspondem a distância entre os municípios, apresentando apenas uma porcentagem total de assertividade.

Comparando os três métodos, apesar de apresentarem resultados próximos, principalmente na análise com margem de erro com faixas de quilometragem de 30 km (18,3% de acerto) e na subdivisão entre municípios (26% de acerto), essa última trouxe resultado mais satisfatório.

Contudo, a interpretação dessa taxa que, a priori, aparenta ser baixa, atingiu um bom resultado levando em consideração que os parâmetros utilizados foram muito específicos, bem como a saída, que corresponde a determinação do km no qual o acidente acontece.

Ao levar em consideração parâmetros que têm muitas variações que se diferenciam em valores de pesos bastante próximos, torna plausível que a rede tenha maiores dificuldades em diferenciá-los e enfrente um problema de generalização de dados.

Quanto à saída da rede, prever o exato quilômetro ou faixas curtas de quilometragens onde um acidente ocorrerá também é muito específico, demonstrando pelos dados que quanto maior essa faixa, maior a probabilidade de acerto da rede.

7. CONCLUSÕES

Levando em consideração os três métodos, foi possível observar que o desempenho da rede neural em trechos maiores tende a apresentar melhores ajustes e, conseqüentemente, melhores resultados.

Outro importante ponto observado foi que no método 2, quando se tem o trecho dividido em trechos menores de extensões fixas, o resultado da rede passa a ser bastante similar para faixas acima de 20 km de extensão, embasando a escolha do método 3 como a mais assertiva, uma vez que todos os municípios considerados na subdivisão apresentam distâncias maiores que 20 km entre si.

No entanto, apesar da rede neural apresentar um bom ajuste de dados de treinamento e de teste, sua interpretabilidade ainda é dificultada e a porcentagem de acertos é baixa. Mesmo apresentando bom resultado, levando em consideração a especificidade dos parâmetros, não corresponde a uma ótima eficiência na previsão de acidentes.

Essa interpretação se baseia também no fato de que a maioria dos acidentes rodoviários acontecem por conta de falhas humanas e veiculares, indo além da interpretação estatística. Dados da planilha de acidentes da PRF trazem como principais causas de acidentes a falta de atenção à condução, ingestão de álcool, ultrapassagem indevida, e outros fatores que não são viários-ambientais.

Contudo, a rede neural foi capaz de mapear os trechos com maior incidência de acidentes ao longo da BR – 230, no trecho situado no estado da Paraíba. Esse mapeamento é de suma importância para embasar as ações de fiscalização da polícia rodoviária federal na rodovia, planejamento de duplicação dos trechos mais críticos que hoje possuem pistas simples e demais projetos que visem mitigar o número de acidentes através de sinalizações ou revitalização das vias.

8. REFERÊNCIAS

KOVÁCS, Zsolt L. **Redes neurais artificiais: fundamentos e aplicações**. 4. ed. São Paulo: Livraria da Física, 2006.

KOPILER, Alberto A.; SILVA, Victor Navarro A. L. da; OLIVEIRA, Luiz Antônio A. de; LINDEN, Ricardo; SILVA, Luis Renato A. de A.; FONSECA, Bruno L. da C. **Redes Neurais Artificiais e suas aplicações no setor elétrico**. Macaé, 2019.

MOREIRA, Sandro. **Rede Neural Perceptron Multicamadas**. 2018. Disponível em: <https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9>. Acesso em: 24 nov. 2020.

FLECK, Leandro et al. REDES NEURAS ARTIFICIAIS: princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**, Medianeira, v. 1, n. 13, p. 47-57, jun. 2016.

Data Science Academy. **Deep Learning Book**, 2019. Disponível em: <http://www.deeplearningbook.com.br/>. Acesso em: 24 nov. 2020.

SILVA, Philippe Barbosa. **MODELOS DE PREVISÃO DE ACIDENTES EM RODOVIA BRASILEIRA DE PISTA DUPLA**. 2019. 342 f. Tese (Doutorado) - Curso de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, 2019.

BRASIL. INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **ACIDENTES DE TRÂNSITO NAS RODOVIAS FEDERAIS BRASILEIRAS: caracterização, tendências e custos para a sociedade**. Brasília: Governo Federal, 2015.

VITORINO, Rafaella Santos; BARROS, Kleber Napoleão Nunes de Oliveira; ALBUQUERQUE, Márcio Augusto de. **Modelo dinâmico de Poisson aplicado a dados de acidentes automobilísticos nas rodovias federais do estado da Paraíba**. Brazilian Journal Of Development. Campina Grande, p. 40583-40602. jun. 2020.

COSTA, J. O. **Desenvolvimento dum modelo de previsão de acidentes**. Tese de Doutorado em Segurança Rodoviária, Escola de Engenharia, Universidade do Minho, Guimarães, Portugal, 2013.

BARROS, Aluísio J. D.; AMARAL, Rodrigo L.; OLIVEIRA, Maria Simone B.; LIMA, Scilla C.; GONÇALVES, Evandro V. **Acidentes de trânsito com vítimas: sub-registro, caracterização e letalidade**. Scielo, Pelotas, v. 1, n. 1, p. 1-1, fev. 2003.

Instituto Brasileiro de Geografia e Estatística. **Pesquisa Nacional de Saúde 2013: acesso e utilização dos serviços de saúde, acidentes e violências.** Brasil, grandes regiões e Unidades da Federação. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2015.

Departamento de Vigilância de Doenças e Agravos Não Transmissíveis e Promoção da Saúde, Secretaria de Vigilância em Saúde, Ministério da Saúde. Viva: **Vigilância de Violências e Acidentes: 2013 e 2014.** Brasília: Ministério da Saúde; 2017.

Malta DC, Felisbino-Mendes MS, Machado IE, Passos VMA, Abreu DMX, Ishitani LH, et al. **Fatores de risco relacionados à carga global de doença do Brasil e Unidades Federadas,** 2015. Rev Bras Epidemiol 2017; 20 Suppl 1:217-32.

Andrade, Flávia Reis de e Antunes, José Leopoldo Ferreira. **Tendência do número de vítimas em acidentes de trânsito nas rodovias federais brasileiras antes e depois da Década de Ação pela Segurança no Trânsito.** Cadernos de Saúde Pública. v. 35, n. 8.

MATSUNAGA, Victoria Yukie. **Curso de Redes Neurais utilizando o MATLAB.** Belém, 2012.

SINAY, M. C. F. & TAMAYO, A. S. **Segurança Viária: Uma Visão Sistêmica.** Anais do III Rio de Transportes, Rio de Janeiro, Brasil. 2005.

SAVOLAINEN, P. T., MANNERING, F. L., LORD, D. & QUDDUS, M. A. (2011) **The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives.** Accident Analysis and Prevention.

HAUER, E. **Statistical Road Safety Modeling.** Transportation Research Record, n. 1897, p. 81-87. 2004.

HAUER, E., COUNCIL, F. & MOHAMMEDSHAD, Y. **Safety models for urban fourlane undivided road segments.** Transportation Research Record, n. 1897, p. 96-105. 2004.

WANG, C., QUDDUS, M. A. & ISON, S. G. **Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model.** Accident Analysis and Prevention, v. 43(6), p. 1979-1990. 2011.

CHANG, L. **Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network.** Safety Science, v. 43, p. 541-557. 2005.

COSTA, J. O. **Desenvolvimento dum modelo de previsão de acidentes. Tese de Doutoramento em Segurança Rodoviária**, Escola de Engenharia, Universidade do Minho, Guimarães, Portugal. 2003.

LORD, D. & MANNERING, F. **The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives**. Transportation Research Part A, v. 44, p. 291–305. 2010.

LI, H., GRAHAM, D. J. & MAJUMDAR, A. **The effects pf congestion charging on road traffic casualties: A causal analysis using difference-in-difference estimation**. Accident Analysis and Prevention, v. 49, p. 366-377. 2012.

CHAPELLE, O., SCHÖLKOPF, B., ZIEN, A. **Semi-supervised learning**. MIT Press, Cambridge. 2006.

HAYKIN, S. **Redes Neurais - Princípios e Práticas**. BOOKMAN, São Paulo, 2ª ed. 2001.

MIRANDA, F. A.; FREITAS, S. R. C.; FAGGION, P. L. **Integração e interpolação de dados de anomalias ar livre utilizando-se a técnica de RNA e krigagem**. Boletim de Ciências Geodésicas, v.15, n.3, p. 428-443, 2009.

MACHADO, W, C.; FONSECA JÚNIOR, E. S. **Redes Neurais Artificiais aplicadas na previsão do VTEC no Brasil**. Boletim de Ciências Geodesicas, v.19, n.2, p. 227-246, 2013.

AMBRÓSIO, P. E. **Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares**. Ribeirão Preto – SP. Dissertação (Mestrado) – Faculdade de Filosofia, Universidade de São Paulo. 2002.

BRAGA, A. de P., LUDERMIR, T. B., CARVALHO, A. C. P. de L. F., **Redes Neurais Artificiais teoria e aplicações**, Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., 2000.

EYNG, E. **Controle Feedforward-Feedback aplicado às colunas de absorção de produção de etanol por fermentação**. 2008. Campinas, SP. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia Química, Universidade Estadual de Campinas.

BARROS, Victor Pedroso Ambiel. **Avaliação do desempenho de algoritmos de retropropagação com redes neurais artificiais para a resolução de problemas não-lineares**. Ponta Grossa. Dissertação – Universidade Tecnológica do Paraná. 2018.

ANEXO

Tabela 16: Classificação do tipo de veículo

Classificação do tipo de veículo em grupos	
Categoria	Veículo
Carros e veículos pequenos	Automóvel
	Utilitário
	Caminhonete
	Camioneta
	Carro de mão
Motos	Bicicleta
	Motocicleta
	Motoneta
	Ciclomotor
Caminhões leves e Ônibus	Caminhão
	Ônibus
	Micro-ônibus
	Carroça-charrete
Caminhões pesados e Especiais	Caminhão-trator
	Chassi-plataforma
	Quadriciclo
	Reboque
	Semireboque
	Trator de esteira
	Trator de rodas
	Trator misto
Triciclo	

Tabela 17: Média e desvio padrão dos dados de faixa etária

Média	39,21352
Desvio padrão	13,88838

Tabela 18: Normalização da faixa etária

Normalização faixa etária	
15	0,0406
16	0,0473
17	0,0549
18	0,0633
19	0,0728
20	0,0833
21	0,0949
22	0,1076
23	0,1215
24	0,1367
25	0,1531
26	0,1707
27	0,1896
28	0,2097
29	0,2310
30	0,2535
31	0,2771
32	0,3017
33	0,3273
34	0,3537
35	0,3808
36	0,4085
37	0,4367
38	0,4652
39	0,4939

40	0,5226
41	0,5512
42	0,5795
43	0,6074
44	0,6348
45	0,6615
46	0,6875
47	0,7125
48	0,7365
49	0,7595
50	0,7813
51	0,8020
52	0,8214
53	0,8396
54	0,8565
55	0,8722
56	0,8866
57	0,8998
58	0,9119
59	0,9229
60	0,9328
61	0,9416
62	0,9496
63	0,9566
64	0,9628
65	0,9683
66	0,9731
67	0,9773
68	0,9809
69	0,9840
70	0,9867
71	0,9890

72	0,9909
73	0,9925
74	0,9939
75	0,9950
76	0,9960
77	0,9967
78	0,9974
79	0,9979
80	0,9983
83	0,9992
89	0,9998

Tabela 19: Média e desvio padrão do quilômetro

Média	335,0487
Desvio padrão	130,1221

Tabela 20: Normalização dos dados de quilômetros

Normalização dos quilômetros					
Valor	Normalizado	Valor	Normalizado	Valor	Normalizado
142	0,0690	239,5	0,2314	406,7	0,7091
146,3	0,0735	240	0,2326	406,8	0,7093
147,8	0,0751	240,2	0,2330	406,9	0,7096
150	0,0775	240,6	0,2340	407	0,7099
150,6	0,0782	240,7	0,2342	407,1	0,7101
150,7	0,0783	242	0,2373	407,5	0,7112
150,8	0,0784	243,6	0,2411	407,9	0,7122
151	0,0786	243,7	0,2413	408	0,7125
151,5	0,0792	245,5	0,2457	409	0,7151
151,6	0,0793	245,6	0,2459	410,6	0,7193
151,8	0,0795	246	0,2469	411	0,7203
151,9	0,0796	246,6	0,2483	411,9	0,7226
152	0,0798	247,2	0,2498	412	0,7229

152,1	0,0799	248,6	0,2532	412,2	0,7234
152,5	0,0803	249,8	0,2562	413,4	0,7265
152,6	0,0804	250,6	0,2582	414,9	0,7303
152,7	0,0806	251	0,2592	415	0,7305
152,8	0,0807	251,8	0,2612	416,8	0,7351
152,9	0,0808	254,5	0,2680	420,1	0,7433
153	0,0809	254,8	0,2687	422,6	0,7495
153,1	0,0810	256,7	0,2735	423,3	0,7512
153,6	0,0816	257	0,2743	424,2	0,7534
153,8	0,0818	258,3	0,2777	427	0,7601
153,9	0,0819	258,5	0,2782	428,8	0,7644
154	0,0821	259,5	0,2808	428,9	0,7646
154,4	0,0825	259,6	0,2810	432,7	0,7735
154,5	0,0826	260	0,2821	434,8	0,7783
154,6	0,0828	262	0,2873	437	0,7833
154,9	0,0831	263	0,2899	439	0,7878
155	0,0832	263,1	0,2902	441,5	0,7933
155,2	0,0835	263,5	0,2912	442,6	0,7958
155,3	0,0836	264,2	0,2931	443,6	0,7979
155,5	0,0838	264,3	0,2933	445,7	0,8024
155,7	0,0841	267,8	0,3026	446	0,8031
155,8	0,0842	269	0,3059	446,2	0,8035
155,9	0,0843	272,9	0,3165	446,6	0,8044
156	0,0844	273	0,3167	446,9	0,8050
156,2	0,0846	275	0,3222	447	0,8052
156,3	0,0848	275,5	0,3236	449,8	0,8111
156,4	0,0849	276,5	0,3264	450,8	0,8131
156,6	0,0851	277,1	0,3280	451,2	0,8140
156,7	0,0852	277,3	0,3286	451,3	0,8142
156,8	0,0854	277,8	0,3300	451,4	0,8144
156,9	0,0855	278	0,3305	453,7	0,8191
157	0,0856	278,4	0,3317	454,8	0,8213

157,2	0,0858	278,5	0,3319	455	0,8217
157,4	0,0861	279,7	0,3353	455,3	0,8223
157,5	0,0862	280,8	0,3384	456	0,8237
157,7	0,0865	280,9	0,3387	458	0,8276
157,8	0,0866	281,6	0,3406	458,3	0,8282
157,9	0,0867	289,2	0,3623	458,6	0,8288
158	0,0868	289,8	0,3640	458,9	0,8294
158,1	0,0869	289,9	0,3643	459	0,8296
158,2	0,0871	290,9	0,3672	459,7	0,8310
158,3	0,0872	291	0,3675	459,9	0,8313
158,5	0,0874	291,1	0,3678	460	0,8315
158,6	0,0875	291,5	0,3689	460,5	0,8325
158,7	0,0877	292	0,3704	460,8	0,8331
158,8	0,0878	292,2	0,3710	461	0,8335
159	0,0880	292,8	0,3727	461,2	0,8338
159,2	0,0883	292,9	0,3730	462	0,8354
159,4	0,0885	293	0,3733	462,2	0,8358
159,5	0,0887	293,2	0,3739	462,7	0,8367
159,6	0,0888	293,6	0,3750	462,8	0,8369
159,9	0,0891	293,7	0,3753	463,1	0,8375
160	0,0893	294,7	0,3782	463,3	0,8378
160,5	0,0899	295,9	0,3818	463,4	0,8380
160,6	0,0900	296	0,3821	464	0,8392
160,8	0,0903	297,1	0,3853	464,2	0,8395
161	0,0905	299,3	0,3918	464,4	0,8399
161,6	0,0913	300,1	0,3941	464,5	0,8401
162	0,0918	304	0,4057	464,6	0,8403
162,1	0,0919	305,9	0,4114	464,7	0,8405
162,5	0,0924	306	0,4117	464,8	0,8407
162,6	0,0925	308,2	0,4183	464,9	0,8408
162,8	0,0928	308,8	0,4201	465,4	0,8418
163	0,0930	309,5	0,4222	466	0,8429

164	0,0943	310	0,4237	466,1	0,8431
164,4	0,0949	312,1	0,4300	466,3	0,8434
164,5	0,0950	312,9	0,4324	466,6	0,8440
165	0,0956	313	0,4327	466,7	0,8442
165,6	0,0964	313,2	0,4333	466,8	0,8444
166	0,0969	314	0,4357	466,9	0,8445
166,2	0,0972	314,5	0,4373	467,5	0,8456
166,3	0,0973	317,2	0,4454	467,7	0,8460
166,5	0,0976	317,9	0,4476	467,8	0,8462
166,7	0,0979	320,6	0,4558	467,9	0,8464
166,8	0,0980	321	0,4570	469	0,8484
167	0,0983	321,7	0,4591	469,7	0,8496
167,5	0,0989	323,8	0,4656	470,8	0,8516
167,7	0,0992	325,1	0,4695	471,6	0,8530
168	0,0996	325,9	0,4720	471,8	0,8534
168,1	0,0997	327,6	0,4772	471,9	0,8535
168,3	0,1000	327,7	0,4775	472	0,8537
168,4	0,1001	328	0,4784	472,6	0,8548
168,5	0,1003	328,7	0,4805	472,7	0,8549
169	0,1010	328,9	0,4812	473,8	0,8569
169,1	0,1011	329,7	0,4836	473,9	0,8570
169,2	0,1012	330,5	0,4861	474,3	0,8577
169,4	0,1015	331,2	0,4882	475,1	0,8591
169,5	0,1016	331,9	0,4903	475,9	0,8605
169,7	0,1019	332	0,4907	476	0,8606
169,8	0,1021	332,2	0,4913	476,1	0,8608
169,9	0,1022	332,3	0,4916	476,8	0,8620
170	0,1023	332,7	0,4928	477	0,8623
170,3	0,1027	333,5	0,4953	477,9	0,8639
170,9	0,1036	334	0,4968	478,3	0,8645
171,2	0,1040	334,7	0,4989	479,2	0,8660
171,5	0,1044	334,8	0,4992	479,5	0,8665

172	0,1051	335	0,4999	480,3	0,8678
172,2	0,1054	335,5	0,5014	481	0,8690
172,8	0,1062	336	0,5029	481,1	0,8692
172,9	0,1064	336,7	0,5051	481,5	0,8698
173	0,1065	337,1	0,5063	483	0,8722
173,5	0,1072	337,7	0,5081	483,7	0,8734
173,9	0,1078	338,6	0,5109	485	0,8754
174	0,1079	338,7	0,5112	486,3	0,8775
174,2	0,1082	339	0,5121	487,7	0,8796
175,9	0,1107	341	0,5182	488	0,8801
176	0,1108	342,2	0,5219	488,7	0,8812
177,5	0,1130	342,5	0,5228	489,8	0,8828
178,2	0,1140	343,5	0,5259	492,6	0,8870
179	0,1152	345	0,5305	493,6	0,8885
181,6	0,1191	345,1	0,5308	493,9	0,8889
182	0,1198	346	0,5335	494,9	0,8904
182,5	0,1205	346,2	0,5341	495	0,8905
183	0,1213	346,3	0,5345	495,1	0,8907
183,1	0,1215	346,4	0,5348	495,2	0,8908
183,2	0,1216	346,6	0,5354	496,3	0,8924
184	0,1229	350	0,5457	496,4	0,8925
184,9	0,1243	351	0,5488	496,6	0,8928
185,1	0,1246	351,3	0,5497	496,7	0,8929
186,5	0,1268	351,9	0,5515	496,9	0,8932
188	0,1292	352,8	0,5543	497,3	0,8938
190	0,1325	352,9	0,5546	497,7	0,8943
190,1	0,1327	354,2	0,5585	497,9	0,8946
190,7	0,1336	354,9	0,5606	499,3	0,8966
190,8	0,1338	360,4	0,5772	500	0,8975
191,7	0,1353	360,8	0,5784	500,6	0,8984
192,3	0,1363	362	0,5820	500,7	0,8985
193,8	0,1388	362,9	0,5847	501,1	0,8990

196,2	0,1430	363	0,5850	501,3	0,8993
198	0,1461	363,4	0,5862	501,8	0,9000
198,9	0,1477	364	0,5880	502	0,9003
199	0,1479	365,2	0,5916	502,2	0,9005
200	0,1497	365,3	0,5919	502,7	0,9012
200,1	0,1498	366,5	0,5955	502,9	0,9015
201,6	0,1525	367	0,5970	503	0,9016
202,8	0,1547	367,3	0,5979	503,1	0,9017
203	0,1551	367,5	0,5985	503,2	0,9019
204	0,1569	368,4	0,6011	503,3	0,9020
204,6	0,1580	368,5	0,6014	503,9	0,9028
204,8	0,1584	368,9	0,6026	504	0,9029
205,5	0,1597	369	0,6029	504,5	0,9036
206,7	0,1620	369,5	0,6044	504,7	0,9038
208	0,1644	370,7	0,6080	504,9	0,9041
208,4	0,1652	371	0,6088	505	0,9042
212,2	0,1726	372,2	0,6124	505,1	0,9044
212,3	0,1728	375	0,6206	505,2	0,9045
212,9	0,1739	375,9	0,6232	505,6	0,9050
213	0,1741	376	0,6235	505,7	0,9052
213,1	0,1743	376,1	0,6238	505,8	0,9053
213,7	0,1755	376,2	0,6241	506	0,9055
213,8	0,1757	377,1	0,6267	506,1	0,9057
214	0,1761	377,3	0,6273	506,2	0,9058
214,6	0,1773	378,7	0,6314	506,5	0,9062
214,9	0,1779	380,7	0,6371	507	0,9068
215	0,1781	381,4	0,6392	507,6	0,9076
215,9	0,1799	384	0,6466	508	0,9081
216,4	0,1809	384,8	0,6489	508,1	0,9082
216,9	0,1819	385,1	0,6498	508,2	0,9084
217	0,1821	386	0,6523	508,3	0,9085
217,4	0,1830	390	0,6636	508,4	0,9086

221	0,1904	390,9	0,6661	508,7	0,9090
221,1	0,1906	391,1	0,6667	508,8	0,9091
221,4	0,1912	391,6	0,6681	509	0,9094
222	0,1925	392,6	0,6709	511	0,9118
223,7	0,1961	395,1	0,6778	511,7	0,9127
224	0,1967	396	0,6803	512	0,9131
225	0,1989	396,4	0,6814	512,1	0,9132
227,2	0,2036	397,2	0,6835	512,2	0,9133
228,3	0,2060	398,7	0,6876	512,6	0,9138
230,5	0,2109	399,7	0,6904	513,6	0,9150
231,1	0,2122	401,9	0,6963	513,8	0,9152
231,2	0,2124	402	0,6966	513,9	0,9154
231,9	0,2140	402,5	0,6979	514	0,9155
232	0,2142	403	0,6992	514,1	0,9156
232,4	0,2151	404	0,7019	515,1	0,9168
233	0,2164	404,7	0,7038	515,6	0,9174
234,2	0,2192	405	0,7046	516	0,9178
234,8	0,2205	405,8	0,7067	516,5	0,9184
235,6	0,2224	406	0,7072	516,7	0,9186
237,5	0,2267	406,1	0,7075	516,9	0,9189
239	0,2302	406,2	0,7077	517,1	0,9191
239,4	0,2311	406,3	0,7080	518,2	0,9204

Tabela 21: Valores alvos e valores obtidos pela rede neural 6

Valores alvo (em km)	Valores obtidos (em km)
494,9	371,7946044
494,9	335,5085649
165	327,1015461
165	399,2895857
177,5	518,0898138

469,7	159,3205527
160,6	146,1789444
337,1	236,856584
454,8	371,7946044
256,7	384,6959041
406	218,6558156
338,6	376,4077537
411	431,2832635
157,7	331,8290638
339	446,3620947
191,7	332,1520591
330,5	443,7648038
330,5	150,0023954
477	482,5185994
239,5	221,1741657
147,8	328,2319074
507	142,1645443
168	494,6055544
338,7	160,0758589
390	348,7599959
228,3	411,1870642
228,3	337,5962
156,8	144,0657357
363	376,6891654
155,9	411,1870642
155,9	144,6908101
157,4	489,9424276
158,5	251,0783153
157	318,8650013
208	366,2192676
147,8	141,9999948
157	293,4081696

455	266,0479378
334,8	331,8290638
154	465,2435094
275,5	337,6777581
404,7	176,1194042
168	179,8896162
245,5	271,0912965
291	305,3697442
156,9	141,9999948
251	303,5019957
407	142,0480203
160,6	497,5302763
181,6	443,6261447
467,7	376,6411058
161,6	142,1762861
408	311,8864429
231,2	206,1761129
156,4	308,0917586
502	508,4823015
337,7	148,4616815
369	497,6726229
155	427,620965
152,8	433,1107186
293	443,4645315
160,8	518,1775809
165	402,7426931
472,6	517,9320426
169,5	516,1882382
167,5	218,163504
240	387,3195629
328,9	338,0562015
153	363,4368453

190,8	477,85256
497,9	141,9999948
504,9	501,5032944
156	143,7462528
466,9	407,6423837
407	366,2192676
505,8	474,284883
292	141,9999948
204,6	308,811154
170,9	142,0969988
365,3	374,9249305
404	386,2358897
325,1	422,9949357
221	316,1692846
377,1	141,9999948
368,4	145,1517826
402,5	518,2039272
172,9	377,5965059
503,9	518,2039272
416,8	141,9999948
230,5	314,6087929
508,1	220,1421286
295,9	262,3482221
501,3	399,2895857
245,6	367,0254586
173	461,3575748
153,8	190,5469932
155,3	335,5085649
497,7	331,4375276
169,8	497,5871916
154	331,8290638