

**REDES NEURAS APLICADAS À PREVISÃO DO VOLUME DAS CHEIAS NO
MUNICÍPIO DE CATAGUASES – MG**
**NEURAL NETWORKS APPLIED TO FORECASTING THE FLOOD VOLUME IN
THE MUNICIPALITY OF CATAGUASES - MG**

Fernanda Florindo. Rede de Ensino Doctum, Unidade Cataguases.
fernandafiorindo1991@gmail.com. Graduando em Sistemas de informação.

Geraldo Furtado Neto. Rede de Ensino Doctum, Unidade Cataguases.
prof.geraldo.neto@doctum.edu.br.

RESUMO

As Redes Neurais Artificiais têm sido utilizadas em diversas áreas na sociedade a fim de possibilitar e facilitar diversas tarefas, seja pela sua capacidade de produzir ou calcular, esta última possibilita a previsão de acontecimentos. O município de Cataguases – Minas Gerais, em época de grande volume de chuvas, é fortemente atingido por enchentes, que causam diversos problemas à população, que não possui uma informação assertiva sobre o volume que o rio irá atingir. Desse modo, este trabalho visa construir uma rede neural artificial para prever o volume aproximado que o Rio Pomba irá atingir na época de cheias na cidade de Cataguases. Para tanto, desenvolveu-se uma rede neural utilizando a linguagem Python, que foi treinada através de dados fornecidos pela Agência Nacional de Águas. Ao final da análise dos resultados das métricas de avaliação percebeu-se que a Rede Neural é um método capaz de prever o nível das inundações.

Palavras Chave: Rede Neural Artificial. Cheias em Cataguases. Previsão.

ABSTRACT

Artificial Neural Networks have been used in several areas in society in order to enable and facilitate various tasks, whether due to their ability to produce or calculate, the latter makes it possible to predict events. The municipality of Cataguases - Minas Gerais, in times of high rainfall, is heavily affected by floods, which cause several problems to the population, which does not have assertive information about the volume that the river will reach. Thus, this work aims to build an artificial neural network to predict the approximate volume that the Pomba River will reach during the flood season in the city of Cataguases. To this end, a neural network was developed using the Python language, which was trained through data provided by the National Water Agency. At the end of the analysis of the results of the evaluation metrics, it was noticed that the network is able to accurately predict the volume of floods.

Keywords: Artificial Neural Networks. Floods in Cataguases. Prevision.

1- Introdução

A cidade de Cataguases sofre, em períodos de chuvas, com alagamentos, deixando famílias desabrigadas e gerando um grande impacto econômico. O principal rio da cidade, rio Pomba, possui muitos afluentes e passa por diversas cidades desde sua nascente até a cidade de Cataguases, recebendo, portanto um grande volume de água decorrido das chuvas nessas localidades. Na cidade algumas residências são ribeirinhas, porém existem muitas que só são atingidas quando há um grande volume de água, um grande transbordamento.

Atualmente encontra-se para dispositivos móveis um aplicativo fornecido pela ANA – Agência Nacional de Águas -, o Hidroweb, porém ele fornece dados referentes à medição somente do volume atual do rio, também fornecidos pela Defesa Civil do município. Isso faz com que a população que é atingida somente por grandes cheias não saiba se é necessário sair de suas residências e empresas, ou não, fazendo com que na maioria dos casos, fiquem esperando, sendo tarde demais, perdendo seus bens. Um grande número de informações chega até a população por redes sociais, informando sobre volume de água em cidades a montante de Cataguases e sobre abertura das comportas das usinas, volume este que chegará e atingirá a cidade. Porém essas informações não são precisas, gerando alarde ou tranquilidade incorreta.

As redes neurais possuem grande relevância em nosso cotidiano, através de sua capacidade de serem treinadas para prever séries temporais, sendo empregadas em diversos casos para solucionar problemas (LUDWIG, 2007; MONTGOMERY, 2007).

Uma solução para o problema apresentado anteriormente seria impedir enchentes, porém isso não é possível, já que o rio e seus afluentes cortam a cidade, e trata-se de um rio com leito muito largo. Não há a possibilidade de se alterar o curso do rio, devido à falta de local e de prejudicar quem dele necessita, nem encobri-lo através de túneis subterrâneos, como ocorre em algumas cidades, tendo um exemplo Belo Horizonte, onde o volume de chuvas em Janeiro de 2020 gerou grande correnteza, fazendo com que o rio rompesse a contenção (MANSUR, 2020; RODRIGUES, 2020).

Para solucionar o problema tratado neste trabalho é necessário que a informação correta chegue à população a tempo. Atualmente não existe algo que transmita essa informação no tempo certo, sem algo que comprove, sendo muitas vezes, informações falsas ou inexatas. Como solução pode-se realizar a previsão do volume de água que chegará à bacia hidrográfica do município, com base em informações de outras cheias, treinando redes neurais, que informarão a população através de software e nas redes oficiais do município, tendo, portanto, uma previsão mais assertiva.

Este trabalho tem o objetivo de desenvolver uma rede neural, treiná-la com dados de outras cheias, e com dados do volume de água em outras cidades pertencentes à bacia hidrográfica do Rio Pomba. Pretende-se que a rede consiga prever o volume que irá atingir a cidade, sendo capaz futuramente, de avisar a população com uma maior precisão, a fim de solucionar o problema em questão.

O presente trabalho se justifica pelo fato de as enchentes causarem grandes prejuízos à população. Ao longo da história de Cataguases, diversas autoridades tentaram tratar desse problema por afetar o cotidiano e a economia do município, porém nenhuma medida aplicada, até o presente momento, teve grande impacto para solucioná-lo.

O trabalho não visa solucionar o problema das enchentes, porém visa amenizar os transtornos gerados à população através do uso da tecnologia, que com a informação correta poderá tomar as medidas necessárias no tempo certo.

O restante deste trabalho está organizado da seguinte forma: na Seção 2 é apresentado o referencial teórico do tema, abordando Redes Neurais e o Estudo Hidrológico de Cataguases, a Seção 3 apresenta a metodologia utilizada, já na Seção 4 é apresentado e discutido os resultados obtidos, e por fim na Seção 5 apresenta-se as considerações finais deste trabalho.

2- Redes Neurais e Estudo Hidrológico de Cataguases

Será apresentado neste tópico o estudo bibliográfico sobre a Hidrologia do município de Cataguases, abordando o histórico de enchentes e o processo de medição do rio, assim como a pesquisa bibliográfica sobre Redes Neurais Artificiais, apresentando conceitos, aprendizagem, otimização, vantagens e desvantagens de uma rede neural, e sobre a linguagem Python.

2-1 Hidrologia De Cataguases

O município de Cataguases situado na região da Zona da Mata no estado de Minas Gerais, apresenta uma área de cerca de 482 km², possuindo 75123 habitantes estimados segundo dados (2019) do IBGE – Instituto Brasileiro de Geografia e Estatística -, é banhado pelo Rio Pomba pertencente à Bacia do Rio Paraíba do Sul.

O Rio Pomba nasce a 1100m de altitude na Serra da Conceição, pertencente à Cadeia da Mantiqueira, território de Barbacena. Atinge a foz do Rio Paraíba do Sul após percorrer 265 km, percorrendo cerca de 35 municípios mineiros e 3 fluminenses (COPPETEC, 2006). Seus principais afluentes são Rio Novo, Piaui, Xopotó, Formoso e Pardo. De sua nascente até Cataguases o Rio Pomba recebe as águas do Rio Formoso, Rio Paraopeba, Rio Xopotó e Rio Novo, além de seus afluentes. Em Cataguases recebe as águas do Ribeirão Meia Pataca e de outros cursos d'água, como o córrego Lava-Pés, Romualdinho e Ribeirão do Cágado. Em seu percurso possui algumas barragens e Hidroelétricas, sendo as principais as usinas Maurício, Ituerê, e a Barra do Braúna, sendo está localizada após Cataguases.

2-1-1 Histórico de Enchentes

Em períodos de grande volume de chuvas na região da bacia do Paraíba do Sul, ocorrem às cheias normais do Rio Pomba, que atingem a população ribeirinha, porém em períodos de elevado nível de chuvas ocorrem às cheias excepcionais do Rio Pomba, que atingem parcelas das áreas urbanas.

Algumas cheias atingiram a cidade ao longo dos anos, sendo normais ou excepcionais, destacam-se a enchente de 1979, 2008 (considerada a pior de sua história), 2012 e em 2020 que ocorreram três enchentes excepcionais, sendo um fenômeno nunca antes visto pela população.

2-1-2 Medição nível do Rio Pomba

A ANA monitora a quantidade e a qualidade da água que entra e sai de uma região, possibilitando que planejamentos sejam realizados. Em Cataguases o processo de medição da quantidade e qualidade do rio é realizado pelo Serviço Geológico Do Brasil, depois repassado a ANA, que disponibiliza os resultados

através do portal Hidro – Telemetria e pelo aplicativo Hidroweb. Ambos disponibilizam o nível do rio nas cidades brasileiras em tempo real (ANA, 2018).

A ANA utiliza a tecnologia *Dopper* para medir a vazão dos rios (ASCON/ANA, 2011). Trata-se de uma tecnologia que utiliza sensores acústicos, que emitem ultrassons que são refletidos ao se chocarem com sólidos ou bolhas para o sensor. Quando a água está em movimento, os ultrassons são refletidos em uma frequência alternada proporcional à velocidade das partículas refletoras presentes na água. Essa velocidade é então calculada pelo equipamento (ZEBEN, 2005).

2-2 Rede Neural Artificial

Com a Revolução Industrial e o advento da tecnologia, vários estudos surgiram com base em desenvolver máquinas que pudessem substituir o homem em diversas atividades. Nos últimos tempos ganhou força os estudos com base em desenvolver técnicas que possam simular o funcionamento do cérebro humano.

Segundo SIMON (2011) uma Rede Neural se assemelha ao cérebro em dois aspectos:

- O conhecimento é adquirido através do ambiente por meio de um processo de aprendizagem.
- Para armazenar o conhecimento adquirido utilizam as forças de conexão entre os neurônios.

Uma Rede Neural Artificial (RNA) é formada por um ou uma combinação de neurônios em uma ou mais camadas que se interligam (SIMON, 2011). Uma RNA geralmente possui:

- Camada de entrada – Sua função é representar a distribuição dos valores das entradas de cada neurônio da camada seguinte;
- Camada oculta – Pode-se ter nenhuma, uma ou várias camadas ocultas em uma RNA. A função dos neurônios nessa camada é capacitar à rede e possibilitar solucionar problemas que não sejam separáveis.
- Camada de saída – É a camada que contém o resultado dos cálculos realizados pela RNA.

2-2-1 Neurônio

O cérebro humano possui cerca de 10 bilhões de neurônios, que são capazes de criar, cada um, até 10000 ligações (sinapses) com outros neurônios (LUDWIG, 2007; MONTGOMERY, 2007).

O neurônio biológico é constituído principalmente pelos componentes:

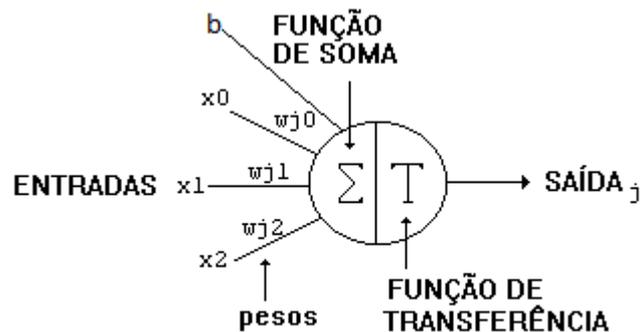
- Dendritos - responsáveis por receber estímulos transmitidos por outros neurônios;
- Soma – é o corpo do neurônio, tem a função de coletar e combinar as informações recebidas de outros neurônios;
- Axônio – tem a função de transmitir os estímulos para as outras células. Pode ter alguns metros de comprimentos, sendo constituído por uma fibra tubular.

Os sinais gerados pelos sentidos (sensores), como pela retina ocular (visão), passam pelos axônios a uma velocidade média de 25 m/s na forma de pulsos. Esses sinais só são transmitidos caso forem superiores a 50 mV (limiar do disparo) aproximadamente, caso não sejam, são bloqueados por serem considerados irrelevantes. Antes de um sinal entrar no próximo neurônio ele passa pelo processo de ligação do axônio e o dendrito, denominado sinapse. Se o sinal for superior a certo limite, denominado *threshold*, ele prossegue, caso contrário é bloqueado (LUDWIG, 2007; MONTGOMERY, 2007).

Um neurônio recebe sinais através de vários dendritos. Em seu corpo (soma) o sinal pode ser amplificado ou atenuado, dependendo de sua origem, pois a cada dendrito está associado um peso – memórias -, que multiplica o sinal. Os valores dos pesos são definidos através do treinamento recebido pelo cérebro durante sua vida (memorização).

Baseado no neurônio biológico tem-se o neurônio matemático, Figura 1, que recebe um ou mais sinais de entrada e retorna um sinal de saída, que pode ser à saída da rede ou um sinal de entrada para outro neurônio.

Figura 1 Neurônio Matemático



Fonte: Centro de Informática da Universidade Federal de Pernambuco (1999)

Cada entrada, assim como nos dendritos de um neurônio natural, possuem um peso simbolizado pela letra w (Encontra-se, também para peso a sua nomenclatura em inglês: *weight*, sendo por isso a letra w representante dos pesos em diagramas.) que as multiplica, tendo-se assim as entradas ponderadas. Esse somatório é calculado pelo neurônio através da função de soma, tendo assim a função de ativação. O valor gerado pela função de ativação é passado para a função de transferência, que dentre outras finalidades, evita que os valores de saída sejam acrescidos progressivamente, ou seja, restringe o valor da amplitude do sinal a um valor finito.

O modelo matemático de um neurônio pode incluir uma variável de polarização, o bias, representada pela letra b . O valor dessa variável é incluído ao somatório da função de ativação, com a finalidade de aumentar a capacidade de aproximação da rede, possibilitando que um neurônio apresente saída não nula ainda que todas as entradas sejam nulas, conforme explicado por LUDWIG (2007); MONTGOMERY (2007). O valor do bias também pode ser ajustado, assim como os pesos, no processo de aprendizagem de uma rede.

2-2-2 Aprendizagem de uma RNA

Uma rede neural aprende sobre seu ambiente através de treinamento, que é o processo de ajustar os valores dos pesos e do bias.

O processo de aprendizagem de uma RNA pode ser de duas formas de acordo com LUDWIG (2007); MONTGOMERY (2007):

- Aprendizado Supervisionado – é quando a rede é treinada com conjunto de valores de entrada e saída.
- Aprendizado Não-Supervisionado - o treinamento é realizado de forma que a rede não recebe informações sobre as saídas desejadas, sendo usados somente valores de entrada.

2-2-3 Percéptron de Múltiplas Camadas

O Percéptron de múltiplas camadas (*Multilayer Perceptron* – MLP) é uma rede que possui no mínimo uma camada interna, ou seja, possui no mínimo uma camada oculta, sendo, portanto, uma rede com uma camada sensorial ou camada de entrada, uma ou mais camadas oculta e uma camada de saída. O sinal de entrada se propaga para frente através das camadas até a camada de saída, ou seja, é uma rede alimentada para frente (LUDWIG, 2007; MONTGOMERY, 2007).

Os Percéptron de múltiplas camadas tem tido um notório sucesso ao resolver problemas complexos por meio do algoritmo de treinamento supervisionado retropropagação de erro (*error back-propagation*) ou regra delta generalizada. A retropropagação de erro consiste em duas etapas através das camadas da rede:

- *Forward* – um passo para frente, a propagação. Um vetor de entrada é aplicado aos nós da rede e seu efeito se propaga através da rede. Um conjunto de saídas é produzido como a resposta real da rede, para cada neurônio tem-se um desvio (erros) entre as saídas desejadas e as obtidas. Durante essa etapa os pesos sinápticos permanecem inalterados.
- *Backward* – um passo para trás, a retropropagação. Produz-se um sinal de erro através da diferença entre a resposta obtida e a desejada, onde esse sinal é propagado para trás através da rede, contra a direção das ligações sinápticas. Os pesos são ajustados por todos os neurônios a cada iteração de acordo com uma regra de correção de erro, com a finalidade de fazer com que a resposta real da rede se aproxime mais da resposta desejada.

Segundo SIMON (2011) um pércéptron de múltiplas camadas tem três características distintas:

- O modelo de cada neurônio da rede possui uma função de ativação não-linear, sendo a não-linearidade suave. A presença da não-linearidade é importante para que a relação entrada-saída não seja reduzida a àquela de um perceptron de camada única.
- A rede possui um ou mais camadas de neurônios ocultos, que capacitam a rede a aprender tarefas complexas extraíndo de forma progressiva as características mais significativas dos padrões de entrada.
- A rede demonstra um alto grau de conectividade, determinado pelas sinapses da rede. Qualquer mudança em alguma dessas conectividades da rede exige uma mudança na população das conexões sinápticas ou de seus pesos.

2-2-4 Método de Otimização

Através de uma função objetivo – função que busca a maximização ou minimização de um determinado objetivo. – técnicas buscam otimizar a solução de um problema avaliado, essas técnicas são chamadas de métodos de otimização.

Os métodos de otimização podem ser classificados em métodos exatos e aproximados. Os métodos exatos são capazes de encontrar a solução ótima de um problema, porém existem problemas mais complexos onde obter suas soluções ótimas torna-se uma tarefa mais difícil, desse modo utiliza-se os métodos aproximados, que são capazes de encontrar soluções próximas a ótimas (FURTADO, 2020).

Alguns métodos de otimização aproximados (heurísticos), podem ser inspirados em fenômenos naturais, como o voo dos pássaros, a evolução das espécies, dentre outros. Cita-se um algoritmo de otimização baseado em enxame de partículas (PSO, *Particle Swarm Optimization*), que é inspirado no voo dos pássaros e insetos, e no movimento dos peixes. Foi desenvolvido em 1995 pelo psicólogo James Kennedy e pelo engenheiro Russel Eberhart.

O PSO é um algoritmo estocástico - aqueles ao qual não se tem conhecimento perfeito sobre os acontecimentos que estão por vir – e funciona da seguinte forma: cada indivíduo (partícula) do conjunto se desloca sobre o espaço de busca com uma velocidade que é ajustada de acordo com suas experiências pessoais e do conjunto (população). Os indivíduos da população possuem uma

memória que registra a melhor posição do indivíduo (*pBest*), e de uma memória coletiva que registra a melhor posição que a população de indivíduos já atingiu (*gBest*). Essas características permitem que os indivíduos possam estimar qual a melhor posição dentro do espaço de busca, visando alcançar os pontos que representam as melhores soluções para o problema. Essa capacidade de memória dos indivíduos determina o processo de aprendizado do algoritmo, permitindo alterar a velocidade e a direção. De acordo com a função objetivo, que é estabelecida conforme o problema a ser solucionado, é definido se a posição do indivíduo se aproxima mais da ótima. O espaço de busca é delimitado de acordo com valores máximos e mínimos definidos segundo o problema a ser solucionado (FURTADO, 2020).

Pode-se utilizar o PSO para otimizar os parâmetros de um rede neural MLP, como o número de camadas ocultas e a função de ativação. Como critério de parada, o algoritmo pode utilizar o número máximo de interações ou um valor de tolerância para a variação do valor da função objetivo.

2-2-5 Vantagens e desvantagens de uma Rede Neural Artificial

Diferente de outros algoritmos que processam apenas o que lhes foi ensinado, as redes neurais são capazes de produzir saídas para entradas que não estavam em seu treinamento, através de uma habilidade denominada generalização. Outra vantagem tem-se o fato de que as redes possam modificar seus pesos em tempo real, possibilitando solucionar problemas onde as estatísticas mudam com o tempo.

Como desvantagem cita-se o fato de as redes não conseguirem fornecer soluções para problemas mais complexos. Uma solução é decompor o problema em tarefas mais simples. Outra desvantagem é o fato de ser impossível saber o porquê uma rede chegou a determinado resultado ou, para redes de múltiplas camadas, qual a relevância de um peso para o resultado, ou seja, não é possível obter uma prova formal para os resultados obtidos.

2-3 Python

Python é uma linguagem de programação de alto nível, multiparadigma, interpretada, robusta, multiplataforma e open source.

A linguagem Python foi criada no final da década de 80 e implantada no início da década de 90 pelo programador e pesquisador do Instituto Nacional de Matemática e Ciência da Computação da Holanda Guido Van Rossum, com a finalidade de aprimorar um software de programação denominado ABC. (LIMA, 2019). Atualmente, a linguagem faz parte de um modelo de desenvolvimento comunitário gerenciado pela organização sem fins lucrativos Python Software Foundation.

A linguagem tem ganhado crescente notoriedade nos últimos anos, devido a sua curva de aprendizado baixa, tornando-se uma ferramenta ágil para o desenvolvimento de software, destacando-se suas funcionalidades para o desenvolvimento de big data, *machine learning* e inteligência artificial.

3- Metodologia

Para a parte bibliográfica foram utilizados livros digitais encontrados na biblioteca Digital da instituição de ensino, e um livro físico que foi a fundamentação do estudo teórico. Utilizou-se também de artigos encontrados através da ferramenta Google acadêmico - Ferramenta do Google que possibilita pesquisar artigos acadêmicos e editoriais universitários. -, dados obtidos através do site da ANA, e sites pesquisados no Google.

Para o desenvolvimento da rede, inicialmente, faz-se necessário montar uma base de dados para o treinamento da rede. Obteve-se os dados através de uma base fornecida pela ANA, onde utilizou-se dados referente a cotas (volume do leito do rio) de cinco estações. As estações foram escolhidas devido à proximidade com Cataguases, destacando que pertencem ao leito do Rio Pomba e são anteriores ao município.

Ressalta-se que durante a montagem da base de dados, utilizada para o treinamento e validação da rede, ocorreram algumas dificuldades, como o fato de não conseguir encontrar dados referentes a algumas estações pertencentes ao leito do Rio Pomba e de seus afluentes na base de dados fornecida pela ANA, tendo como exemplo a estação pertencente ao leito do Rio Paraopeba, que não está presente nessa base de dados. Outra dificuldade encontrada cita-se o fato de que algumas estações não possuem dados referentes a algumas datas, onde se destaca

a estação de Guarani que possui apenas dados referentes ao período de 2008 ao primeiro semestre de 2010, e o do ano de 2012.

Utilizando o aplicativo Hidroweb separaram-se as estações que seriam utilizadas no mapa e utilizando o código de cada estação escolhida, fornecido no aplicativo, realizou-se a filtragem na base de dados da ANA. Os dados disponibilizados pela ANA encontra-se em um arquivo .mdb que, neste trabalho, os organizou-se em planilhas .xls, uma para cada estação. Após, organizou-se os dados referentes a cinco dias em cada estação anterior a Cataguases e para o município separou-se um dia após os cinco das demais estações, contemplando um período de seis dias de observação, em uma tabela, que após converteu-se em um arquivo .csv utilizado para o treinamento e validação da rede.

Foi desenvolvida uma rede neural artificial de múltiplas camadas ou Perceptrons Multicamadas (MLPs). Neste trabalho utilizaram-se redes que possuem até três camadas internas. A topologia utilizada foi a *feed-forward* – redes neurais em que as sinapses não formam um ciclo. As informações movimentam-se para frente na rede, inicialmente pelos nós da camada de entrada, depois pelos nós das camadas ocultas e, após, pelos nós da camada de saída. –, unidirecional simples. Dentre os vários algoritmos para treinamento de redes neurais existentes, nesse trabalho utilizou-se o L-BFGS (*Broyden-Fletcher-Goldfarb-Shanno*), que se trata de um algoritmo de otimização capaz de solucionar problemas com um grande número de variáveis controlando a quantidade de memória utilizada.

Como função de ativação da rede foram utilizadas as funções a seguir:

- Linear (Identidade)

$$\phi(v) = v \quad (3.1)$$

- Linear retificada (ReLU)

$$\varphi(v) = \max(0, x) \quad (3.2)$$

- Logística

$$\varphi(v) = \frac{1}{1 + e^{-v}} \quad (3.3)$$

- Tangente Hiperbólica (Tanh)

$$\varphi(v) = \tanh(v) \quad (3.4)$$

O PSO (*Particle Swarm Optimization*) foi o método de otimização utilizado para esta rede, ou seja, para obter-se os parâmetros da rede e utilizou-se o número máximo de interações como critério de parada.

Utilizou-se a técnica de validação cruzada – que é uma técnica utilizada para avaliar o desempenho de estimadores, a generalização da rede, utilizando a totalidade dos dados do conjunto. – k-fold, que consiste em dividir a base de dados em k conjuntos de tamanhos iguais, e realizar o ajuste do modelo em k-1 conjuntos de dados e a validação no conjunto restante. Esse processo ocorre k vezes, e o modelo é validado para cada parte do conjunto de dados. Geralmente escolhe-se o valor de k entre 5 e 10, onde a escolha é realizada dependendo do tamanho do conjunto de amostras.

O método de validação cruzada é utilizado para minimizar os erros de previsão causados por *overfitting* (sobre-ajuste) – situação onde o modelo aprende sinais e ruídos nos dados de treinamento, fato que prejudica seu desempenho diante de dados não utilizados em seu treinamento. (LUCIAN, 2020) – e é indicado para a validação de bases de dados que possuem um número reduzido de dados amostrais.

Para avaliar a qualidade da previsão da rede desenvolvida nesse trabalho foram usadas métricas de avaliação estatística baseadas no erro de previsão. As métricas podem ser definidas de acordo com o problema a ser avaliado. Neste trabalho utilizaram-se três métricas, escolhidas devido a sua fácil interpretação e difusão na literatura.

Estabelece-se \hat{y} como a saída estimada, y como rótulo das amostras, \bar{y} como a média dos rótulos das amostras e N o número de amostras, tem-se a definição das métricas utilizadas:

- Coeficiente de determinação (R^2) – trata-se de uma medida estatística de quão próximo os dados estão da linha de regressão ajustada e todos os pontos de dados. Sobretudo, esse coeficiente indica o quanto o modelo foi capaz de explicar os dados coletados. É definido pela equação:

$$R^2(y, \hat{y}) = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (3.5)$$

Onde (R^2) varia de 0 a 1, sendo que quanto mais próximo de 1 melhor é a qualidade de generalização do modelo.

- Raiz do erro médio quadrático (RMSE, *Root Mean Squared Error*) – é a distância quadrática média da linha de regressão, ou seja, é a alteração em redor da linha de regressão. É comumente usada para expressar a exatidão dos resultados. O RMSE é definido na mesma unidade que a saída do método, facilitando sua interpretação. A métrica pode ser definida como sendo o desvio padrão entre os valores previstos e amostrais. É obtido através da equação:

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.6)$$

- Erro percentual absoluto médio (MAPE, *Mean Percentual Error*) – Seu valor é dado em porcentagem, facilitando a comparação da qualidade da previsão do método. É definido através da equação:

$$\text{MAPE}(y, \hat{y}) = \frac{1}{N} \left(\sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|} \right) 100 \quad (3.7)$$

3-1 Código computacional

O desenvolvimento do código computacional deste trabalho ocorreu na linguagem Python. Para o desenvolvimento da rede nessa linguagem faz-se necessário a utilização da biblioteca “*sklearn.neural_network*” (*scikit-learn*), um classificador Multi-layer Perceptron. E para a preparação e análise dos dados foi utilizada a biblioteca Pandas, uma biblioteca especializada para tais funções, que oferece estrutura de dados e operações para manipular dados numéricos em formato tabular e séries temporais (LOPES, 2019).

Inicialmente a rotina cria uma pasta onde será armazenado um arquivo .csv contendo o resultado para cada rodada da rede. Após, é feita a leitura da base de dados e a preparação da mesma para ser utilizada pela rede. Para a normalização dos dados, uma parte importante da preparação devido ao fato de permitir obter

melhores resultados, foi estabelecido um limite mínimo e máximo para os dados, respectivamente 0,15 e 0,85. Depois da escolha desses limites, foi realizada a normalização de acordo com as equações:

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (3.8)$$

$$Y_n = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (3.9)$$

Onde X_{\min} e Y_{\min} são os valores mínimos, X_{\max} e Y_{\max} são os valores máximos e X_n e Y_n são os valores normalizados.

Após a normalização dos dados, inicia-se a etapa de otimização dos parâmetros da técnica de aprendizagem. Essa otimização é realizada pelo PSO com o auxílio da técnica de validação cruzada. O PSO verifica qual o melhor valor quanto à taxa de aprendizado, a função de ativação (Podendo ser Linear, Linear Retificada, Logística e Tangente Hiperbólica), o número de camadas ocultas (Sendo de 1 até 3 camadas) e o número de neurônios em cada camada (Podendo ser de 50 a 200 neurônios). Todo o processo descrito nesse parágrafo possui uma parcela de aleatoriedade, para tanto são realizadas 30 execuções do mesmo para garantir uma amostra significativa de resultados.

Calculadas as métricas, explicadas anteriormente, para os resultados das execuções do algoritmo, para cada uma das 30 execuções, são calculados, os valores de média e desvio padrão de cada métrica, o que possibilitou a análise da capacidade de previsão da rede.

4 Resultados e Discussões

Após executar o método 30 vezes, obteve-se como a configuração da rede para a melhor rodada os seguintes parâmetros: taxa de aprendizado: 0,345; número de camadas: 2; número de neurônios em cada camada: 55 para a 1ª e 124 para a 2ª; e função de ativação: logística.

Na Tabela 1 são apresentados os resultados das métricas através da média e do desvio padrão (valores entre parênteses) para as trinta execuções realizadas.

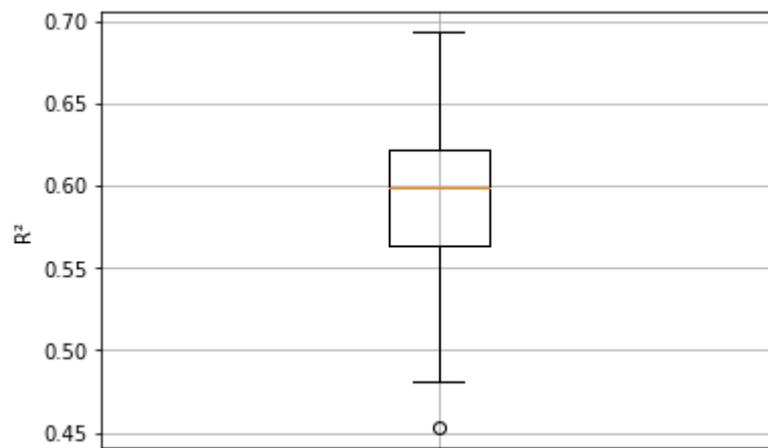
Tabela 1 Resultado Métricas de Avaliação

Métrica	MAPE (%)	RMSE (cm)	R ²
	20,25 (1,99)	66,09 (4,35)	0,59 (0,05)

Fonte: Elaborada pelo autor (2020)

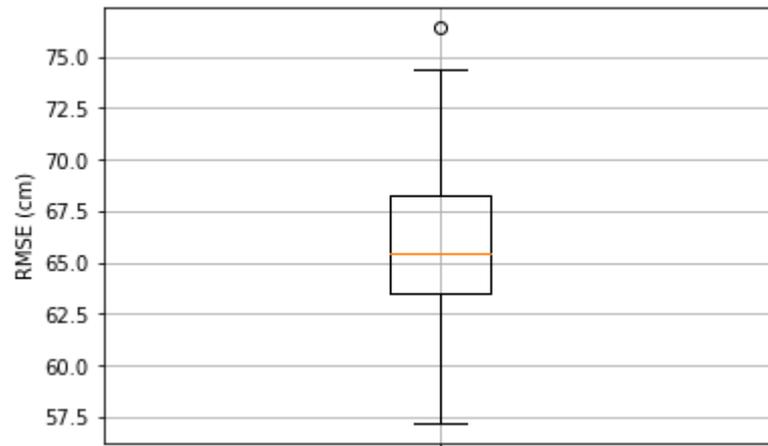
Pelos valores das métricas percebe-se que a variância em relação ao previsto pela rede e os dados reais é pequena, demonstrando que o método utilizado é preciso.

Para auxiliar a análise dessa variância pode-se observar os Gráficos 1 a 3 que demonstram gráficos de *boxplot* dos resultados das métricas da melhor rodada.

Gráfico 1- Boxplot R²

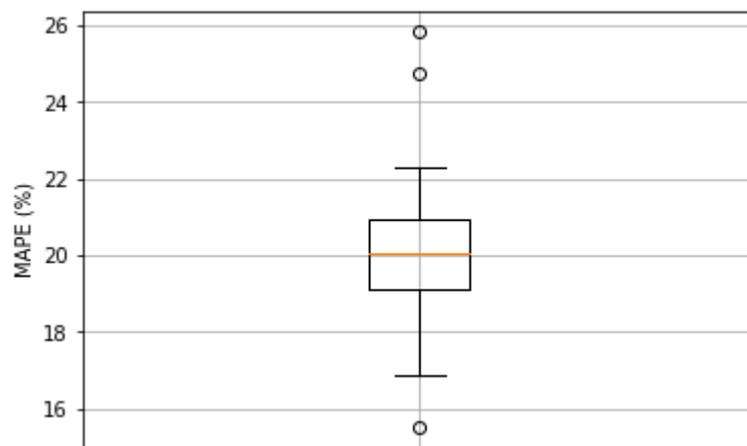
Fonte: Elaborado pelo autor (2020)

No Gráfico 1 observa-se que o erro médio quadrático é 0,60, e a variação dos resultados é aproximadamente 0,05 para baixo e para cima.

Gráfico 2 - Boxplot RMSE

Fonte: Elaborado pelo autor (2020)

No Gráfico 2 observa-se que o erro médio é de aproximadamente 65cm, e a variação dos resultados é de aproximadamente 5 cm para baixo e para cima.

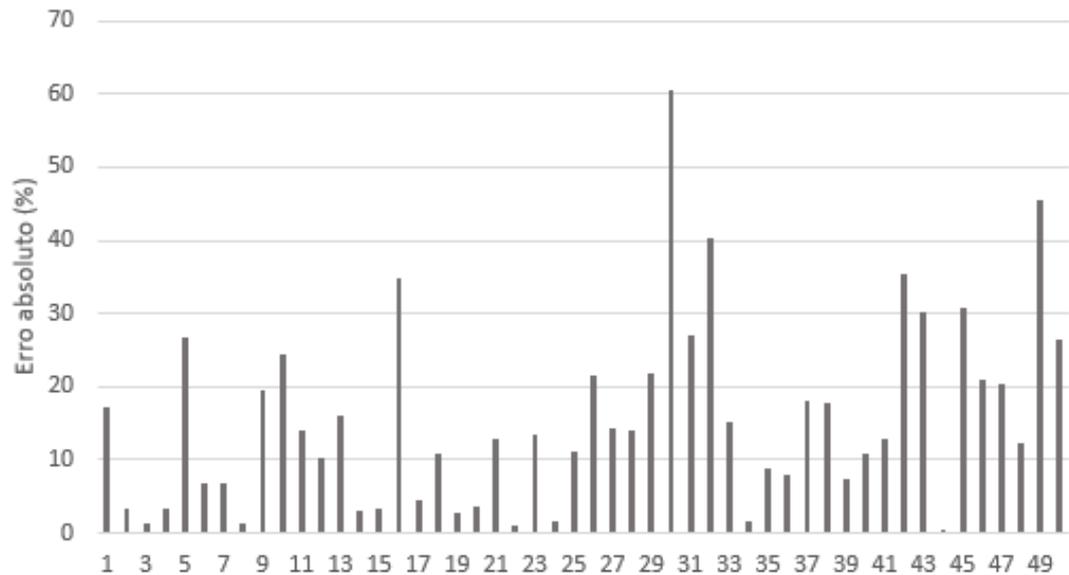
Gráfico 3 - Boxplot MAPE

Fonte: Elaborado pelo autor (2020)

No Gráfico 3 observa-se que o erro médio é de 20% e que a variação dos resultados é de aproximadamente 2% para cima e 3% para baixo.

O Gráfico 4 apresenta o erro absoluto porcentual de aprendizagem da rede na melhor rodada, sendo que devido a parcela de aleatoriedade foram realizadas 30 execuções. O erro porcentual é dado pelo coeficiente entre a diferença entre os valores da base de dados (valores desejados) e os valores previstos pela rede (valores obtidos) e os valores desejados.

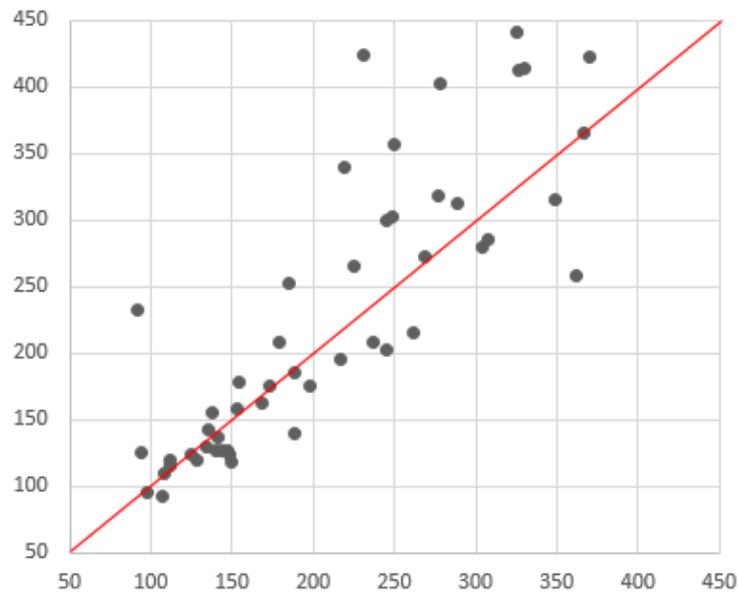
Gráfico 4 Erro Absoluto Porcentual



Fonte: Elaborado pelo autor (2020)

O erro médio absoluto percentual está entre 15% a 16%. Observa-se que para valores menores de cotas o percentual de erro foi menor, tendo picos de diferença para valores maiores. Isso se dá pelo fato de que a base de dados fornecida pela ANA é escassa e, também, por períodos de cheias serem mais esporádicos, tendo valores de cotas menores na maior parte do ano.

O gráfico 5 demonstra a diferença entre os valores preditos do real, onde também pode-se observar que para valores maiores a diferença foi maior em relação à valores menores. Desse modo, implica-se que para que a rede possa prever melhor para maiores volumes de cotas faz-se necessário treina-la com um número maior de dados.

Gráfico 5 Valores Preditos x Valores Reais

Fonte: Elaborado pelo autor (2020)

Destaca-se que para obter uma previsão mais assertiva seria necessário ter um número maior de dados com grandes valores de cotas, além de que a uma base de dados maior. A base de dados fornecida pela ANA utilizada para treinamento e validação da rede era escassa, sendo que muitas das estações avaliadas não possuíam muitos dados. Destaca-se também, que foram utilizados somente dados de estações pertencentes ao leito do rio Pomba, pois não foram encontrados dados referentes a cotas do leito de afluentes, como por exemplo, um dos que mais influenciam no volume de cheias do município, o Rio Xopotó. Desse modo, faz-se necessário para obter-se melhores resultados obter dados das cotas desses afluentes.

5 Considerações Finais

O presente trabalho buscou investigar o uso de redes neurais artificiais na previsão do volume de cheias no município de Cataguases. Foi desenvolvido uma rede neural artificial percéptron múltiplas camadas treinada através do algoritmo L-BFGS, e otimizada pelo PSO com o propósito de otimizar o desempenho dos métodos de aprendizado de máquina. A rede foi treinada e validada pelo método de validação cruzada K-fold. O processo de treinamento, validação dos resultados e

cálculo das métricas de avaliação foram executados trinta vezes visando obter uma melhor análise estatística dos resultados.

O método demonstra ser preciso, apresentando resultados médios de 20,25% no MAPE, 66,09 cm no RMSE e 0,59 no R^2 . As análises dos gráficos de boxplot demonstrou que obteve-se bons resultados de desvio padrão. A variância dos resultados pode ser verificada também pelos gráficos de dispersão, onde percebe-se que a variação dos resultados é menor para valores menores de cotas do que para valores maiores. Essa variação de resultados deve-se a escassez de dados na base fornecida pela ANA.

Dessa forma, percebe-se que o método utilizado é eficaz na resolução do problema tratado nesse trabalho, sendo uma rede precisa na predição dos dados desejados. Para um melhor resultado faz-se necessário utilizar um maior número de dados com valores altos de cotas e dados referentes a cotas de estações além das pertencentes ao leito do rio Pomba, a de seus afluentes.

Referências

AGÊNCIA NACIONAL DE ÁGUAS – ANA. *Hidroweb v3.1.1 – Rede Hidrometeorológica Nacional*. Disponível em: < <http://www.snirh.gov.br/hidroweb/apresentacao>>. Acesso em 28 de maio de 2020.

ASCOM/ANA. *Agência explica como é feito monitoramento de rios. 15 de março de 2019*. Disponível em: < <https://www.ana.gov.br/noticias-antigas/agaancia-explica-como-a-c-feito-monitoramento-de.2019-03-15.3285844819>>. Acesso em 28 de maio de 2020.

FUNDAÇÃO COPPETEC LABORATÓRIO DE HIDROLOGIA E ESTUDOS DE MEIO AMBIENTE. *Plano de Recursos Hídricos da Bacia do Rio Paraíba do Sul – Caderno de Ações Bacia do Rio Pomba*. Relatório Contratual R-10. 2 p. Disponível em: < <http://www.ceivap.org.br/downloads/cadernos/Caderno%205%20-%20Pomba.pdf>>. Acesso em 28 de maio de 2020.

FURTADO, GERALDO. *Metamodelo com arquitetura em duas camadas para a previsão de propriedades mecânicas de misturas de concreto*. 2020.

LIMA, GUSTAVO. *O que é Python e pra que serve?*. Disponível em: < <https://www.voitto.com.br/blog/artigo/python>>. Acesso em 28 de novembro de 2020.

LOPES, ALEXANDRE. *O Que é Pandas e Seu Funcionamento Para Análise de Dados*. Disponível em: < <https://aprendizadodemaquina.com/artigos/o-que-e-pandas/>>. Acesso em 29 de novembro de 2020.

LUCIAN, BRUNO. *Overfitting: o que é e como evitar*. Disponível em: <<https://www.dadosaleatorios.com.br/post/overfitting/>>. Acesso em 28 de novembro de 2020.

LUDWIG, OSWALDO; MONTGOMERY, Eduard. *Redes Neurais – Fundamentos e Aplicações com Programas em C*. Rio de Janeiro: Editora Ciência Moderna, 2007. 9788573936193.

MANSUR, RAFAELA. RODRIGUES, GABRIEL. *Estrutura que cobre o Arrudas na Andradas pode romper, alertam especialistas*. Disponível em: < <https://www.otempo.com.br/cidades/estrutura-que-cobre-o-arrudas-na-andradas-pode-romper-alertam-especialistas-1.2289431#:~:text=A%20cena%20do%20rio%20Arrudas,arrancadas%20pela%20for%C3%A7a%20da%20%C3%A1gua.>>. Acesso em 01 de dezembro de 2020.

SERVIÇO GEOLÓGICO DO BRASIL – CPRM. *Programa Recursos Hídricos Superficiais*. Disponível em: < <http://www.cprm.gov.br/publique/Hidrologia/Apresentacao-34>>. Acesso em 28 de maio de 2020.

SIMON, H. *Redes Neurais*. Porto Alegre: Bokman ; Grupo A, 01/2011. 9788577800865. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788577800865/>>. Acesso em: 27 de maio de 2020.

ZEBEN. *Como funciona a tecnologia Doppler?*. Disponível em: < <https://zeben.pt/pt/como-funciona-a-tecnologia-doppler>>. Acesso em 28 de maio de 2020.