



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL

José Pedro Lima Alexandre

Inteligência artificial para modelar o processo chuva vs. vazão natural afluyente ao reservatório de Sobradinho

São Cristóvão – SE

2023

José Pedro Lima Alexandre

Inteligência artificial para modelar o processo chuva vs. vazão natural afluente ao reservatório de Sobradinho

Trabalho de Conclusão de Curso apresentado ao
Departamento de Engenharia Civil da Univer-
sidade Federal de Sergipe (UFS) como requisito
para o título de Bacharel em Engenharia Civil

Universidade Federal de Sergipe (UFS)
Centro de Ciências Exatas e Tecnologia (CCET)
Departamento de Engenharia Civil (DEC)

Orientador: Prof. Dr. Alcigeimes Batista Celeste

São Cristóvão – SE
2023

José Pedro Lima Alexandre.

Inteligência artificial para modelar o processo chuva vs. vazão natural afluente ao reservatório de Sobradinho/ José Pedro Lima Alexandre, São Cristóvão – SE, 2023. 59 p. : il.

Trabalho de Conclusão de Curso. Universidade Federal de Sergipe, São Cristóvão – SE.

I. Universidade Federal de Sergipe. II. Título.



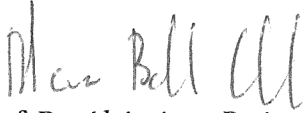
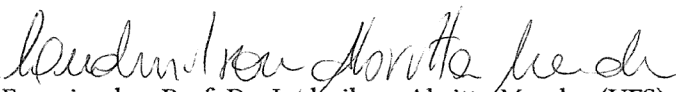
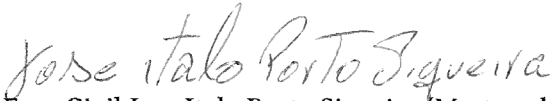
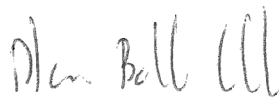
ATA DE DEFESA

José Pedro Lima Alexandre

Inteligência artificial para modelar o processo chuva vs. vazão natural afluente ao reservatório de Sobradinho

Trabalho de Conclusão de Curso apresentado ao Departamento de Engenharia Civil da Universidade Federal de Sergipe (UFS) como requisito para o título de Bacharel em Engenharia Civil.

Aprovada em: 29 de maio de 2023

Banca Examinadora	Nota
 Orientador: Prof. Dr. Alcigeimes Batista Celeste (UFS)	<u>8,8</u>
 Examinador: Prof. Dr. Ludmilson Abritta Mendes (UFS)	<u>8,5</u>
 Examinador: Eng. Civil Jose Italo Porto Siqueira (Mestrando PROEC/UFS)	<u>9,0</u>
	Média Final: <u>8,8</u>
 Prof. Dr. Alcigeimes Batista Celeste (UFS) Orientador	

*Este trabalho é todo dedicado aos meus pais,
pois é graças aos esforços deles que hoje posso concluir o meu curso.*

Agradecimentos

Agradeço primeiramente a Deus, por me guiar até aqui, me dando forças para conquistar o meu sonho.

Aos meus pais, Josefina e Pedrito, por todo apoio, carinho e esforços.

A minha irmã, Juiliana, por toda força, ajuda e fraternidade.

Aos meus avós, Maria Helena, Josefa (*in memoriam*) e João, por todo carinho e apoio que sempre tiveram juntos comigo.

Aos meus familiares, agradeço a todos que sempre estiveram me apoiando.

Aos amigos e colegas que fiz durante a graduação, vocês foram fundamentais na minha formação, com momentos de muitos estudos, desesperos e de alegria.

Ainda assim, ao meu orientador, Prof. Alcigeimes, por ter me aceitado e por todo aprendizado, pela dedicação e disponibilidade.

*"Tudo o que um sonho precisa para ser realizado
é alguém que acredite que ele possa ser
realizado."*

(Roberto Shinyashiki)

Resumo

O Brasil possui uma grande quantidade de água disponível. Porém, parte da região do Nordeste sofre com a falta do recurso. A regularização do regime de vazões do rio São Francisco com a construção do reservatório de Sobradinho contribuiu para a diminuição das abundantes cheias nas regiões a jusante. Este trabalho apresenta a aplicação de redes neurais artificiais (RNA) e do algoritmo dos k -vizinhos mais próximos (KNN) para modelagem do processo chuva vs. vazão natural afluente ao reservatório de Sobradinho. Os dados de precipitação foram coletados do portal HidroWeb e os dados de afluências naturais do portal da Câmara de Comercialização de Energia Elétrica (CCEE). Os conjuntos de dados foram divididos entre calibração e validação, em 70% e 30%, respectivamente, selecionados de maneira randômica. As simulações foram realizadas com o software de aprendizado de máquina Weka, e testadas quatro formulações para análise mensal. Por fim, os resultados são mostrados através da eficiência das simulações realizadas, verificadas por meio do coeficiente Nash-Sutcliffe. O objetivo inicial era modelar o processo vazão-vazão em Sobradinho de forma a prever a vazão total do próximo ano baseada nas vazões de anos passados. Foram feitas diversas modelagens, com várias configurações de atributos, porém, os resultados foram todos insatisfatórios para as quatro formulações da análise anual com vazões. Por essa razão, decidiu-se utilizar a análise mensal, com os dados de chuva e vazão. Assim, foram obtidos resultados bons e muito bons para as quatro formulações, em ambos os modelos trabalhados, RNA e KNN. Na modelagem, a formulação com atributos chuva vs. vazão de até três períodos anteriores apresentou os melhores resultados para a KNN e a formulação com apenas atributos de precipitação apresentou os melhores resultados para o RNA, com os índices de eficiência e classificação de muito bom.

Palavras-chaves: Modelagem chuva-vazão. Modelagem vazão-vazão. Redes neurais artificiais. Algoritmo dos k -vizinhos mais próximos. Weka. Vazões naturais.

Abstract

Brazil has a large amount of available water. However, part of the Northeast region suffers from a lack of this resource. The regularization of the flow regime of the São Francisco river with the construction of the Sobradinho reservoir contributed to the reduction of the abundant floods in the downstream region. This work presents the application of artificial neural networks (ANN) and the k-nearest neighbors algorithm (KNN) for modeling the rainfall-runoff process considering the natural flow to the Sobradinho reservoir. Precipitation data were collected from the HidroWeb portal and natural inflow data from the Câmara de Comercialização de Energia Elétrica (CCEE) portal. The set of data were divided into calibration (70%) and validation (30%), at random. Simulations were performed by using the Weka machine learning software, and four formulations were tested for monthly analysis. The goodness of fit of the results are shown by means of the Nash-Sutcliffe coefficient. The initial objective was to model the runoff-runoff process in Sobradinho in order to predict the total inflow of the next year based on the flows of past years. Several models were tested, with several configurations of attributes, however, the results were all unsatisfactory for the four formulations of the annual analysis with flows. For this reason, it was decided to use the monthly analysis, with rainfall and flow data. Thus, good and very good results were obtained for the four formulations, in both models investigated, ANN and KNN. In modeling, the formulation with rainfall-runoff attributes from three previous periods showed the best results for ANN and the formulation with only precipitation attributes showed the best results for KNN, with efficiency indices and classification of very good.

Keywords: Rainfall-runoff modeling. Runoff-runoff modeling. Artificial neural networks. *k*-nearest neighbors algorithm. Weka. Natural flows.

Lista de ilustrações

Figura 1	– Representação simplificada de uma rede neural artificial	18
Figura 2	– Detalhes de um neurônio artificial na camada oculta	19
Figura 3	– Menu de acesso	21
Figura 4	– Seleção de séries históricas de chuva para a sub-bacia 43	23
Figura 5	– Estações Convencionais de chuva da Sub-bacia 43	23
Figura 6	– Importação de Dados	24
Figura 7	– Seleção para importação de arquivo baixado no HidroWeb	25
Figura 8	– Coleta de dados	25
Figura 9	– Totais mensais da série da estação Fazenda Samburá - Posto 02046025 . . .	26
Figura 10	– Configuração de um arquivo ARFF - <i>allData</i> da Formulação 1 da análise mensal	29
Figura 11	– Seleção da aba <i>Preprocess</i>	30
Figura 12	– Localização de DDM da RNA	31
Figura 13	– Exemplo de formulação 1 da visualização de quantidade de neurônios e camadas para o RNA, para análise mensal	31
Figura 14	– Localização de DDM do KNN	32
Figura 15	– Calibração da RNA	33
Figura 16	– Exemplo de resultados obtidos para RNA	34
Figura 17	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação de RNA (análise anual)	36
Figura 18	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação de RNA (análise anual)	37
Figura 19	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação de RNA (análise anual)	37
Figura 20	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação de RNA (análise anual)	38
Figura 21	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 5 com aplicação de RNA (análise anual)	38
Figura 22	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação do KNN (análise anual)	39
Figura 23	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação do KNN (análise anual)	39
Figura 24	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação do KNN (análise anual)	40
Figura 25	– Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação do KNN (análise anual)	40

Figura 26 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 5 com aplicação do KNN (análise anual)	41
Figura 27 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação de RNA (análise mensal)	41
Figura 28 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação de RNA (análise mensal)	42
Figura 29 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação de RNA (análise mensal)	42
Figura 30 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação de RNA (análise mensal)	43
Figura 31 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação do KNN (análise mensal)	43
Figura 32 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação do KNN (análise mensal)	44
Figura 33 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação do KNN (análise mensal)	44
Figura 34 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação do KNN (análise mensal)	44
Figura 35 – Melhores índices de desempenho nas Formulações (análise anual)	45
Figura 36 – Melhores índices de desempenho nas Formulações (análise mensal)	46

Lista de tabelas

Tabela 1	– Quantidade de postos pluviométricos nas sub-bacias a montante do barramento de Sobradinho	26
Tabela 2	– Períodos para os quais a chuva média foi obtida por sub-bacia	27
Tabela 3	– Formulações para análise de vazões anuais via DDM	27
Tabela 4	– Formulações para análise mensal via DDM	27
Tabela 5	– Configurações das camadas ocultas na RNA para as cinco formulações da análise anual	30
Tabela 6	– Configurações das camadas ocultas na RNA para a formulação 1 para análise mensal	30
Tabela 7	– Configurações das camadas ocultas na RNA para a formulação 2 para análise mensal	32
Tabela 8	– Configurações das camadas ocultas na RNA para a formulação 3 para análise mensal	32
Tabela 9	– Configurações das camadas ocultas na RNA para a formulação 4 para análise mensal	32
Tabela 10	– Configurações da quantidade de k -vizinhos aplicados no KNN para análise anual	33
Tabela 11	– Configurações da quantidade de k -vizinhos aplicados no KNN para análise mensal	33

Lista de abreviaturas e siglas

DDM	Modelo guiado por dados
ANA	Agência Nacional de Águas e Saneamento Básico
IBL	Aprendizado baseado em instâncias
MLP	<i>Multilayer Perceptron</i>
RNA	Rede neural artificial
KNN	Algoritmo dos k -vizinhos mais próximos

Sumário

1	INTRODUÇÃO	15
1.1	Considerações Iniciais	15
1.2	Objetivos	16
1.2.1	Objetivo Geral	16
1.2.2	Objetivos Específicos	16
2	REVISÃO BIBLIOGRÁFICA	17
2.1	Modelagem Chuva-Vazão	17
2.2	Redes Neurais Artificiais	18
2.3	Algoritmo dos <i>k</i>-vizinhos Mais Próximos	19
2.4	Eficiência de Nash-Sutcliffe	20
2.5	Weka	21
3	MATERIAL E MÉTODOS	22
3.1	Área de Estudo	22
3.2	Dados de Precipitação e Vazão	22
3.2.1	Aquisição de Dados	22
3.2.2	Formulações para Análise DDM	27
3.2.3	Classificação dos Dados das DDMs	28
3.3	Aplicação do Weka	28
3.3.1	Arquivo em ARFF	28
3.3.2	Escolha do Método e Configuração	28
3.3.2.1	Configuração da RNA	28
3.3.2.2	Configuração do KNN	32
3.3.3	Execução dos Modelos	33
3.4	Verificação do Desempenho da Modelagem	34
4	RESULTADOS E DISCUSSÃO	36
4.1	Resultados das Modelagens	36
4.1.1	Análise Anual	36
4.1.1.1	Resultados da RNA	36
4.1.1.2	Resultados do KNN	38
4.1.2	Análise Mensal	39
4.1.2.1	Resultados da RNA	39
4.1.2.2	Resultados do KNN	40
4.2	Discussão dos Resultados	43

4.2.1	Análise Anual	43
4.2.2	Análise Mensal	45
	Conclusão	47
	REFERÊNCIAS	48
	 APÊNDICES	 50
	APÊNDICE A – EFICIÊNCIAS DE NASH-SUTCLIFFE DAS FORMULAÇÕES EM RNA PARA ANÁLISE ANUAL	51
	APÊNDICE B – EFICIÊNCIAS DE NASH-SUTCLIFFE NA FORMULAÇÃO 1 EM RNA PARA ANÁLISE MENSAL	53
	APÊNDICE C – EFICIÊNCIAS DE NASH-SUTCLIFFE NA FORMULAÇÃO 2 EM RNA PARA ANÁLISE MENSAL	54
	APÊNDICE D – EFICIÊNCIAS DE NASH-SUTCLIFFE NA FORMULAÇÃO 3 EM RNA PARA ANÁLISE MENSAL	55
	APÊNDICE E – EFICIÊNCIAS DE NASH-SUTCLIFFE NA FORMULAÇÃO 4 EM RNA PARA ANÁLISE MENSAL	56
	APÊNDICE F – EFICIÊNCIAS DE NASH-SUTCLIFFE DAS FORMULAÇÕES EM KNN PARA ANÁLISE ANUAL	57
	APÊNDICE G – EFICIÊNCIAS DE NASH-SUTCLIFFE NAS FORMULAÇÕES 1, 2, 3 E 4 NO KNN PARA ANÁLISE MENSAL	59

1 Introdução

1.1 Considerações Iniciais

O Brasil é abundante em termos de quantidade de água disponível, mas é fato o difícil acesso ao recurso hídrico na região Nordeste do país. Essa região possui o maior volume de água represado em regiões semiáridas do mundo. Por conseguinte, particularmente pelas variações climáticas, falta uma política coerente para que se tenha a distribuição desse volume de água de forma efetiva e rápida, para o atendimento das necessidades das populações que não têm acesso à água (GONÇALVES; OLIVEIRA, 2009).

A regularização do regime de vazões do rio São Francisco começou com a construção do reservatório de Sobradinho, que contribuiu para a diminuição das abundantes cheias nas regiões após o reservatório. Contudo, as atividades tradicionais das populações das regiões foram afetadas com a ocupação vasta da região com a construção da barragem (MARTINS et al., 2011).

A estimativa da vazão natural em certa seção de uma bacia hidrográfica pode ser feita através do processo de reconstituição. Esse processo pondera a vazão observada no local e as informações referentes às alterações feitas pelo homem, tais como: evaporações em reservatórios, regularizações de vazões processadas pelos reservatórios, desvios de água, e uso frequente de água para abastecimento na zona urbana, rural e industrial, além do uso na irrigação e criação de animais (ALMEIDA et al., 2016).

A modelagem chuva-vazão contribui na coordenação dos diversos usos da água, como também, na previsão de chuvas e estiagens em determinada região (DANTAS, 2022). É possível se notar os recentes desenvolvimentos da inteligência artificial, no campo de aprendizagem de máquina, tendo em vista a vasta expansão da capacidade da modelagem empírica. A área abordada nesta pesquisa é chamada de modelagem guiada por dados (*DDM: data driven modeling*), que se baseia na análise dos dados de um sistema (CELESTE; MENEZES; FONTES, 2014).

É indispensável que se tenha os conhecimentos dos processos hidrológicos e aspectos físicos da bacia hidrográfica para a utilização dos modelos conceituais de transformação da chuva em vazão. Porém, tais processos e aspectos citados não são necessários para utilizar DDM (SOUSA; SOUSA, 2006).

Para se fazer a modelagem chuva-vazão, vários tipos de DDM podem ser empregadas. Redes neurais artificiais (RNA) são modelos empíricos amplamente utilizados por terem a capacidade de modelar vasta quantidade de dados e sistemas não lineares complexos no processo de chuva-vazão (VILANOVA, 2017). Alternativamente, o aprendizado baseado em instância (IBL: *instance-based learning*) é outro procedimento utilizado para modelagem chuva-vazão tendo como mais tradicional modelo o algoritmo dos *k*-vizinhos mais próximos (KNN) (AKBARI;

OVERLOOP; AFSHAR, 2011).

Este trabalho apresenta a aplicação de RNA e KNN para a modelagem do procedimento de chuva vs. vazão natural afluente no reservatório de Sobradinho.

1.2 Objetivos

1.2.1 Objetivo Geral

Propor a utilização da modelagem chuva vs. vazão através das metodologias de redes neurais artificiais e do algoritmo dos k-vizinhos mais próximos no reservatório de Sobradinho, localizado no Estado da Bahia.

1.2.2 Objetivos Específicos

- Averiguar a eficiência das estratégias DDM para análise vazão-vazão anual, i.e., considerando apenas vazões passadas para prever vazões de anos futuros em escala anual;
- Averiguar a eficiência das estratégias DDM para análise chuva-vazão em escala mensal;
- Analisar a influência da escolha randômica dos dados de calibração e validação.

2 Revisão Bibliográfica

2.1 Modelagem Chuva-Vazão

A modelagem chuva-vazão é o processo que relaciona a precipitação na bacia hidrográfica com a vazão escoada em um rio (SCHEIDT; BRUNETTO, 2011). Essa modelagem é muito utilizada para previsões de vazões de reservatórios e bacias hidrográficas, ocorrências de cheias, falha ou falta de dados de estações pluviométricas e fluviométricas (ALMEIDA; SERRA, 2017).

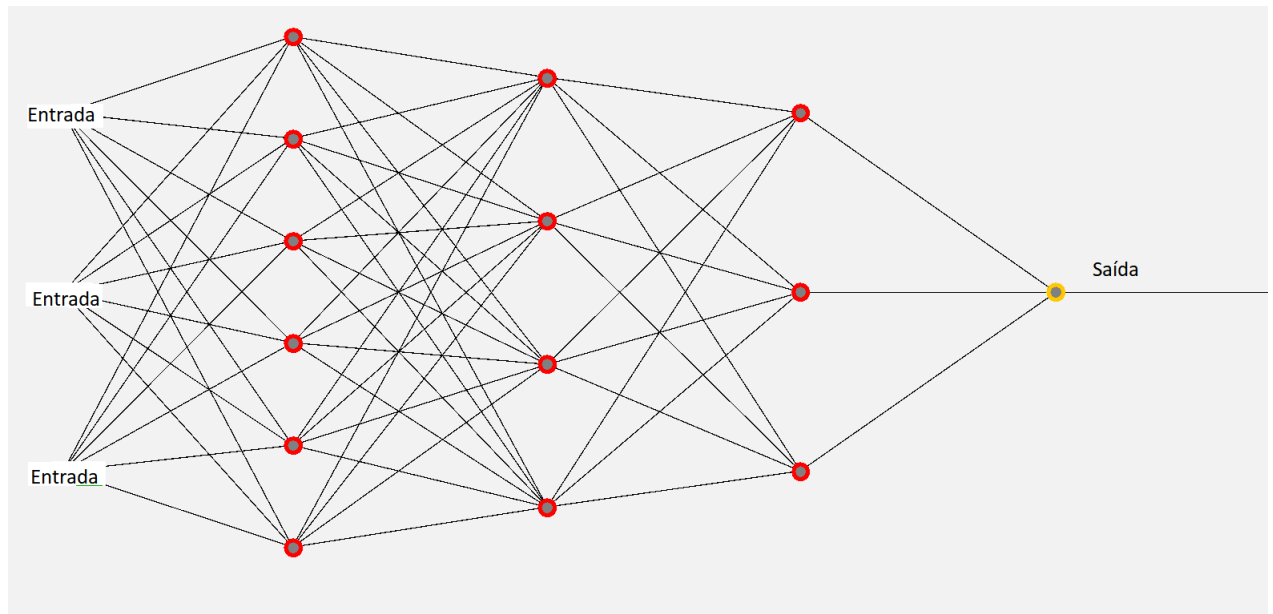
Os modelos hidrológicos podem ser classificados em empíricos, matemáticos e físicos. Os primeiros usam a abordagem da modelagem guiada por dados que assegura a análise das correlações dos dados de entrada e saída. Além disso, usa-se um conjunto de dados de treinamento. Após a realização do treinamento, posteriormente, testa-se o modelo com dados independentes (CELESTE; MENEZES; FONTES, 2014; SOLOMATINE; MASKEY; SHRESTHA, 2008). Ao longo dos anos, foram realizados diversos trabalhos relacionados à modelagem chuva-vazão.

Dantas (2022) utilizou a aplicação das redes neurais artificiais e do algoritmo dos k-vizinhos mais próximos para modelar o processo chuva-vazão na bacia hidrográfica do rio Piauitinga, em Sergipe. O autor utilizou os dados de entrada na modelagem do posto fluviométrico do município de Estância e do posto pluviométrico do município de Salgado. Os dados foram divididos entre calibração e validação, selecionados de maneira randômica. Foi utilizado o software Weka para realizar as simulações. A eficiência das simulações foi medida a partir do coeficiente de Nash-Sutcliffe e se obtiveram resultados consideráveis, principalmente nos modelos baseados em redes neurais artificiais.

Vilanova (2017) fez a utilização das redes neurais artificiais para simulação da vazão na bacia hidrográfica do Rio Itapemirim-ES. Nesse trabalho, o objetivo era aplicar e testar a viabilidade de utilização de RNA como escolha para simular a vazão na bacia hidrográfica do Rio Itapemirim-ES. Os resultados obtidos a partir dos testes realizados mostraram que a estimativa a vazão diária na bacia estudada se deu de forma satisfatória com a utilização de RNA e dos dados de precipitação como variáveis de entrada.

Celeste e Chaves (2014) fizeram o uso do modelo conceitual Tank Model para simular o processo de chuva-vazão na bacia hidrográfica do rio Japaratuba, em Sergipe. Nesse estudo, para realizar a calibração automática dos parâmetros do modelo foram analisados diferentes algoritmos otimizadores com funções objetivas. Em suma, os autores obtiveram resultados do trabalho e realizaram a comparação com uma pesquisa anterior, que fez o uso da combinação única de otimizador e funções objetivas, e com o modelo conceitual MODHAC (VIEGAS FILHO et al., 1999), realizado para estudo do Plano Estadual de Recursos Hídricos do estado de Sergipe. Por fim, os autores concluíram que foram alcançados resultados mais eficientes do que os obtidos no estudo anterior e no MODHAC, e que podem ser calibrados parâmetros do modelo

Figura 1 – Representação simplificada de uma rede neural artificial



Fonte: O autor

por diversos arranjos de otimizadores e funções objetivas.

2.2 Redes Neurais Artificiais

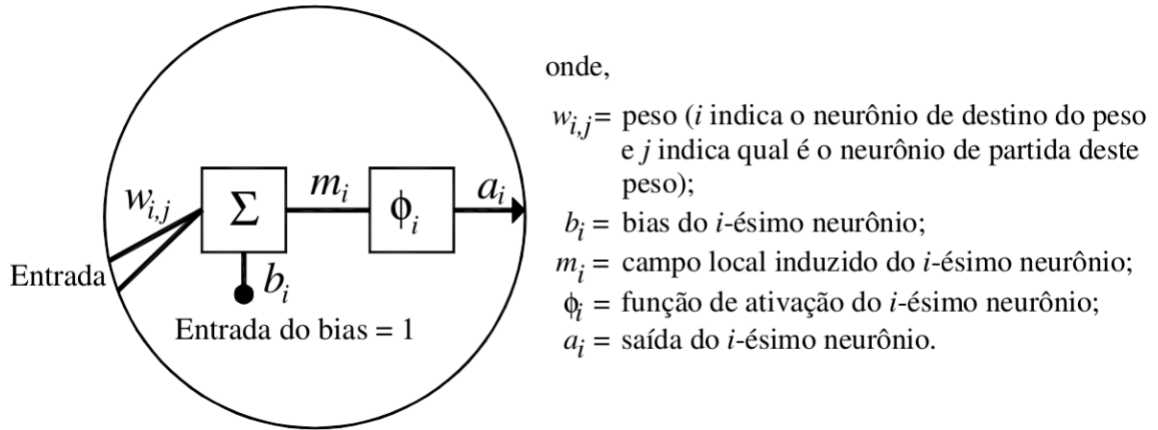
As redes neurais artificiais são sistemas computacionais habilitados a estabelecer conexões não lineares entre informações de entrada e saída de um sistema. Essa esquematização é composta por neurônios artificiais que utilizam informações para processamento dos dados e gerar uma saída (CELESTE; MENEZES; FONTES, 2014).

É possível notar que as redes neurais artificiais são caracterizadas por sua arquitetura e durante o processo de aprendizado, os pesos associados às ligações são ajustados iterativamente. A arquitetura utilizada nesta pesquisa é do tipo *feedforward* (Figura 1), composta por camada única ou múltiplas (FERNEDA, 2006). Neste tipo de RNA, a quantidade de neurônios na camada de entrada altera-se de acordo com a quantidade de variáveis de entrada inseridas. É visto também que elementos de entrada estão ligados aos neurônios na camada oculta e os dados das camadas ocultas estão ligados ao neurônio de saída (composta por um neurônio apenas, nesta pesquisa) (CELESTE; MENEZES; FONTES, 2014).

Os pesos das conexões são calibrados no processo de treinamento, com o objetivo de reduzir os erros no processo de saída da RNA (MACHADO, 2005).

Na Figura 2 é possível notar que em cada camada oculta há um neurônio artificial cuja configuração é conectada a cada elemento de entrada. Dessa forma, o campo local induzido (m_i) é gerado pela soma dos dados de entrada multiplicados pelos pesos (w_{ij}) e do *bias* (b_i). Por fim, para que os dados sejam transferidos aos neurônios adjacentes, sabe-se que o campo local induzido é submetido a uma função de ativação ϕ_i , a partir da qual obtém-se o valor de

Figura 2 – Detalhes de um neurônio artificial na camada oculta



Fonte: Farias, Santos e Celeste (2010)

entrada do neurônio subsequente da topologia (CELESTE; MENEZES; FONTES, 2014).

A função de ativação ϕ_i proporciona que a RNA tenha relações entre as variáveis de entrada e saída (MACHADO, 2005). Um exemplo tradicional de função de transferência é a função sigmoide, mostrada abaixo:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

2.3 Algoritmo dos k -vizinhos Mais Próximos

O algoritmo dos k -vizinhos mais próximos é um dos tipos de aprendizado baseado em instâncias que possui uma base na combinação de dados já existente em seu banco de dados (registros históricos, instâncias, resultados de um experimento, etc.) Com esses dados é possível realizar a previsão da saída para uma nova consulta de entrada (que não estão necessariamente armazenadas previamente no banco de dados) (SOLOMATINE; MASKEY; SHRESTHA, 2008).

O IBL decorre do método chamado *lazy learning*, composto pelo armazenamento de dados de treinamento. Ao obter nova entrada, o modelo procura na memória um conjunto de instâncias relacionadas cujas saídas são usadas para obtenção de uma nova saída (SOLOMATINE; MASKEY; SHRESTHA, 2008).

O algoritmo dos vizinhos mais próximos (*nearest neighbor*) aceita um vetor de atributos que não é incluído no conjunto de treinamento (*query vector* X_q) (AKBARI; OVERLOOP; AFSHAR, 2011). Posteriormente, uma norma Euclidiana ponderada é normalmente usada para fazer a medição de proximidade entre os vetores X_i (armazenados no banco de dados de treinamento) e X_q , respectivamente, vetor de atributos e vetor de consulta:

$$d(X_q, X_i) = \sqrt{\sum_{j=1}^n w_j^A (x_{ij} - x_{qj})^2} \quad (2.2)$$

onde, w_j^A é o peso do j -ésimo atributo ($i = 1, \dots, n$).

Com os valores de saída y_i dos k pontos mais próximos (vizinhos), o valor previsto para y_q é calculado pela média ponderada:

$$y_q = \frac{\sum_{i=1}^k w_i^V y_i}{\sum_{i=1}^k w_i^V} \quad (2.3)$$

onde, w_i^V é o peso do i -ésimo vizinho.

Na versão mais simples, $w_i^V = 1/k$ e a estimativa é o valor médio de k vizinhos mais próximos. Entretanto, o algoritmo dos k vizinhos mais próximos possui uma versão mais sofisticada, com o uso das funções *kernel*, em que cada vizinho recebe um peso de acordo com a distância $d(X_q, X_i)$, na qual o vizinho mais distante recebe um peso menor, tendo menor efeito entre os vizinhos que estão mais próximos.

Algumas funções *kernel* que decrescem de forma invariável à medida que sua distância aumenta, são listadas abaixo (AKBARI; OVERLOOP; AFSHAR, 2011):

$$\text{Linear: } w_i^V = 1 - d(X_q, X_i)$$

$$\text{Exponencial: } w_i^V = \exp[-d(X_q, X_i)]$$

$$\text{Gaussiana: } w_i^V = \exp\{-[d(X_q, X_i)]^2\}$$

$$\text{Inversa: } w_i^V = [d(X_q, X_i)]^{-1}$$

$$\text{Inversa quadrática: } w_i^V = \{[d(X_q, X_i)]^2\}^{-1}$$

Contudo, não existem indícios que uma função kernel única seja superior às outras, de acordo com Atkeson (1997, apud Akbari, Overloop e Afshar, 2011).

O número de vizinhos é escolhido empiricamente, por validação cruzada ou pelo conhecimento prático, de acordo com Kang e Cho (2008, apud Akbari, Overloop e Afshar, 2011).

2.4 Eficiência de Nash-Sutcliffe

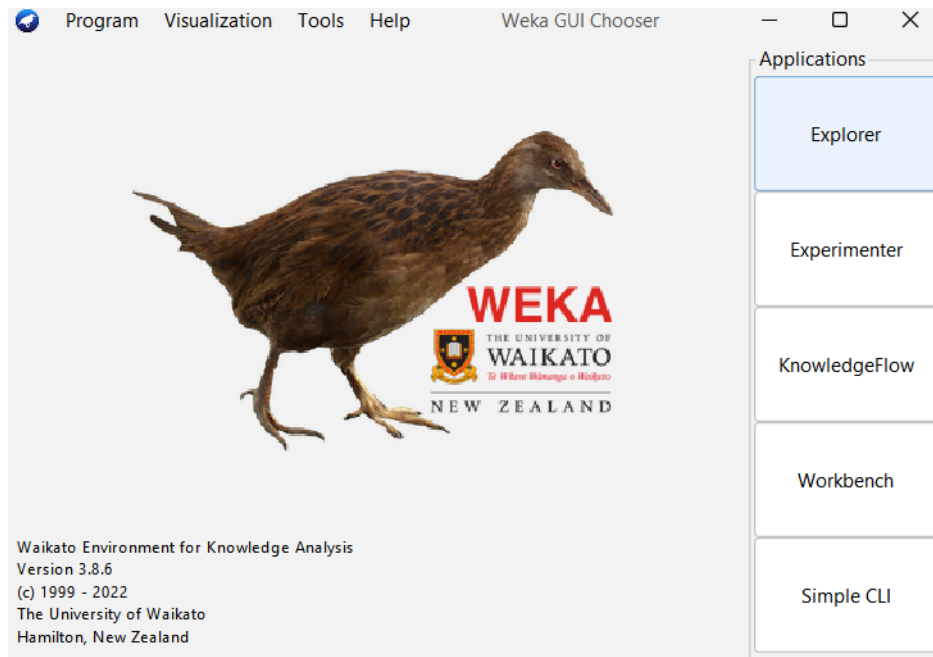
A performance da modelagem pode ser medida por meio da eficiência de Nash-Sutcliffe (E):

$$E = 1 - \frac{\sum_{i=1}^m (y_i - Y_i)^2}{\sum_{i=1}^T (y_i - \bar{y})^2} \quad (2.4)$$

onde,

- y_i é o valor observado
- Y_i é o valor modelado
- $\bar{y}$ é a média dos valores observados
- m é o número de amostras (instâncias).

Figura 3 – Menu de acesso



Fonte: Software Weka

Sabe-se que a eficiência de Nash-Sutcliffe pode variar de $-\infty$ a 1 e, como consequência, quando mais próximo de 1, o modelo é melhor.

Quando $E = 1$ diz-se que os dados observados do modelo têm um ajuste perfeito. Se a eficiência estiver inferior a zero é notório que a média observada fornece resultado melhor que o modelo. Já quando $E = 0$ diz-se que os dados observados do modelo teve previsão tão boa quando à média (CELESTE; CHAVES, 2014).

2.5 Weka

O software de aprendizado de máquina Weka (*Waikato Environment for Knowledge Analysis*) teve advento na Universidade de Waikato na Nova Zelândia, possui código aberto, utiliza a linguagem de programação Java e fornece diversos algoritmos para a mineração de dados. Possui uma interface gráfica que facilita a manipulação pelo usuário na aplicação dos diversos algoritmos (ALCÂNTARA, 2012).

O Weka é um software livre sob a licença GPL e é disponibilizado no seguinte site: (<https://www.cs.waikato.ac.nz/ml/weka/>).

Além disso, o Weka aplica técnicas de *machine learning* sem a obrigação de precisar escrever códigos de programação (DANTAS, 2022). O presente trabalho utilizou esse software para a aplicação de RNA e KNN na modelagem chuva-vazão afluente ao reservatório de Sobradinho. A janela de boas vindas inicial do software é mostrada na Figura 3.

3 Material e Métodos

3.1 Área de Estudo

A área de estudo deste trabalho trata-se da bacia de contribuição ao reservatório de Sobradinho, na Bahia, o qual se encontra dentro da bacia hidrográfica do rio São Francisco.

De acordo com Mendes (2012), a bacia do São Francisco possui uma área de drenagem de 636.920 km² e está dividida em quatro regiões fisiográficas – Alto, Médio, Submédio e Baixo São Francisco. O rio percorre sete unidades da federação (Minas Gerais, Goiás, Bahia, Pernambuco, Alagoas, Sergipe e parte do Distrito Federal) ao longo de seus 2.863 km de extensão.

Para esse estudo, foram utilizadas as seguintes 8 sub-bacias, enumeradas conforme classificação da Agência Nacional de Águas e Saneamento Básico (ANA), todas localizadas a montante de Sobradinho, que fazem, portanto, parte de sua área de drenagem: 40, 41, 42, 43, 44, 45, 46 e 47. Tais sub-bacias contribuem ao reservatório de Sobradinho. Os mapas de todas essas sub-bacias com estações da rede hidrometeorológica nacional (fluviométricas e pluviométricas) podem ser encontrados no portal de metadados da ANA: (<<http://metadados.ana.gov.br>>).

3.2 Dados de Precipitação e Vazão

3.2.1 Aquisição de Dados

Nesta pesquisa, um objetivo inicial era modelar o processo vazão-vazão em Sobradinho de forma a prever a vazão total do próximo ano baseada nas vazões de anos passados. Com a previsão da vazão anual futura, os dados seriam desagregados em estimativas mensais por meio do Método dos Fragmentos (SVANIDZE, 1980). Foram feitas diversas modelagens, com várias configurações de atributos (apenas vazão), porém, os resultados foram todos insatisfatórios. Por essa razão, decidiu-se utilizar a análise mensal, com os dados de chuva e vazão.

Para se fazer a modelagem chuva vs. vazão é necessária a obtenção de dados de vazão e precipitação, visto que essas informações fazem parte dos atributos de entrada e saída no software.

Os dados pluviométricos aplicados na modelagem foram obtidos do portal HidroWeb (<<https://www.snirh.gov.br/hidroweb>>), uma plataforma da ANA que permite o acesso às informações de estações hidrometeorológicas. Por outro lado, os dados de vazões naturais afluentes ao reservatório de Sobradinho foram baixados no portal da Câmara de Comercialização de Energia Elétrica (CCEE) (<<https://www.ccee.org.br/>>).

Assim sendo, foram baixados do HidroWeb os arquivos de registros correspondentes as séries históricas de chuva dos postos das sub-bacias afluentes ao reservatório de Sobradinho

Figura 4 – Seleção de séries históricas de chuva para a sub-bacia 43

Fonte: HidroWeb

Figura 5 – Estações Convencionais de chuva da Sub-bacia 43

Selecionar	Código	Nome Estação	Tipo Estação
☒	1545004	GAUCHOS	Pluviométrica
☒	1545006	VÁRZEA BONITA	Pluviométrica
☒	1545008	PARQUE DE EXPOSIÇÕES	Pluviométrica
☒	1546000	ARINOS - MONTANTE	Pluviométrica
☒	1546001	BURITIS - JUSANTE	Pluviométrica

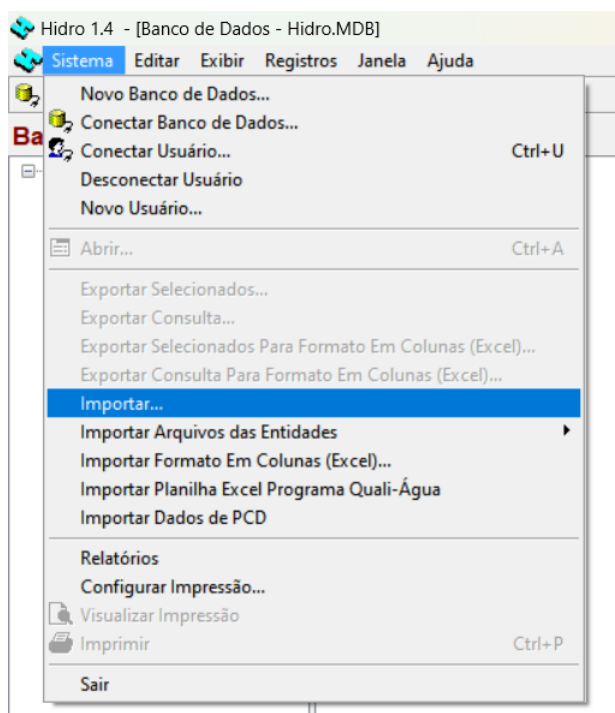
Fonte: HidroWeb

(bacias 40 até 47). Acessando o portal HidroWeb, seleciona-se a aba de *séries históricas*, a qual fornece tela para pesquisar por estação. Assim, preenchem-se os dados com o tipo de estação, bacia, sub-bacia, órgão responsável e a seleção dos postos que estão operando no momento. Posteriormente, clica-se em *consultar*. A Figura 4 mostra exemplo do procedimento de preenchimento para a sub-bacia 43.

Após a consulta, aparece na tela da plataforma as estações convencionais disponíveis. Assim, selecionam-se as estações e escolhe-se o arquivo em formato de banco de dados a ser baixado, conforme observa-se na Figura 5.

O arquivo no formato de banco de dados pode ser importado pelo software Hidro, também disponível no portal HidroWeb, que implementa várias funções de análise dos dados.

Figura 6 – Importação de Dados



Fonte: Hidro

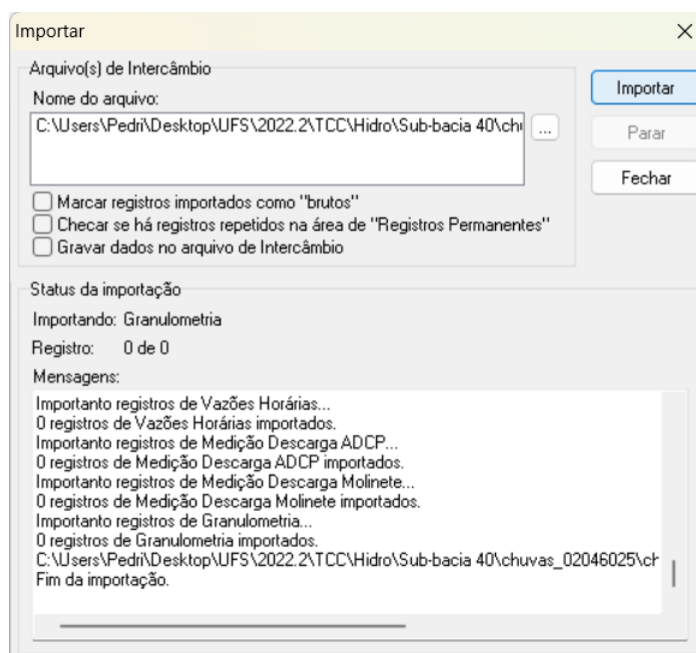
A Figura 6 mostra os passos a se seguir. Para fazer a importação dos dados, seleciona-se na aba *sistema* a opção *importar*. Logo após, escolhe-se o arquivo a ser importado e por fim seleciona-se a opção *importar* novamente, conforme mostrado na Figura 7.

Para fazer a análise dos dados importados para o Hidro é necessário selecionar a aba *funções* e escolher a opção *chuvas diárias*, como mostra na Figura 8. Logo depois, escolhe-se o arquivo importado com dados brutos e assim aparecem na tela os dados, conforme Figura 9. Para copiar a série, seleciona-se a aba *editar* e clica-se na opção *copiar série*. Nesta pesquisa, copiaram-se todos os dados das sub-bacias em estudo para uma planilha a fim de se estimar a chuva média na bacia de contribuição à Sobradinho. Foi utilizado o método da média aritmética.

A quantidade de postos de chuva utilizados das sub-bacias a montante do reservatório de Sobradinho pode ser observada na Tabela 1. Inicialmente, para cada sub-bacia foi calculada a chuva média com base em seus postos. Quando havia ausência ou falha em algum dado, esse era desconsiderado no cálculo da média, ou seja, a média de determinada data era calculada apenas com base nos dados sem falhas dos postos da bacia. Por fim, chegaram-se a valores médios de chuva para cada sub-bacia nos períodos mostrados na Tabela 2.

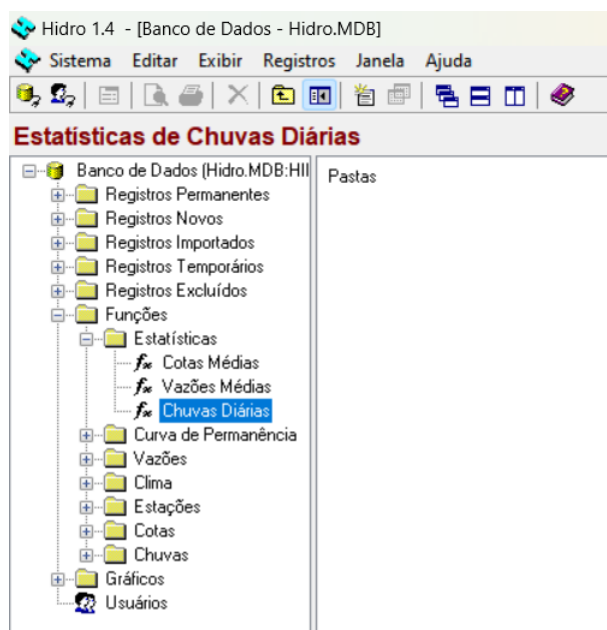
A partir disso, foi realizada as médias das chuvas médias nas sub-bacias para se chegar à série temporal de precipitações médias que contribuem para a área de drenagem de Sobradinho. Essa série final consistiu de valores de 1931 até 2022.

Figura 7 – Seleção para importação de arquivo baixado no HidroWeb



Fonte: Hidro

Figura 8 – Coleta de dados



Fonte: Hidro

Figura 9 – Totais mensais da série da estação Fazenda Samburá - Posto 02046025

Hidro 1.4 - [Estatísticas de Chuvas]

Sistema Editar Exibir Janela Ajuda

Fechar

Séries

Série: FAZENDA SAMBURÁ 02046025 (Importado, Bruto, 02/1984 - 09/2022)

Totais Diários Gráfico de Totais Diários Totais Mensais Máximas Mensais Dias de Chuva Mensais Leituras Mensais Situação Mensal

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
1984			177,1	85,5	65,2	2,6	1,0	55,9	114,3	96,0	159,8	340,8	
1985	457,5	126,9	238,2	97,5	20,3	5,1	0,0	2,1	46,4	70,8	180,7	302,4	1547,9
1986	354,5	207,4	127,4	31,2	69,7	1,0	43,7	123,2	20,1	29,3	182,8	481,5	1671,8
1987	254,4	176,0	90,2	162,8	14,7	9,2	2,2	2,1	129,0	154,0	248,4	299,4	1542,4
1988	118,9	377,6	132,2	53,2	51,6	4,0	0,0	0,0	34,0	108,2	140,6	244,0	1264,3
1989	224,8	294,9	90,7	57,6	21,0	34,0	12,1	50,0	111,3	83,1	240,2	264,0	1483,7
1990	131,0	64,2	128,3	125,9	74,8	0,0	29,6	79,1	39,8	77,6	105,1	111,7	967,1
1991	473,3	210,1	376,6	175,9	62,7	1,4	0,0	0,0	51,7	97,4	125,3	281,3	1855,7
1992	489,2	369,7	152,7	73,7	76,8	0,0	4,1	12,2	158,6	192,8	156,0	338,2	2024,0
1993	183,7	460,7	355,6	64,5	2,8	2,8	0,0	21,0	76,1	106,4	18,2	177,6	1469,4
1994	361,3	90,8	154,0	23,7	33,2	0,6	0,0	0,0	1,5	160,5	229,8	250,2	1305,6
1995	31,1	408,9	99,3	47,7	9,8	11,3	2,1	0,0					
1996													
1997													
1998													
1999						29,5	3,5	0,0	68,0	46,2	165,7	317,9	
2000	379,5	216,7	195,4	15,1	2,4	0,0	4,9	5,8	135,8	67,8	280,7	298,0	1602,1
2001	157,5	84,8	197,4	8,0	53,5	0,0	8,0	70,7	65,3	105,6	122,0	404,9	1277,7
2002	195,0	356,4	94,7	9,8	125,4?	0,0	35,4	0,0	66,3	47,8	180,0	342,0	1452,8?
2003	566,5	98,4	187,1	32,5	42,0	0,0	0,4	11,2	35,4	62,0	222,8	209,1	1467,4
2004	283,4	538,5	199,0	128,7	23,6	9,7	41,5	0,0	34,2	177,3	120,4	375,3	1931,6
2005	377,1	93,7	184,6	36,7	110,8	18,4	10,9	3,2	130,9	48,2	311,1	354,7	1680,3
2006	190,0	137,0	281,8	36,9	12,1	6,5	0,1	31,7	45,8	263,8	110,5	392,1	1508,3
2007	535,4	189,7	88,9	165,5	29,3	0,0	19,2	0,0	0,0	129,9	215,5	217,4	1590,8
2008	480,2	276,6	254,2	153,9	8,1	11,9	0,0	24,3	118,1	47,0	119,4	482,5	1976,2
2009	292,5	267,6	258,4	95,2	46,1	45,1	0,4	50,8	56,6	173,9	105,3	224,0	1615,9
2010	176,9	84,4	368,8	24,0	43,2	1,1	1,2	0,0	69,5	214,1	257,8	221,6	1462,6
2011	219,3	140,6	318,6	97,3	5,8	18,2	0,0	0,0	2,6	85,6	54,6	319,4	1262,0
2012	411,1	74,7	139,3	32,7	41,7	119,9	5,3	0,0	28,0	60,6	223,3		
2013	371,2	108,6	228,7	54,4	47,2	17,3	11,8	17,0	53,6	119,6	258,3	163,9	1451,6
2014	79,0	53,8	147,0	126,8	26,8	0,0	32,2	0,0	11,7	97,1	264,4	172,3	1011,1
2015	133,7	253,7	328,8?	52,4	46,2	21,5	0,0	0,0	138,5	46,1	201,7	380,4	1603,0?
2016	328,2	105,9	81,9	10,1	4,2	83,1	0,0	13,0	15,1	99,2	152,3	257,2	1150,2
2017	230,4	118,5	45,6	49,7	81,9	35,9	0,0	0,0	11,0	152,3	291,7	258,5	1275,5
2018	98,8	172,3	95,4	51,5	27,7	0,2	0,0	27,8	64,1	99,8	596,5	139,1	1373,2
2019	91,5	241,6	190,2	108,0	2,5	5,5	0,0	3,8	57,8	94,0	213,4	177,6	1185,9
2020	417,6	425,7	106,0	91,0	73,2	0,0	0,0	0,5	13,4	123,8	95,7	219,8	1566,7
2021	163,0	133,7								254,2	181,7	231,7	
2022	353,4	377,5	104,7	36,6	32,7	0,0	0,0	7,5	43,8				
Média	282,7	215,8	182,9	71,1	40,9	14,2	7,7	17,5	60,2	111,5	192,1	280,3	1485,9

* - estimado; ? - duvidoso; \$ - acumulado

Fonte: Hidro

Tabela 1 – Quantidade de postos pluviométricos nas sub-bacias a montante do barramento de Sobradinho

Sub-bacia	Total	Operando	Responsável-ANA	Dados utilizados
40	299	183	60	58
41	211	130	46	40
42	175	100	39	37
43	48	30	14	13
44	179	101	44	31
45	142	94	52	36
46	281	169	87	40
47	170	68	24	24

Fonte: O autor

Tabela 2 – Períodos para os quais a chuva média foi obtida por sub-bacia

Sub-bacia	Período
40	1941-2021
41	1938-2021
42	1958-2021
43	1954-2021
44	1938-2021
45	1937-2021
46	1931-2021
47	1972-2021
Fonte: O autor	

Tabela 3 – Formulações para análise de vazões anuais via DDM

1	$Q(t) = h(Q(t-1), Q(t-2))$
2	$Q(t) = h(Q(t-1), Q(t-2), Q(t-3))$
3	$Q(t) = h(Q(t-1), Q(t-2), Q(t-3), Q(t-4))$
4	$Q(t) = h(Q(t-1)^2, Q(t-2)^2, Q(t-1)^3, Q(t-2)^3)$
5	$Q(t) = h(Q(t-2) \cdot Q(t-1), Q(t-3) \cdot Q(t-1), Q(t-3) \cdot Q(t-2))$
Fonte: O autor	

Tabela 4 – Formulações para análise mensal via DDM

1	$Q(t) = h(Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5))$
2	$Q(t) = h(P(t), P(t-1), P(t-2), P(t-3), P(t-4), P(t-5))$
3	$Q(t) = h(P(t), P(t-1), P(t-2), P(t-3), Q(t-1), Q(t-2), Q(t-3))$
4	$Q(t) = h(P(t-1), P(t-2), P(t-3), P(t-4), Q(t-1), Q(t-2), Q(t-3), Q(t-4))$
Fonte: O autor	

3.2.2 Formulações para Análise DDM

As modelagens RNA e KNN foram submetidas a quatro tipos de formulações, da seguinte forma:

$$h(x) = h(x_1, x_2, \dots, x_n) \quad (3.1)$$

onde $h(x)$ é a (hipótese para a) vazão do mês seguinte em função dos atributos $x_1, x_2, \dots, x_n$, os quais podem ser chuvas (do mês atual e/ou de meses passados) e/ou de vazões passadas. Quando a chuva atual é incluída, tem-se um modelo chuva-vazão tradicional que serve para extensão de séries de vazão com base em séries de chuvas observadas. Quando apenas dados passados são utilizados nos atributos, tem-se um modelo de previsão da vazão futura com base no passado.

Os dados de precipitação e vazão são organizados de maneira matricial em uma planilha. Desse modo, a partir de ferramentas de planilhas, foi realizada a modificação dos dados para forma de vetor, que organiza os atributos de entrada e os atributos de saída na planilha, para melhor apuração dos resultados. A escolha do número e dos tipos de atributos foi feita de forma empírica, conforme mostrado na Tabela 3 para análise anual e na Tabela 4 para análise mensal.

3.2.3 Classificação dos Dados das DDMs

Para realizar a modelagem de chuva vs. vazão, os dados coletados foram divididos em calibração e validação. A calibração serve para treinar o modelo e a validação serve para verificar se o modelo é válido para dados não usados no treinamento. A quantidade de dados foi dividida em 70% para calibração e 30% para validação.

3.3 Aplicação do Weka

3.3.1 Arquivo em ARFF

Para fazer a modelagem de chuva vs. vazão com o Weka, é preciso fazer a exportação dos dados para o software. Desse modo, como o Weka utiliza-se de dados no formato próprio ARFF, forçando a conversão dos dados da planilha para esse formato.

O formato ARFF deve seguir o modelo da Figura 10. Os dados devem ser separados por vírgula e o arquivo deve conter informações que são especificadas usando marcações. Por exemplo, os nomes do conjunto de dados e dos atributos são mostrados por meio das marcações *@relation* e *@attribute*, respectivamente. Já os dados em si são delimitados através da marcação *@data* (FRANK; HALL; WITTEN, 2016). Para colocar o arquivo no padrão ARFF, o manuseio é realizado por meio de um editor de texto. Dessa forma, o arquivo será salvo no formato ARFF.

Dessa maneira, para aplicar a modelagem foram necessários ser criados dois tipos de arquivos ARFF, o primeiro com nome *trainData*, composto pelos dados de calibração, e o segundo denominado de *allData*, que contém todo o conjunto de dados, composto dos dados de calibração e validação.

Para abrir um arquivo no Weka, deve-se selecionar opção *Explorer* na tela inicial do software como observa-se na Figura 3. Em seguida, carrega-se o arquivo *trainData* selecionando-se na opção *Open File* na aba *Preprocess*. É possível ver se os atributos do arquivo estão corretos, pois os mesmos são carregados conforme a Figura 11.

3.3.2 Escolha do Método e Configuração

A próxima etapa da utilização do Weka é a escolha do método de aprendizado de máquina e sua configuração. Para isso, é necessário clicar na guia *Classify*, escolher a opção *Choose*, para assim fazer a escolha do método, neste caso RNA ou KNN.

3.3.2.1 Configuração da RNA

Para aplicação da RNA no Weka, após clicar em *Choose*, clica-se na opção *functions* e faz-se a escolha ao clicar em *MultilayerPerceptron*, como mostra a Figura 12.

O Weka permite configurar para a RNA os neurônios e camadas ocultas. Por padrão, o Weka realiza de forma automática uma configuração para camadas ocultas, ou seja, o programa

Figura 10 – Configuração de um arquivo ARFF - *allData* da Formulação 1 da análise mensal

```

@RELATION allData

@ATTRIBUTE Q(t-1) NUMERIC
@ATTRIBUTE Q(t-2) NUMERIC
@ATTRIBUTE Q(t-3) NUMERIC
@ATTRIBUTE Q(t-4) NUMERIC
@ATTRIBUTE Q(t-5) NUMERIC
@ATTRIBUTE Q(t) NUMERIC

@DATA
4467,7333,6630,4805,4676,2397
2397,4467,7333,6630,4805,1870
1870,2397,4467,7333,6630,1619
1619,1870,2397,4467,7333,1448
1448,1619,1870,2397,4467,1376
1376,1448,1619,1870,2397,1935
1935,1376,1448,1619,1870,2207
2207,1935,1376,1448,1619,4036
4036,2207,1935,1376,1448,4531
4531,4036,2207,1935,1376,3214
3214,4531,4036,2207,1935,1976
1976,3214,4531,4036,2207,1425
1425,1976,3214,4531,4036,1292
1292,1425,1976,3214,4531,1176
1176,1292,1425,1976,3214,995
995,1176,1292,1425,1976,905
905,995,1176,1292,1425,982
982,905,995,1176,1292,2107
2107,982,905,995,1176,3594
3594,2107,982,905,995,5014
5014,3594,2107,982,905,5516
5516,5014,3594,2107,982,3429
3429,5516,5014,3594,2107,2692
2692,3429,5516,5014,3594,1905
1905,2692,3429,5516,5014,1357
1357,1905,2692,3429,5516,1250
1250,1357,1905,2692,3429,1072
1072,1250,1357,1905,2692,915
915,1072,1250,1357,1905,1194
1194,915,1072,1250,1357,1799

```

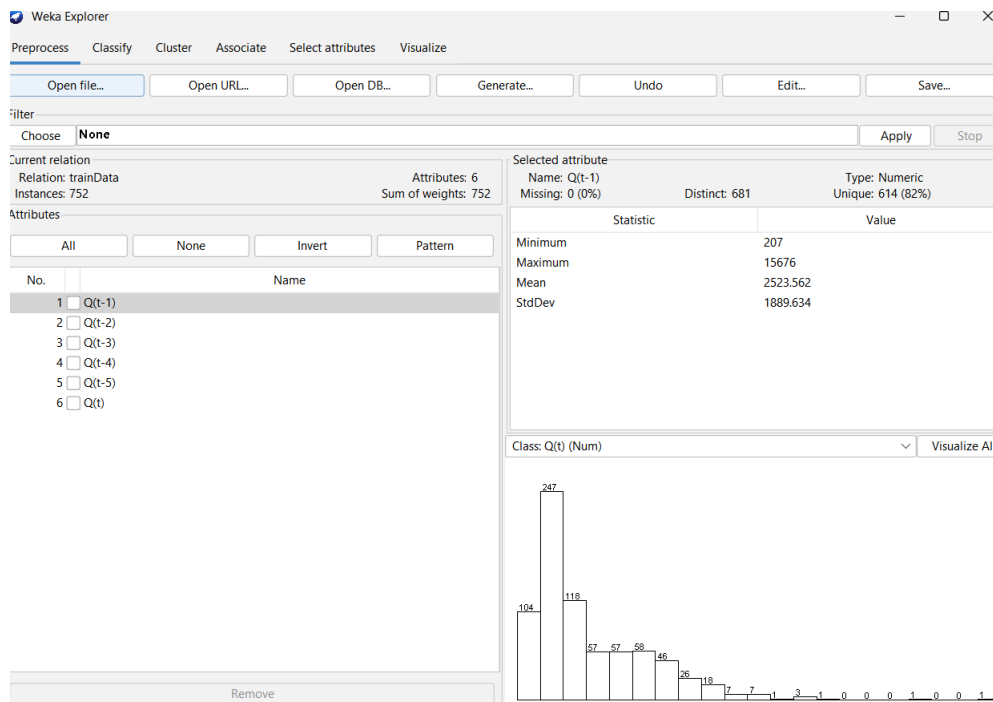
Fonte: Software Weka

roda com uma configuração com determinada quantidade de camadas e neurônios, baseada no número de atributos de entrada.

Por exemplo, na análise mensal, a configuração padrão da RNA da formulação 1 mostra 3 neurônios em uma única camada, como observado na Figura 13. Outrossim, o programa permite que o usuário do software faça alteração na quantidade de neurônios e camadas ocultas, possibilitando que o usuário teste qualquer configuração para o modelo.

No caso da análise anual, o Weka de forma automática aplicou três neurônios artificiais dispostos em uma única camada oculta para as formulações 1, 2, 3, 4 e 5. Para as cinco formulações foram utilizadas as configurações da Tabela 5.

Para a análise mensal, o Weka de forma automática aplicou três neurônios artificiais dispostos em uma única camada oculta para a formulação 1. Para formulação 2 aplicou seis

Figura 11 – Seleção da aba *Preprocess*

Fonte: Software Weka

Tabela 5 – Configurações das camadas ocultas na RNA para as cinco formulações da análise anual

RNA-2	RNA-3	RNA-5
RNA-2,2	RNA-3,2	RNA-5,3
RNA-2,2,2	RNA-3,2,2	RNA-5,3,2

Fonte: O autor

Tabela 6 – Configurações das camadas ocultas na RNA para a formulação 1 para análise mensal

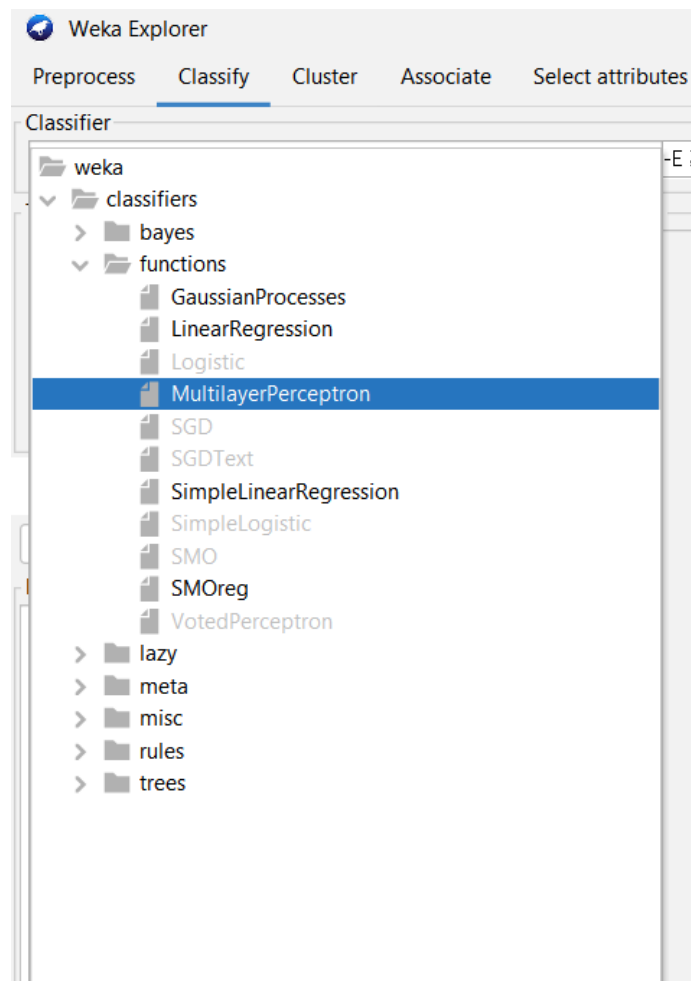
RNA-3	RNA-4	RNA-5	RNA-6
RNA-3,2	RNA-4,2	RNA-5,3	RNA-6,3
RNA-3,3	RNA-4,4	RNA-5,5	RNA-6,6
RNA-3,3,1	RNA-4,4,2	RNA-5,5,3	RNA-6,3,3
RNA-3,3,3	RNA-4,4,4	RNA-5,5,5	RNA-6,6,6

Fonte: O autor

neurônios artificiais dispostos em uma única camada oculta. Ademais, aplicou quatro neurônios artificiais dispostos em uma única camada oculta para as formulações 3 e 4.

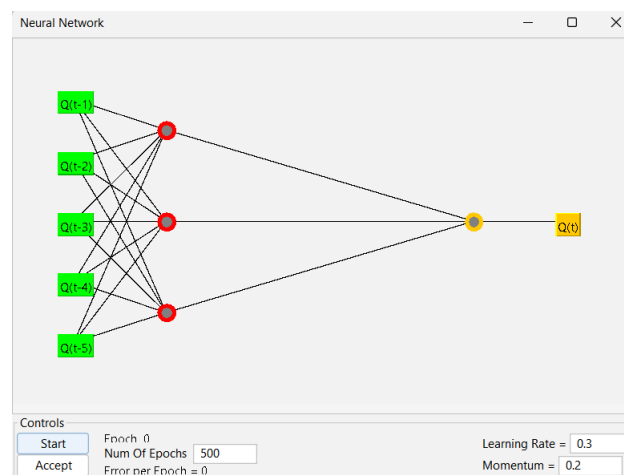
Além das configurações automáticas, foram feitas diversas simulações com outras configurações. As configurações foram escolhidas após a realização dos testes de uma única camada modificando a quantidade de neurônios. Por fim, foram escolhidas as configurações finais mostrada nas tabelas 6 até 9 para as formulações 1 até 4, respectivamente. Nas tabelas, os números representam a quantidade de neurônios por camada. Cada número representa uma camada, e.g., RNA-9,6,3 significa uma rede com 9 neurônios na primeira camada, 6 na segunda, e 3 na terceira.

Figura 12 – Localização de DDM da RNA



Fonte: Software Weka

Figura 13 – Exemplo de formulação 1 da visualização de quantidade de neurônios e camadas para o RNA, para análise mensal



Fonte: Software Weka

Tabela 7 – Configurações das camadas ocultas na RNA para a formulação 2 para análise mensal

RNA-6	RNA-8	RNA-9	RNA-10
RNA-6,3	RNA-8,4	RNA-9,6	RNA-10,5
RNA-6,6	RNA-8,8	RNA-9,9	RNA-10,10
RNA-6,3,3	RNA-8,8,4	RNA-9,6,3	RNA-10,7,5
RNA-6,6,6	RNA-8,8,8	RNA-9,9,9	RNA-10,10,10

Fonte: O autor

Tabela 8 – Configurações das camadas ocultas na RNA para a formulação 3 para análise mensal

RNA-4	RNA-5	RNA-7	RNA-8
RNA-4,4	RNA-5,5	RNA-7,7	RNA-8,8
RNA-4,2,1	RNA-5,3,2	RNA-7,5,3	RNA-8,4,4
RNA-4,2,2	RNA-5,3,3	RNA-7,5,5	RNA-8,8,4
RNA-4,4,4	RNA-5,5,5	RNA-7,7,7	RNA-8,8,8

Fonte: O autor

Tabela 9 – Configurações das camadas ocultas na RNA para a formulação 4 para análise mensal

RNA-3	RNA-4	RNA-5	RNA-7
RNA-3,2	RNA-4,2	RNA-5,3	RNA-7,5
RNA-3,3	RNA-4,4	RNA-5,5	RNA-7,7
RNA-3,3,2	RNA-4,4,2	RNA-5,5,3	RNA-7,7,5
RNA-3,3,3	RNA-4,4,4	RNA-5,5,5	RNA-7,7,7

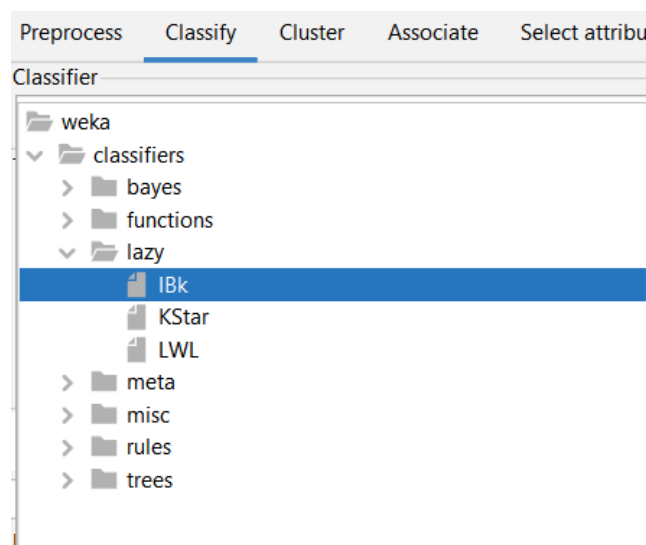
Fonte: O autor

3.3.2.2 Configuração do KNN

Para a aplicação do algoritmo KNN no Weka, após clicar em *Choose*, localiza-se e clica na pasta *lazy* e assim escolhe-se a opção *IBk*, como é mostrado na Figura 14.

A quantidade de *k*-vizinhos mais próximos foi escolhida de forma empírica. Para análise

Figura 14 – Localização de DDM do KNN



Fonte: Software Weka

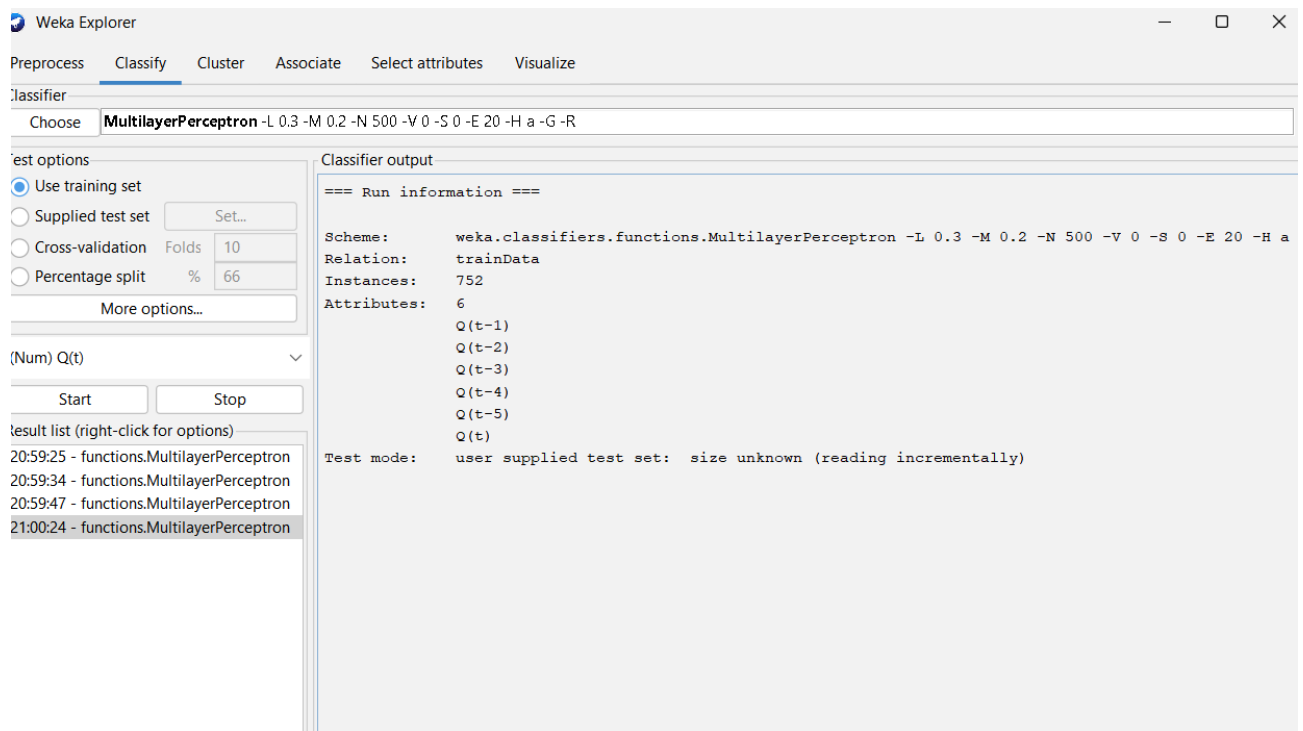
Tabela 10 – Configurações da quantidade de k -vizinhos aplicados no KNN para análise anual

1-NN	2-NN	3-NN
4-NN	5-NN	10-NN
Fonte: O autor		

Tabela 11 – Configurações da quantidade de k -vizinhos aplicados no KNN para análise mensal

1-NN	4-NN	10-NN
2-NN	5-NN	15-NN
3-NN	8-NN	20-NN
Fonte: O autor		

Figura 15 – Calibração da RNA



Fonte: Software Weka

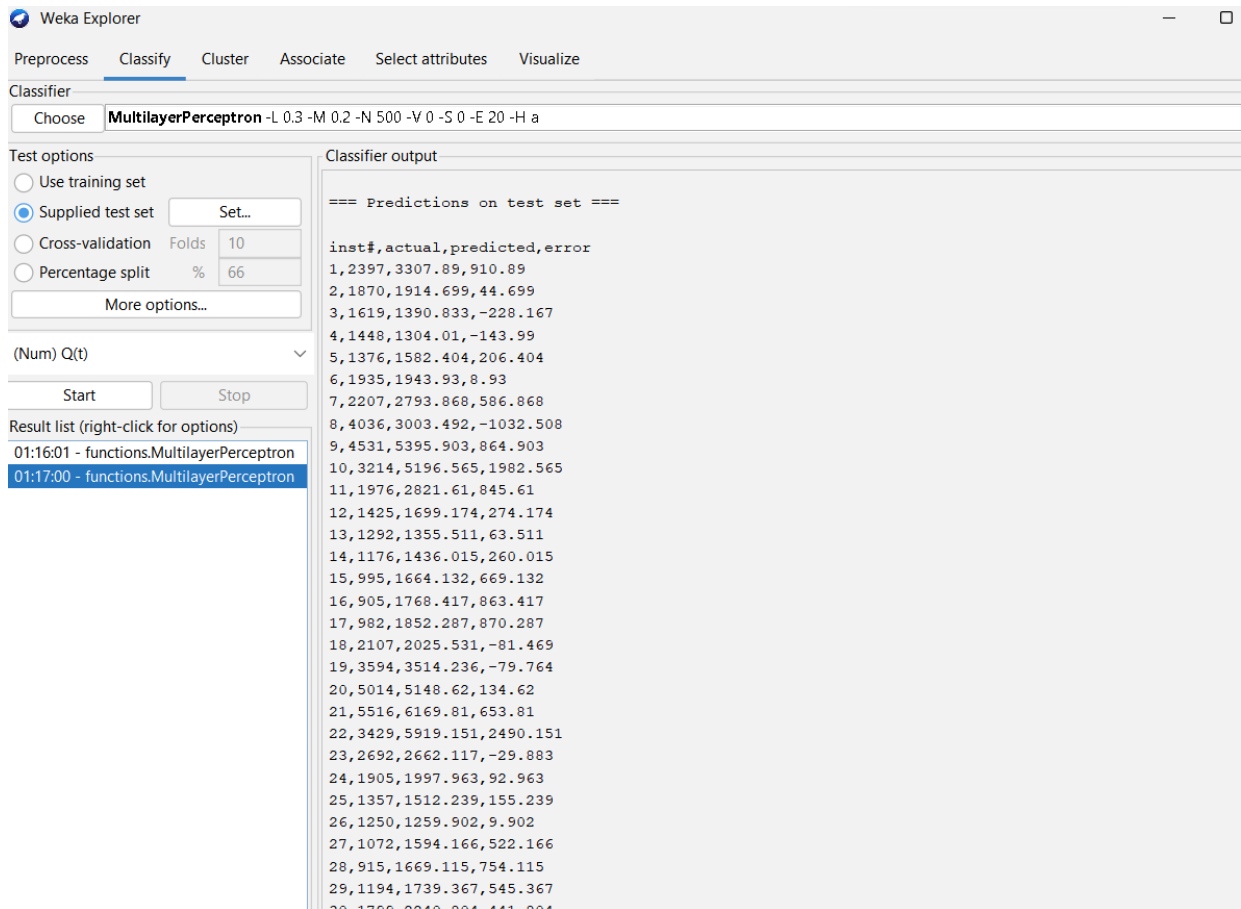
anual, foram realizadas várias simulações para o KNN conforme observação da Tabela 10. Na análise mensal, foram realizadas várias simulações para o KNN conforme observação da Tabela 11.

3.3.3 Execução dos Modelos

Depois de se escolher o modelo a rodar no Weka, escolhe-se a opção *Use training set*, carregando-se as informações de calibração contidas no arquivo *trainData.arff*, como mostra a Figura 15. Assim, é realizado o treinamento do método escolhido. Logo após, seleciona-se a opção *Supplied test set* e carregam-se os dados contidos no arquivo *allData.arff*, composto dos dados de calibração e validação.

Por fim, podem-se coletar os dados, copiá-los em um anotador de texto e deixá-los no

Figura 16 – Exemplo de resultados obtidos para RNA



Fonte: Software Weka

formato para serem copiados para uma planilha onde serão processados os resultados. A Figura 16 mostra os resultados finais fornecidos na tela do Weka a serem posteriormente copiados para a planilha que os processará, compostos pelas vazões simuladas, vazões observadas e o erro entre as vazões.

3.4 Verificação do Desempenho da Modelagem

A performance das modelagens RNA e KNN foi medida a partir dos resultados obtidos através do coeficiente de eficiência de Nash-Sutcliffe (Equação 2.4). Os valores foram calculados com os resultados gerados no Weka e processados em planilha separada. Foram calculados as eficiências de calibração, validação e completa (todos os dados).

Os valores dos resultados dos coeficientes de Nash-Sutcliffe foram classificados de acordo com Baltokoski et al. (2010 apud Lima, 2022), da seguinte forma:

- Muito Bom: $E > 0,75$;
- Bom: $0,65 < E \leq 0,75$;
- Satisfatório: $0,50 < E \leq 0,65$;

- Insatisfatório: $E \leq 0,50$.

Os resultados das modelagens e suas eficiências serão mostrados e discutidos no próximo capítulo.

4 Resultados e Discussão

4.1 Resultados das Modelagens

Os resultados das modelagens chuva vs. vazão afluente ao reservatório de Sobradinho foram desenvolvidos conforme descrito no capítulo anterior e para verificar seus desempenhos foram construídos gráficos a partir dos dados obtidos. Para a apresentação dos valores de índice de eficiência, também foram utilizados gráficos. Os gráficos apresentam o coeficiente de Nash-Sutcliffe para dados de validação como resultados da modelagem.

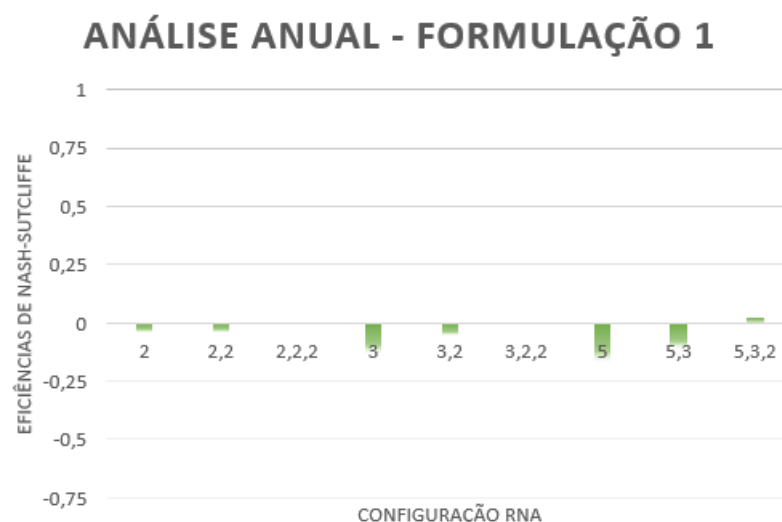
Nos apêndices, é possível observar adicionalmente os resultados de calibração e completo, que contribuem para a discussão final.

4.1.1 Análise Anual

4.1.1.1 Resultados da RNA

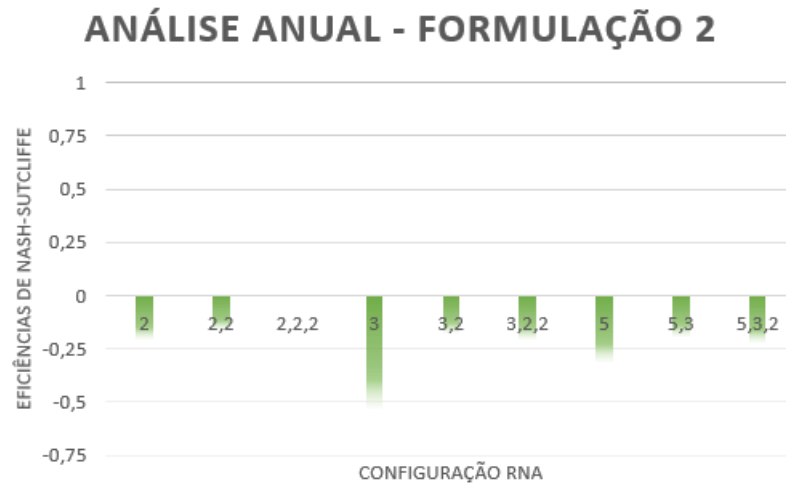
As figuras 17, 18, 19, 20 e 21 ilustram resultados de eficiência de Nash-Sutcliffe para os dados de validação das Formulações 1, 2, 3, 4 e 5 respectivamente, para análise anual via RNA. Além disso, no Apêndice A é possível observar os coeficientes de desempenho para os dados de validação, calibração e completo (validação e calibração) nas Formulações 1, 2, 3, 4 e 5.

Figura 17 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação de RNA (análise anual)



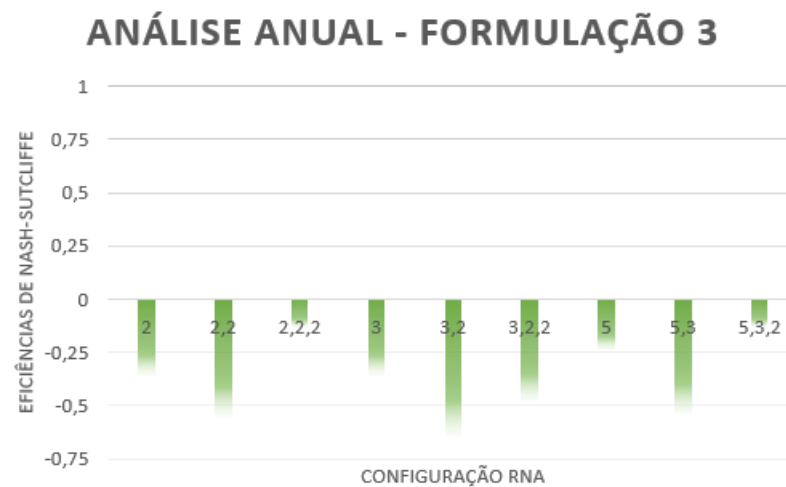
Fonte: O autor

Figura 18 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação de RNA (análise anual)



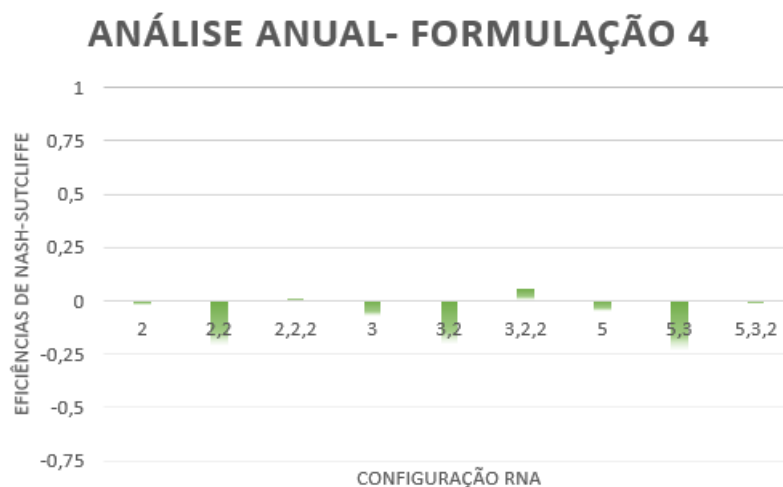
Fonte: O autor

Figura 19 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação de RNA (análise anual)



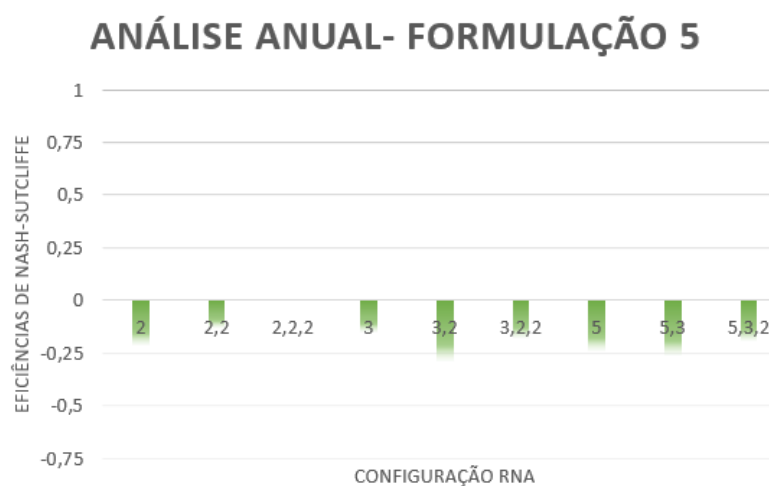
Fonte: O autor

Figura 20 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação de RNA (análise anual)



Fonte: O autor

Figura 21 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 5 com aplicação de RNA (análise anual)

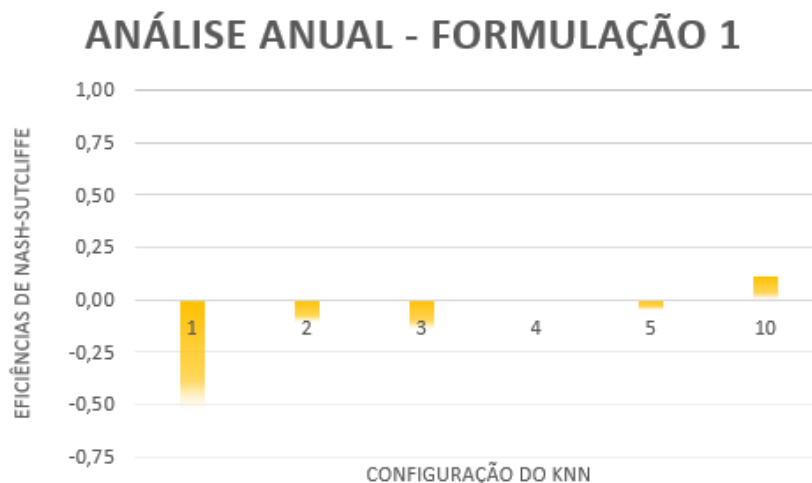


Fonte: O autor

4.1.1.2 Resultados do KNN

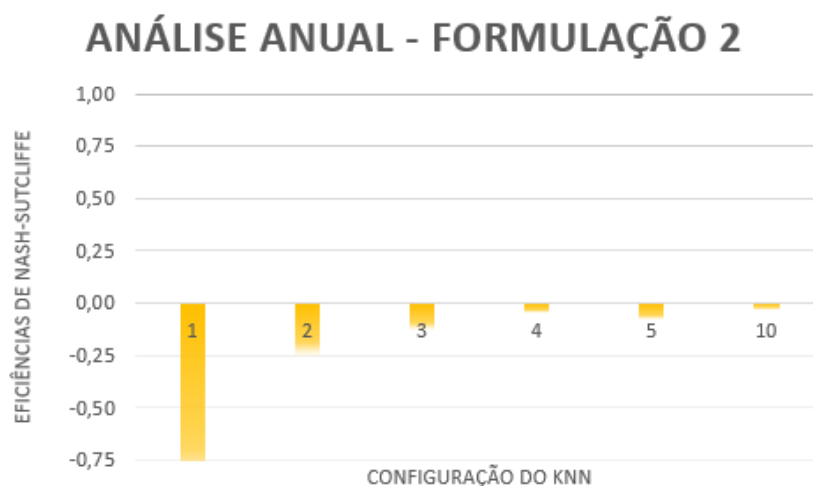
As figuras 22, 23, 24, 25 e 26 ilustram os resultados de eficiência de Nash-Sutcliffe referentes aos dados de validação das Formulações 1, 2, 3, 4 e 5 respectivamente, para análise anual via KNN. Além disso, no Apêndice F, é possível observar os coeficientes de desempenho para os dados de validação, calibração e completo nas Formulações 1, 2, 3, 4 e 5.

Figura 22 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação do KNN (análise anual)



Fonte: O autor

Figura 23 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação do KNN (análise anual)



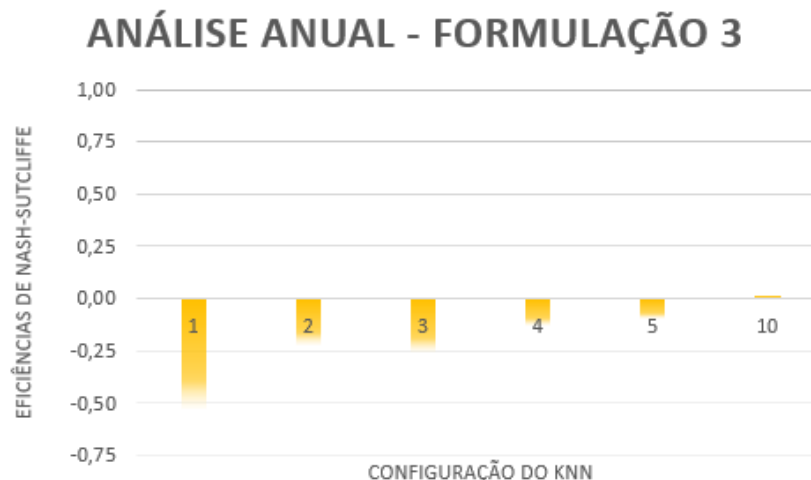
Fonte: O autor

4.1.2 Análise Mensal

4.1.2.1 Resultados da RNA

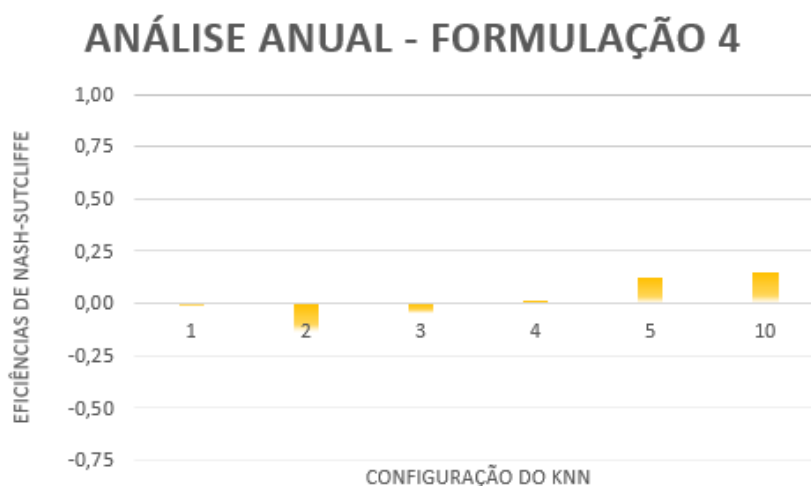
As figuras 27, 28, 29 e 30 ilustram resultados de eficiência de Nash-Sutcliffe para os dados de validação das Formulações 1, 2, 3 e 4, respectivamente, para análise mensal via RNA. Além disso, nos Apêndices B, C, D e E é possível observar os coeficientes de desempenho para os dados de validação, calibração e completo nas Formulações 1, 2, 3 e 4, respectivamente.

Figura 24 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação do KNN (análise anual)



Fonte: O autor

Figura 25 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação do KNN (análise anual)

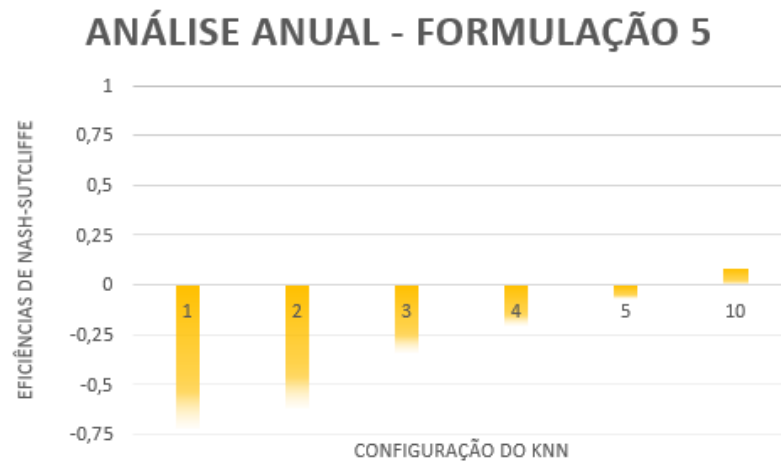


Fonte: O autor

4.1.2.2 Resultados do KNN

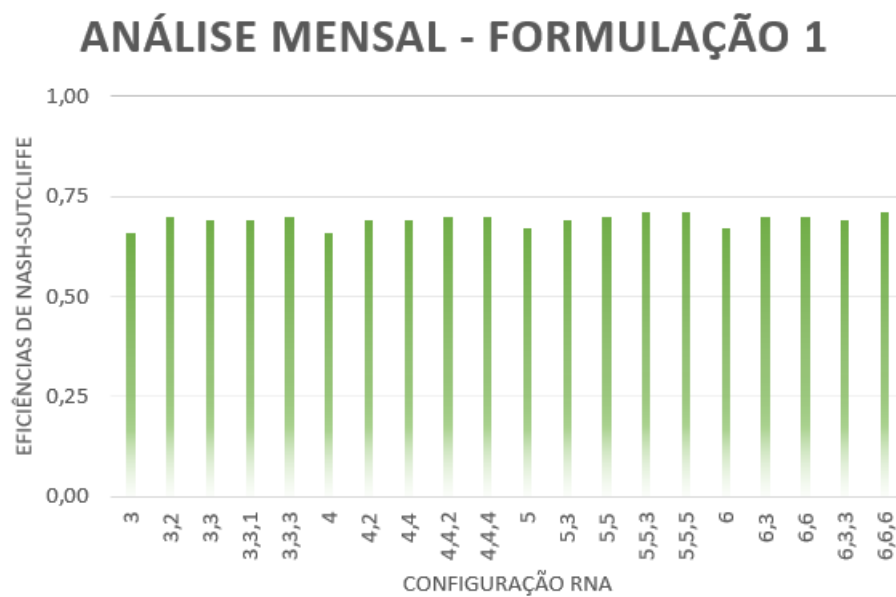
As Figuras 31, 32, 33 e 34 ilustram os resultados de eficiência de Nash-Sutcliffe referentes aos dados de validação das Formulações 1, 2, 3 e 4, respectivamente, para análise mensal via KNN. Além disso, no Apêndice G, é possível observar os coeficientes de desempenho para os dados de validação, calibração e completo nas Formulações 1, 2, 3 e 4.

Figura 26 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 5 com aplicação do KNN (análise anual)



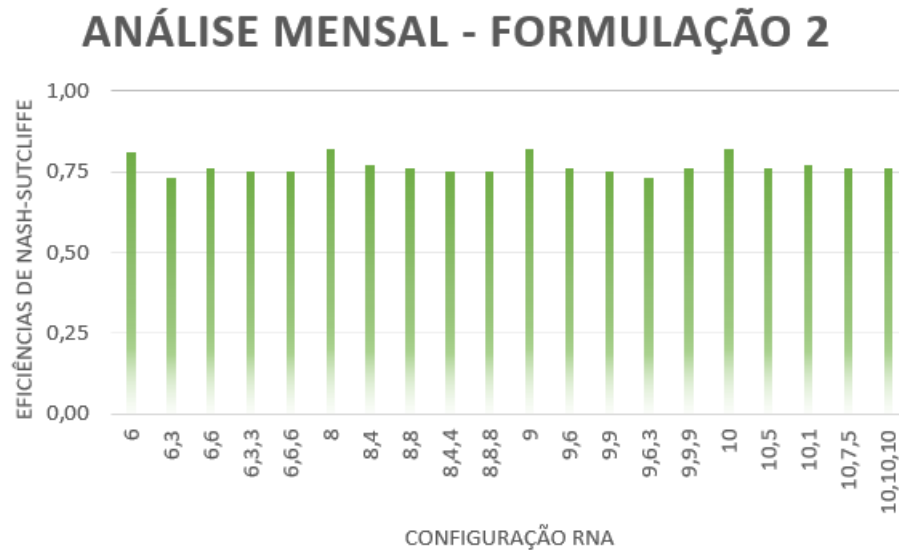
Fonte: O autor

Figura 27 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação de RNA (análise mensal)



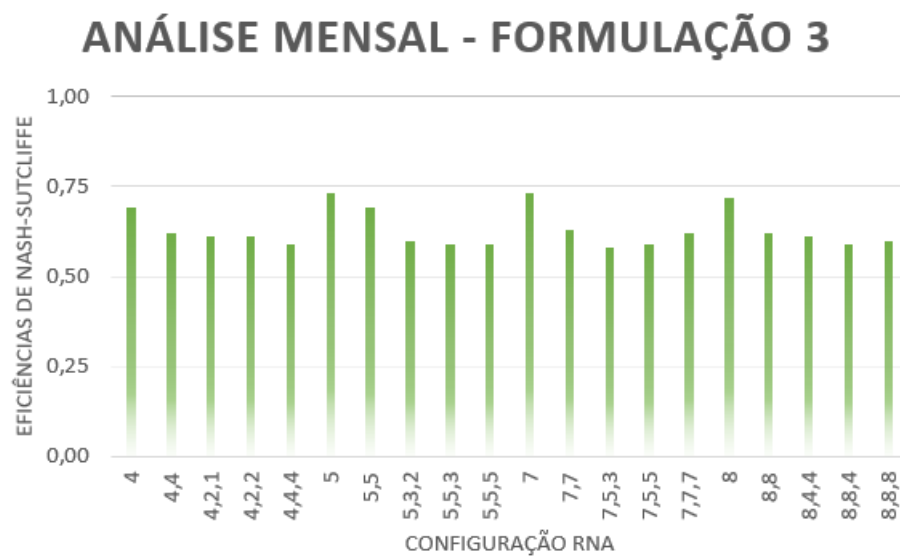
Fonte: O autor

Figura 28 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação de RNA (análise mensal)



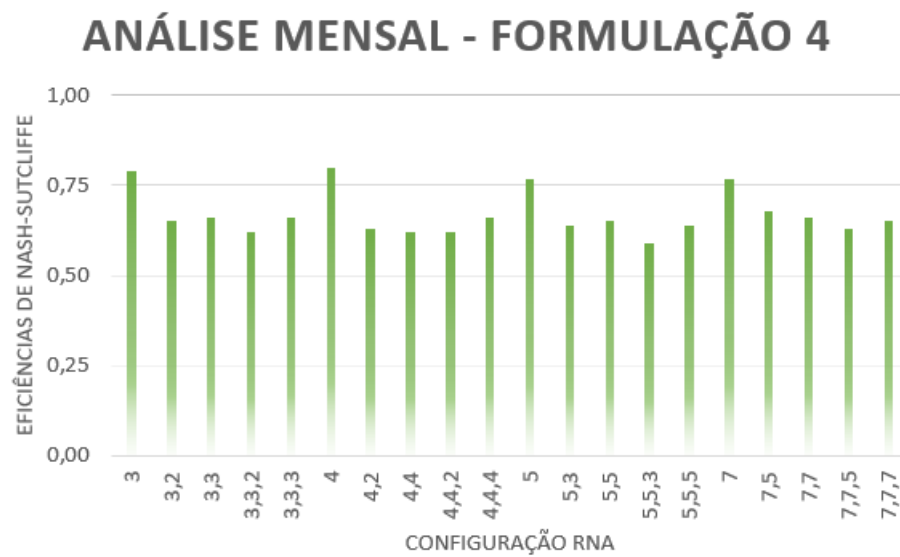
Fonte: O autor

Figura 29 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação de RNA (análise mensal)



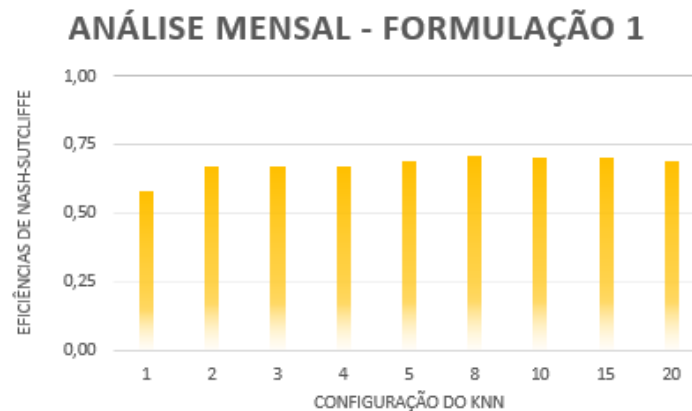
Fonte: O autor

Figura 30 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação de RNA (análise mensal)



Fonte: O autor

Figura 31 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 1 com aplicação do KNN (análise mensal)



Fonte: O autor

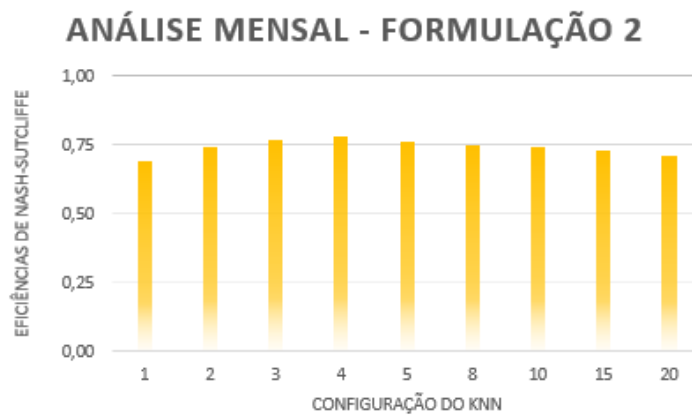
4.2 Discussão dos Resultados

4.2.1 Análise Anual

Os resultados com os melhores valores de coeficiente de Nash-Sutcliffe para os dados de validação das modelagens RNA e KNN para análise anual são listados a seguir:

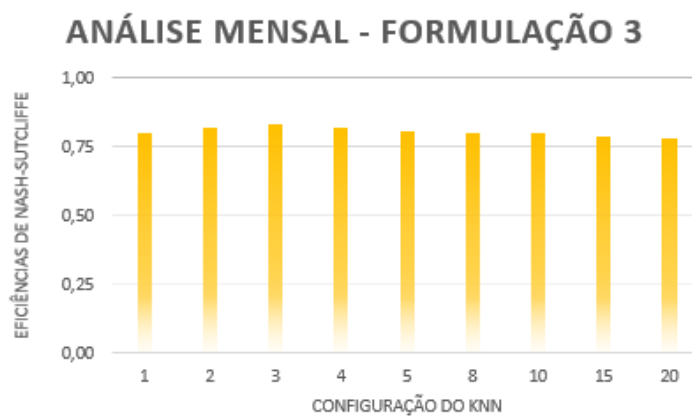
- Formulação 1: RNA-5,3,2 ($E = 0,02$); 10-NN ($E = 0,11$);
- Formulação 2: RNA-2,2,2 ($E = 0,00$); 10-NN ($E = 0,03$);
- Formulação 3: RNA-5,3,2 e RNA-3,2,2 ($E = -0,13$); 10-NN ($E = 0,01$);
- Formulação 4: RNA-3,2,2 ($E = 0,06$); 10-NN ($E = 0,15$).

Figura 32 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 2 com aplicação do KNN (análise mensal)



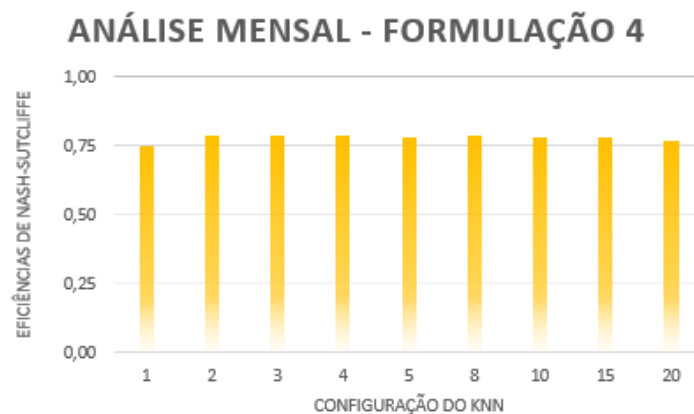
Fonte: O autor

Figura 33 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 3 com aplicação do KNN (análise mensal)



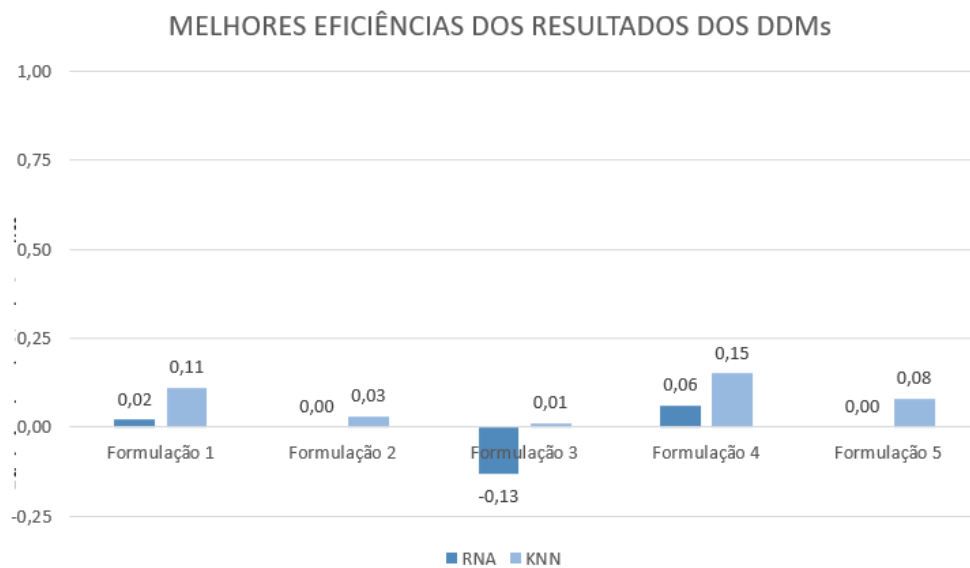
Fonte: O autor

Figura 34 – Eficiências de Nash-Sutcliffe nas configurações da Formulação 4 com aplicação do KNN (análise mensal)



Fonte: O autor

Figura 35 – Melhores índices de desempenho nas Formulações (análise anual)



Fonte: O autor

- Formulação 5: RNA-2,2,2 ($E = 0,00$); 10-NN ($E = 0,08$).

Na Figura 35 podem-se observar os melhores índices de desempenho em cada formulação. Percebe-se que os resultados foram todos insatisfatórios, pois obtiveram valores abaixo de 0,50.

4.2.2 Análise Mensal

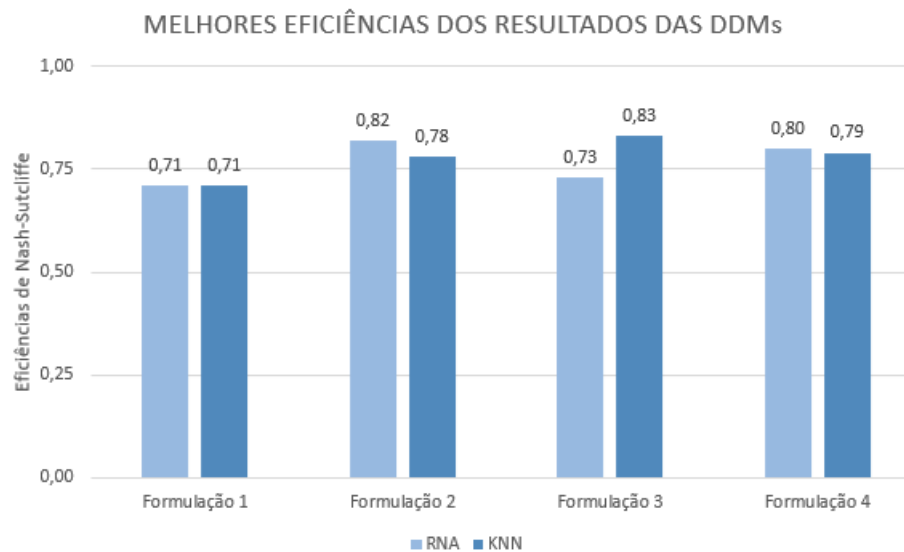
Os resultados com os melhores valores de coeficiente de Nash-Sutcliffe para os dados de validação das modelagens RNA e KNN para análise mensal são listados a seguir:

- Formulação 1: RNA-5,5,3, RNA-5,5,5 e RNA-6,6,6 ($E = 0,71$); 8-NN ($E = 0,71$);
- Formulação 2: RNA-8, RNA-9 e RNA-10 ($E = 0,82$); 4-NN ($E = 0,78$);
- Formulação 3: RNA-5 e RNA-7 ($E = 0,73$); 3-NN ($E = 0,83$);
- Formulação 4: RNA-4 ($E = 0,80$); 2-NN, 3-NN, 4-NN e 5-NN ($E = 0,79$).

Algumas formulações acima obtiveram mais de uma configuração como melhor eficiência nos resultados. Na Figura 36 podem-se observar os melhores índices de desempenho em cada formulação. Percebe-se que a Formulação 2 obteve melhor desempenho na RNA e a Formulação 3 obteve performance superior no KNN.

É possível salientar que os resultados de um modo geral foram classificados como bom ou muito bom, conforme a Seção 3.4. Para o modelo RNA, é notório que o melhor desempenho foi na Formulação 2 ($E = 0,82$), classificado como muito bom. Porém, a Formulação 4 ($E = 0,80$) também apresentou um resultado muito bom. Outrossim, as formulações 1 ($E = 0,71$) e 3 ($E = 0,73$) obtiveram resultados considerados bons. Em contrapartida, o modelo KNN teve seu

Figura 36 – Melhores índices de desempenho nas Formulações (análise mensal)



Fonte: O autor

melhor desempenho na Formulação 3 ($E = 0,83$), com classificação considerada como muito boa. As formulações 2 ($E = 0,78$) e 4 ($E = 0,79$) também obtiveram resultados com eficiências classificadas como muito boas. Além disso, a Formulação 1 ($E = 0,71$) obteve um resultado igual ao RNA, considerado como bom.

Como mencionado na Seção 4.1.2, os modelos obtiveram resultados muito insatisfatórios para a análise anual, apenas com as vazões. Assim sendo, na Formulação 1, para análise mensal obtiveram-se resultados bons, com eficiência de 71% para ambos RNA e KNN, que se torna uma opção viável para se trabalhar com dados de vazão como atributos de entradas, sem necessidade do uso dos dados de precipitação. Então, se tornou possível obter um resultado bom com uma modelagem mais simplificada, para ambos os modelos RNA e KNN.

A Formulação 3 com o KNN obteve o melhor resultado para modelagem de chuva vs. vazão, com o resultado de 83% para o índice de eficiência para validação, classificado como muito bom. Na RNA a formulação apresentou um resultado classificado como bom, porém com índice de eficiência menor de 73%.

A Formulação 2 com RNA obteve o melhor resultado na modelagem de chuva-vazão, porém só com os atributos de precipitação, com eficiência de 82% classificada como muito boa. No KNN, a formulação obteve um resultado muito bom também, com eficiência próxima do resultado da RNA, com 78%.

A Formulação 4 obteve resultados muito bons com RNA e KNN, com eficiências próximas, 80% e 79%, respectivamente.

Conclusões

A quantidade de água à afluir aos reservatórios pode ser prevista através da modelagem chuva-vazão. Esta pesquisa foi realizada para verificar o processo de chuva vs. vazão na área estudada visto que, em pesquisas anteriores, foram obtidos resultados interessantes com os modelos guiados por dados aplicados em outros estudos de caso.

Não foi possível modelar satisfatoriamente o processo vazão-vazão em Sobradinho para prever a vazão total do próximo ano baseada nas vazões de anos passados. Foram realizadas diversas modelagens, com várias configurações de atributos (apenas vazão), porém, os resultados foram todos insatisfatórios. Por causa disso, como alternativa, foi realizada a análise mensal, com os dados de chuva e vazão.

Assim sendo, com os dados de precipitações e vazões mensais, observou-se que através do software Weka é possível realizar a aplicação da modelagem chuva vs. vazão natural ao utilizar modelos baseados em redes neurais artificiais e no algoritmo dos k -vizinhos mais próximos tendo como estudo de caso o reservatório de Sobradinho, na bacia hidrográfica do rio São Francisco. Por meio da eficiência para os dados de validação, foi possível verificar nas quatro formulações propostas, através de várias simulações, que a modelagem fornece um bom resultado. Nas formulações 2 e 4 foram encontrados os melhores resultados para a RNA. Já na Formulação 3 encontrou-se a melhor performance para o KNN. E a Formulação 1 teve mesmo desempenho em ambos modelos. Em suma, a Formulação 3, composta por atributos de chuva e vazão, obteve a melhor eficiência no KNN, durante análise mensal. Entretanto, a Formulação 2, com apenas atributos de chuva de até cinco meses passados, obteve a melhor eficiência na RNA.

Para estudos futuros, uma sugestão é utilizar outras estratégias de DDM para o mesmo tipo de modelagem como forma de comparação. Adicionalmente, uma vez que a modelagem vazão-vazão em escala anual realizada inicialmente nesta pesquisa não se mostrou satisfatória, é interessante realizar uma investigação mais abrangente sobre o tema de forma a tentar buscar uma forma de modelagem que possa obter boas eficiências na previsão de vazões futuras sem a necessidade de atributos passados de chuva.

Referências

- AKBARI, M.; OVERLOOP, P. J. v.; AFSHAR, A. Clustered k nearest neighbor algorithm for daily inflow forecasting. *Water Resources Management*, Springer, v. 25, n. 5, p. 1341–1357, 2011.
- ALCÂNTARA, A. d. *Estudo de ferramentas baseadas no Weka para mineração de dados distribuída em ambiente de grid*. [S.l.], 2012.
- ALMEIDA, L.; SERRA, J. C. V. Modelos hidrológicos, tipos e aplicações mais utilizadas. *Revista da FAE*, v. 20, n. 1, p. 129–137, 2017.
- ALMEIDA, T. F. et al. Análise das precipitações totais anuais no alto e médio São Francisco e sua correlação com a vazão natural. In: Comitê da Bacia Hidrográfica do Rio São Francisco (CBHSF). *I Simpósio da Bacia Hidrográfica do Rio São Francisco*. Juazeiro – BA, 2016.
- BALTOKOSKI, V. et al. Calibração de modelo para simulação de vazão e fósforo total, nas sub-bacias dos Rios Conrado e Pinheiro - Rio Branco (PR). *Revista Brasileira de Ciência e Solo*, v. 34, n. 1, p. 253–261, 2010.
- CELESTE, A.; CHAVES, V. S. Avaliação de algoritmos de otimização e funções objetivo para calibração automática do modelo chuva-vazão tank model. *Ciência e Natura*, v. 36, n. 3, p. 527–537, 2014.
- CELESTE, A. B.; MENEZES, T.; FONTES, V. Modelagem do processo chuva-vazão pelo tank model e por redes neurais artificiais na bacia do rio Japarutuba, Sergipe. *Scientia Plena*, v. 10, n. 7, 2014.
- DANTAS, L. A. M. *Aplicação de redes neurais artificiais e do algoritmo dos k-vizinhos mais próximos para modelar o processo chuva-vazão na Bacia do Rio Piauítinga (SE)*. São Cristóvão – SE, 2022. 48 f. Trabalho de Conclusão de Curso (TCC) – Departamento de Engenharia Civil. Universidade Federal de Sergipe.
- FARIAS, C. A. S.; SANTOS, C. A. G.; CELESTE, A. B. Rede neural com atraso de tempo aplicada à previsão mensal de vazões na bacia hidrográfica do rio Piancó. In: Associação Brasileira de Recursos Hídricos (ABRH). *Proc. X Simpósio de Recursos Hídricos do Nordeste*. Fortaleza, 2010.
- FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. *Ciência da Informação*, SciELO Brasil, v. 35, p. 25–30, 2006.
- FRANK, E.; HALL, M. A.; WITTEN, I. H. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Fourth edition. [S.l.]: Morgan Kaufmann, 2016.
- GONÇALVES, C. U.; OLIVEIRA, C. F. de. Rio São Francisco: as águas correm para o mercado. *Boletim Goiano de Geografia*, Universidade Federal de Goiás, v. 29, n. 2, p. 113–125, 2009.
- LIMA, R. G. S. *Modelo chuva-vazão baseado na aplicação do método dos resíduos ponderados ao balanço hídrico*. São Cristóvão – SE, 2022. 33 f. Trabalho de Conclusão de Curso (TCC) – Departamento de Engenharia Civil. Universidade Federal de Sergipe.

- MACHADO, F. W. *Modelagem chuva-vazão mensal utilizando redes neurais artificiais*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Universidade Federal do Paraná, 2005.
- MARTINS, D. d. M. et al. Impactos da construção da usina hidrelétrica de sobradinho no regime de vazões no baixo são francisco. *Revista Brasileira de Engenharia Agrícola e Ambiental*, SciELO Brasil, v. 15, p. 1054–1061, 2011.
- MENDES, L. A. *O impacto dos usos consuntivos na operação de sistemas de reservatórios para produção de energia elétrica*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2012.
- SCHEIDT, F. A.; BRUNETTO, M. A. de C. Modelagem chuva-vazão utilizando redes neurais artificiais e algoritmos genéticos. In: *Congresso da Sociedade Brasileira de Computação*. [S.l.: s.n.], 2011. v. 31.
- SOLOMATINE, D. P.; MASKEY, M.; SHRESTHA, D. L. Instance-based learning compared to other data-driven methods in hydrological forecasting. *Hydrological Processes: An International Journal*, Wiley Online Library, v. 22, n. 2, p. 275–287, 2008.
- SOUSA, F.; SOUSA, W. Rede neural aplicada ao processo chuva-vazão da bacia hidrográfica de Caraúbas–PB. In: *Congresso Brasileiro de Meteorologia*. [S.l.: s.n.], 2006. v. 14.
- SVANIDZE, G. *Mathematical modeling of hydrologic series (for hydroelectric and water resources computations)*. Fort Collins, Colorado: Water Resources Publications, 1980. ISBN 9780918334329.
- VIEGAS FILHO, J. S. et al. MODHAC - Modelo hidrológico auto-calibrável, versão 98 para Windows. In: *Anais do Congresso Brasileiro de Engenharia Agrícola*. Pelotas: [s.n.], 1999.
- VILANOVA, R. S. *Redes neurais artificiais aplicadas a simulação de vazão na bacia Hidrográfica do rio Itapemirim-ES*. Jerônimo Monteiro - ES, 2017. 69 f. Dissertação de Mestrado – Programa de Pós-Graduação em Ciências Florestais. Universidade Federal do Espírito Santo.

Apêndices

APÊNDICE A – Eficiências de Nash-Sutcliffe das formulações em RNA para análise anual

Configuração	Formulação 1 - Análise Anual			Formulação 2 - Análise Anual		
	Completo	Calibração	Validação	Completo	Calibração	Validação
RNA-2	0,30	0,45	-0,04	0,11	0,44	-0,21
RNA-2,2	0,29	0,44	-0,04	0,27	0,46	-0,16
RNA-2,2,2	0,00	-0,01	0,00	0,00	-0,01	0,00
RNA-3	0,30	0,49	-0,13	0,15	0,46	-0,53
RNA-3,2	0,28	0,43	-0,05	0,26	0,46	-0,18
RNA-3,2,2	0,00	-0,01	0,00	0,22	0,41	-0,21
RNA-5	0,28	0,47	-0,16	0,21	0,44	-0,31
RNA-5,3	0,23	0,39	-0,11	0,26	0,45	-0,19
RNA-5,3,2	0,29	0,41	0,02	0,22	0,41	-0,22

Configuração	Formulação 3 - Análise Anual			Formulação 4 - Análise Anual		
	Completo	Calibração	Validação	Completo	Calibração	Validação
RNA-2	0,23	0,50	-0,36	0,26	0,38	-0,02
RNA-2,2	0,08	0,36	-0,55	0,27	0,48	-0,21
RNA-2,2,2	-0,06	-0,03	-0,13	0,02	0,01	0,01
RNA-3	0,24	0,51	-0,36	0,34	0,53	-0,07
RNA-3,2	0,14	0,49	-0,64	0,27	0,48	-0,20
RNA-3,2,2	0,02	0,24	-0,48	0,30	0,40	0,06
RNA-5	0,26	0,48	-0,24	0,35	0,53	-0,05
RNA-5,3	0,06	0,33	-0,54	0,27	0,49	-0,23
RNA-5,3,2	-0,02	0,02	-0,13	0,27	0,40	-0,01

Configuração	Formulação 5 - Análise Anual		
	Completo	Calibração	Validação
RNA-2	0,24	0,41	-0,21
RNA-2,2	0,29	0,48	-0,13
RNA-2,2,2	0,00	-0,01	0,00
RNA-3	0,29	0,49	-0,15
RNA-3,2	0,27	0,38	-0,29
RNA-3,2,2	0,24	0,42	-0,18
RNA-5	0,25	0,47	-0,24
RNA-5,3	0,20	0,41	-0,26
RNA-5,3,2	0,23	0,42	-0,19

APÊNDICE B – Eficiências de Nash-Sutcliffe na Formulação 1 em RNA para análise mensal

Configuração	Formulação 1 - Análise Mensal		
	Completo	Calibração	Validação
RNA-3	0,63	0,62	0,66
RNA-3,2	0,71	0,71	0,70
RNA-3,3	0,70	0,71	0,69
RNA-3,3,1	0,69	0,69	0,69
RNA-3,3,3	0,72	0,72	0,70
RNA-4	0,63	0,62	0,66
RNA-4,2	0,70	0,70	0,69
RNA-4,4	0,71	0,72	0,69
RNA-4,4,2	0,71	0,71	0,70
RNA-4,4,4	0,72	0,72	0,70
RNA-5	0,68	0,69	0,67
RNA-5,3	0,70	0,71	0,69
RNA-5,5	0,71	0,71	0,70
RNA-5,5,3	0,71	0,72	0,71
RNA-5,5,5	0,72	0,73	0,71
RNA-6	0,70	0,72	0,67
RNA-6,3	0,70	0,70	0,70
RNA-6,6	0,71	0,71	0,70
RNA-6,3,3	0,70	0,70	0,69
RNA-6,6,6	0,72	0,73	0,71

APÊNDICE C – Eficiências de Nash-Sutcliffe na Formulação 2 em RNA para análise mensal

Configuração	Formulação 2 - Análise Mensal		
	Completo	Calibração	Validação
RNA-6	0,80	0,80	0,81
RNA-6,3	0,71	0,69	0,73
RNA-6,6	0,75	0,74	0,76
RNA-6,3,3	0,74	0,73	0,75
RNA-6,6,6	0,74	0,73	0,75
RNA-8	0,82	0,81	0,82
RNA-8,4	0,75	0,74	0,77
RNA-8,8	0,75	0,74	0,76
RNA-8,4,4	0,73	0,72	0,75
RNA-8,8,8	0,74	0,73	0,75
RNA-9	0,81	0,81	0,82
RNA-9,6	0,75	0,74	0,76
RNA-9,9	0,74	0,73	0,75
RNA-9,6,3	0,70	0,69	0,73
RNA-9,9,9	0,74	0,73	0,76
RNA-10	0,81	0,81	0,82
RNA-10,5	0,75	0,74	0,76
RNA-10,10	0,75	0,74	0,77
RNA-10,7,5	0,74	0,73	0,76
RNA-10,10,10	0,74	0,73	0,76

APÊNDICE D – Eficiências de Nash-Sutcliffe na Formulação 3 em RNA para análise mensal

Configuração	Formulação 3 - Análise Mensal		
	Completo	Calibração	Validação
RNA-4	0,66	0,64	0,69
RNA-4,4	0,59	0,57	0,62
RNA-4,2,1	0,57	0,56	0,61
RNA-4,2,2	0,57	0,56	0,61
RNA-4,4,4	0,55	0,53	0,59
RNA-5	0,70	0,68	0,73
RNA-5,5	0,66	0,64	0,69
RNA-5,3,2	0,57	0,55	0,60
RNA-5,5,3	0,56	0,54	0,59
RNA-5,5,5	0,56	0,54	0,59
RNA-7	0,70	0,68	0,73
RNA-7,7	0,59	0,57	0,63
RNA-7,5,3	0,55	0,53	0,58
RNA-7,5,5	0,55	0,54	0,59
RNA-7,7,7	0,59	0,57	0,62
RNA-8	0,69	0,68	0,72
RNA-8,8	0,58	0,57	0,62
RNA-8,4,4	0,58	0,56	0,61
RNA-8,8,4	0,56	0,54	0,59
RNA-8,8,8	0,57	0,55	0,60

APÊNDICE E – Eficiências de Nash-Sutcliffe na Formulação 4 em RNA para análise mensal

Configuração	Formulação 4 - Análise Mensal		
	Completo	Calibração	Validação
RNA-3	0,75	0,73	0,79
RNA-3,2	0,62	0,61	0,65
RNA-3,3	0,63	0,62	0,66
RNA-3,3,2	0,60	0,58	0,62
RNA-3,3,3	0,61	0,59	0,66
RNA-4	0,76	0,74	0,80
RNA-4,2	0,60	0,58	0,63
RNA-4,4	0,59	0,57	0,62
RNA-4,4,2	0,57	0,54	0,62
RNA-4,4,4	0,61	0,58	0,66
RNA-5	0,74	0,72	0,77
RNA-5,3	0,61	0,59	0,64
RNA-5,5	0,61	0,58	0,65
RNA-5,5,3	0,55	0,53	0,59
RNA-5,5,5	0,61	0,59	0,64
RNA-7	0,73	0,72	0,77
RNA-7,5	0,64	0,63	0,68
RNA-7,7	0,63	0,61	0,66
RNA-7,7,5	0,60	0,58	0,63
RNA-7,7,7	0,62	0,61	0,65

APÊNDICE F – Eficiências de Nash-Sutcliffe das formulações em KNN para análise anual

Configuração	Análise Anual					
	Formulação 1			Formulação 2		
	Completo	Calibração	Validação	Completo	Calibração	Validação
1-NN	0,52	1,00	-0,52	0,39	1,00	-0,94
2-NN	0,47	0,73	-0,11	0,40	0,70	-0,26
3-NN	0,36	0,58	-0,14	0,37	0,60	-0,13
4-NN	0,37	0,53	0,00	0,34	0,52	-0,05
5-NN	0,34	0,51	-0,05	0,32	0,50	-0,08
10-NN	0,35	0,45	0,11	0,33	0,47	0,03

Configuração	Formulação 3			Formulação 4		
	Completo	Calibração	Validação	Completo	Calibração	Validação
1-NN	0,52	1,00	-0,53	0,56	1,00	-0,01
2-NN	0,42	0,71	-0,22	0,43	0,69	-0,14
3-NN	0,30	0,55	-0,26	0,38	0,57	-0,05
4-NN	0,33	0,53	-0,13	0,40	0,57	0,01
5-NN	0,30	0,48	-0,10	0,38	0,50	0,12
10-NN	0,29	0,41	0,01	0,36	0,45	0,15

Configuração	Análise Anual		
	Formulação 5		
	Completo	Calibração	Validação
1-NN	0,46	1,00	-0,72
2-NN	0,29	0,70	-0,62
3-NN	0,27	0,55	-0,34
4-NN	0,27	0,49	-0,21
5-NN	0,31	0,47	-0,07
10-NN	0,35	0,47	0,08

APÊNDICE G – Eficiências de Nash-Sutcliffe nas Formulações 1, 2, 3 e 4 no KNN para análise mensal

Análise Mensal						
Configuração	Formulação 1			Formulação 2		
	Completo	Calibração	Validação	Completo	Calibração	Validação
1–NN	0,86	1,00	0,58	0,89	1,00	0,69
2–NN	0,81	0,88	0,67	0,84	0,89	0,74
3–NN	0,77	0,82	0,67	0,82	0,85	0,77
4–NN	0,76	0,80	0,67	0,80	0,82	0,78
5–NN	0,76	0,79	0,69	0,79	0,80	0,76
8–NN	0,75	0,77	0,71	0,76	0,76	0,75
10–NN	0,74	0,77	0,70	0,75	0,75	0,74
15–NN	0,73	0,75	0,70	0,75	0,75	0,73
20–NN	0,72	0,74	0,69	0,73	0,73	0,71

Análise Mensal						
Configuração	Formulação 3			Formulação 4		
	Completo	Calibração	Validação	Completo	Calibração	Validação
1–NN	0,93	1,00	0,80	0,92	1,00	0,75
2–NN	0,89	0,93	0,82	0,88	0,92	0,79
3–NN	0,88	0,90	0,83	0,85	0,88	0,79
4–NN	0,86	0,88	0,82	0,84	0,86	0,79
5–NN	0,85	0,87	0,81	0,83	0,86	0,78
8–NN	0,83	0,85	0,80	0,82	0,83	0,79
10–NN	0,82	0,84	0,80	0,81	0,83	0,78
15–NN	0,81	0,82	0,79	0,80	0,81	0,78
20–NN	0,79	0,80	0,78	0,78	0,79	0,77