

Modelos de Aprendizado na Predição de Nível de Rios para Detecção de Falhas em Estações Hidrotelemétricas

Roni Rodrigues¹, Claudia Aparecida Martins¹

¹Instituto de Computação – Universidade Federal de Mato Grosso (UFMT)
CEP: 78060-900 – Cuiabá – MT – Brasil

roni.rodrigues@live.com, claudia@ic.ufmt.br

Abstract. *This study conducted a comparative analysis of the Prophet, Random Forest, Support Vector Regression (SVR), and XGBoost models for hydrometric level forecasting aimed at sensor fault detection. Using data from the SNIRH (2022–2025) for stations along the Rio M, the models were updated daily with the mean of observed values and evaluated using standard metrics: MAE, MSE, RMSE, MAPE, and R². The results showed that XGBoost achieved the best performance, more effectively capturing variations. In this context, machine learning models outperformed the others, demonstrating strong potential to enhance the predictive maintenance of hydrotelemetering stations.*

Resumo. *Este trabalho realizou uma análise comparativa dos modelos Prophet, Random Forest, Suport Vector Regression (SVR), XGBoost na previsão de níveis hidrométricos para detecção de falhas em sensores. Utilizando dados do SNIRH (2022-2025) das estações do Rio M, os modelos foram retroalimentados diariamente com a média dos valores observados e avaliados pelas métricas padrão MAE, MSE, RMSE, MAPE e R². Os resultados mostraram que o XGBoost teve melhor desempenho, capturando melhor as variações. Neste trabalho, os modelos de aprendizado de máquina tiveram um melhor desempenho para essa aplicação, com potencial para aprimorar a manutenção preditiva de estações hidrotelemétricas.*

1. Introdução

O monitoramento hidrológico desempenha um papel fundamental na gestão sustentável dos recursos hídricos, possibilitando a previsão de eventos críticos, como enchentes e secas, e fornecendo informações essenciais para a tomada de decisões estratégicas [ANA 2016]. No Brasil, o Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH) é a principal plataforma de coleta e armazenamento desses dados, utilizando estações hidrotelemétricas distribuídas pelo território nacional [SNIRH 2021]. Essas estações operam de forma automatizada, transmitindo dados de nível e precipitação em tempo real para os centros de monitoramento, como a Agência Nacional de Águas e Saneamento Básico (ANA) e instituições parceiras [EPAGRI/CIRAM-ANA 2024].

Entretanto, a confiabilidade dessas medições, de coleta e armazenamento, pode ser comprometida por falhas nos sensores, interferências ambientais e erros na transmissão dos dados [ANA 2021]. O diagnóstico precoce de inconsistências nesses registros é crucial para evitar a propagação de erros em análises hidrológicas e evitar ações equivocadas

na gestão da água. Atualmente, a identificação de falhas é baseada em verificações manuais e na comparação com valores históricos, um processo que pode ser demorado e sujeito a erros humanos [SNIRH 2021].

Nesse contexto, técnicas de predição de séries temporais surgem como uma alternativa promissora para identificar padrões anômalos e antecipar possíveis falhas nos sensores. Modelos estatísticos e de aprendizado de máquina podem auxiliar na detecção de desvios nos níveis monitorados, permitindo a implementação de estratégias proativas de manutenção [ANA 2022]. Entre os métodos mais utilizados, destacam-se o Prophet, um modelo baseado em decomposição aditiva desenvolvido pela Meta, e o Random Forest, um algoritmo de aprendizado de máquina baseado em múltiplas árvores de decisão.

O objetivo deste artigo é comparar o desempenho dos modelos Prophet, Random Forest, SVR e GXBoost na previsão de níveis hidrológicos, avaliando sua precisão na detecção de falhas em sensores hidrotelemétricos. Para isso, foram utilizados dados históricos do SNIRH, cobrindo o período de 2022 a 2025, com realimentação contínua dos modelos para aprimorar suas previsões. A análise será estruturada com base na metodologia do *Cross Industry Standard Process for Data Mining* (CRISP-DM), um modelo amplamente utilizado para o desenvolvimento de soluções baseadas em ciência de dados [Shearer 2000].

O artigo está organizado em seis seções: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 discute os fundamentos do monitoramento hidrológico e dos modelos aplicados; a Seção 4 detalha a metodologia, desde o pré-processamento até a avaliação; a Seção 5 apresenta e analisa os resultados obtidos; e a Seção 6 traz as conclusões e propostas de continuidade da pesquisa.

2. Materiais e métodos

O monitoramento hidrológico desempenha um papel fundamental na gestão dos recursos hídricos, fornecendo dados essenciais para a prevenção de desastres, planejamento de infraestrutura e conservação ambiental [ANA 2016]. No Brasil, a coleta de informações hidrológicas é coordenada pelo SNIRH, que agrega dados de estações hidrotelemétricas espalhadas pelo território nacional [SNIRH 2021].

As estações hidrotelemétricas utilizam sensores para medir variáveis como nível da água, vazão e precipitação. No entanto, essas medições estão sujeitas a falhas decorrentes de problemas mecânicos, calibração inadequada ou interferências ambientais. Erros comuns incluem níveis com baixa oscilação, inconsistências totais, oscilações anormais e leituras fora dos limites históricos [EPAGRI/CIRAM-ANA 2024]. A detecção e correção desses problemas exigem ferramentas avançadas de análise de dados.

A previsão de séries temporais é amplamente utilizada em estudos hidrológicos para identificar padrões, tendências e anomalias [Manandhar et al. 2024]. Neste trabalho, foram avaliados quatro modelos representativos: Prophet, voltado à captura de sazonalidades e tendências [Manandhar et al. 2024]; *Random Forest*, baseado em árvores de decisão [Flyrank 2023]; SVR, eficaz em relações não lineares [Sharma and Goel 2024]; e XGBoost, modelo baseado em *boosting* com alto desempenho em regressão [Niazkar et al. 2024]. A escolha refletiu a diversidade de abordagens e sua aplicabilidade em contextos hidrológicos recentes.

2.1. Modelos de predição

O Prophet é um modelo aditivo desenvolvido pela Meta, projetado para lidar com séries temporais de maneira robusta e interpretável [Manandhar et al. 2024]. Ele decompõe os dados em três componentes principais: tendência, sazonalidade e efeitos externos. Essa estrutura torna o modelo altamente eficiente para séries temporais com padrões sazonais e eventos específicos, permitindo ajustes automáticos mesmo em séries com intervalos irregulares.

Estudos recentes demonstram que o Prophet tem sido aplicado com sucesso em previsão de demanda energética e hidrológica, apresentando bons resultados quando há padrões claros e estabilidade nos dados [Barzegar et al. 2021]. No entanto, sua capacidade preditiva pode ser limitada em cenários com alta variabilidade e mudanças abruptas.

O *Random Forest* (RF) é um algoritmo de aprendizado de máquina baseado em um conjunto de árvores de decisão construídas de forma aleatória, combinando os resultados de cada árvore para gerar uma previsão mais estável e precisa. Sua estrutura baseada na combinação de múltiplas árvores de decisão, conhecida como método de agregação, confere ao modelo uma alta robustez contra *overfitting* e sensibilidade a ruídos, tornando-o adequado para problemas com variabilidade significativa nos dados [Flyrank 2023].

Embora originalmente desenvolvido para tarefas de classificação e regressão em dados tabulares, o RF pode ser adaptado à modelagem de séries temporais por meio da introdução de variáveis temporais, como hora, dia, mês e ano, e defasagens históricas (*lags*) dos dados de entrada. Essa abordagem permite ao modelo capturar padrões sazonais, tendências locais e interações complexas entre variáveis. Trabalhos recentes, como em [Jaimes Campos and López Zúñiga 2021], destacam o potencial de *Random Forest* na previsão de demanda em séries temporais irregulares, enquanto o trabalho em [Manandhar et al. 2024] aponta desempenho superior em relação ao Prophet em cenários de alta flutuação.

Support Vector Regression (SVR) é uma técnica de aprendizado de máquina baseada no algoritmo de *Support Vector Machine* (SVM), adaptada para tarefas de regressão. Sua abordagem consiste em encontrar uma função que aproxime os dados de entrada dentro de uma margem de tolerância definida, buscando ao mesmo tempo minimizar a complexidade do modelo. Essa característica permite ao SVR lidar com relações não lineares entre variáveis, mesmo em conjuntos de dados com ruído ou comportamento irregular [Sharma and Goel 2024].

Aplicações na área de recursos hídricos, o SVR tem se mostrado eficaz na modelagem de séries temporais, como a previsão de vazões e níveis de água em rios e reservatórios. Conforme apresentado em [Sharma and Goel 2024], a técnica apresenta desempenho promissor em contextos hidrológicos, destacando-se pela capacidade de capturar padrões complexos de comportamento hidrodinâmico.

Já o *Extreme Gradient Boosting* (XGBoost) é um algoritmo de regressão baseado em árvores de decisão, que utiliza o princípio de *boosting* para construir modelos preditivos robustos a partir da combinação sequencial de múltiplos modelos fracos. Uma de suas principais vantagens está na eficiência computacional, aliada à capacidade de lidar com dados heterogêneos e com alta colinearidade entre variáveis [Niazkar et al. 2024].

Na engenharia de recursos hídricos, o XGBoost tem sido amplamente empregado em tarefas como previsão de vazão, qualidade da água e níveis hidrométricos, devido à sua alta acurácia e estabilidade preditiva. Segundo [Niazkar et al. 2024], a técnica apresenta vantagens significativas quando aplicada a séries temporais hidrológicas, superando modelos tradicionais em termos de desempenho e adaptabilidade.

2.2. Trabalhos Relacionados

A previsão de variáveis hidrológicas por meio de modelos de aprendizado de máquina tem sido tema recorrente na literatura recente, com destaque para sua aplicação em cenários com alta variabilidade e complexidade temporal. Dentre as abordagens exploradas, o modelo Prophet vem sendo utilizado para detectar tendências e padrões sazonais em séries temporais. Em [Manandhar et al. 2024] o Prophet foi avaliado comparativamente a *Random Forest* e outros modelos para previsão de carga elétrica, destacando sua eficácia na identificação de componentes sazonais, embora com limitações na detecção de variações abruptas.

No contexto hidrológico, modelos baseados em SVR mostraram desempenho promissor. No trabalho de [Sharma and Goel 2024] foi aplicado SVR para prever a vazão do rio no reservatório Tehri, na Índia, utilizando variáveis como precipitação, temperatura, cobertura de neve e descarga. No trabalho foi apresentado que SVR é capaz de capturar com precisão as variações da série temporal em escalas diárias, superando modelos tradicionais, especialmente quando calibrada com validação cruzada e *kernels* adequados.

O XGBoost também tem se consolidado como uma das técnicas mais eficazes em tarefas de previsão hidrológica. No trabalho de [Niazkar et al. 2024] foi apresentada uma revisão sistemática sobre o uso do XGBoost em engenharia de recursos hídricos, apontando que 74% dos estudos analisados obtiveram os melhores resultados com esse algoritmo, inclusive em aplicações como previsão de qualidade da água, gerenciamento de aquíferos e modelagem de reservatórios. Sua eficiência computacional e capacidade de lidar com dados esparsos o tornam especialmente adequado para séries temporais complexas.

Nestes trabalhos, foi evidenciada uma tendência crescente por abordagens baseadas em aprendizado de máquina, em substituição ou complementação com modelos estatísticos tradicionais, visando maior precisão e adaptabilidade em cenários hidrológicos diversos.

3. Metodologia

Neste trabalho, foram investigados o comportamento do nível hidrométrico em estações situadas no Rio M [omitido para revisão], utilizando dados do SNIRH no período de 2022 a 2025. Para isso, como comentado, foram aplicados quatro modelos de previsão: Prophet, RF, SVR e XGBoost. Uma análise comparativa dos métodos visa testar a eficácia dos diferentes paradigmas de modelagem, desde métodos estatísticos até algoritmos de aprendizado de máquina, com o objetivo de identificar quais modelos apresentam melhor desempenho em diferentes condições hidrológicas.

Essa análise poderá ser combinada com técnicas de reconhecimento de padrões operacionais, como detecção de anomalias, variações abruptas fora dos limites de secas e cheias históricas ou comportamento atípico, para o desenvolvimento, posteriormente, de

uma ferramenta automatizada de análise de falhas em estações hidrométricas, conforme diretrizes de monitoramento da ANA [ANA 2016, SNIRH 2021].

Seguindo o modelo CRISP-DM [Shearer 2000], a metodologia foi estruturada em três grandes fases: pré-processamento, modelagem e avaliação, conforme representado na Figura 1. Antes dessas etapas, o processo incluiu o entendimento do problema — voltado à detecção de falhas em sensores hidrológicos — e o entendimento dos dados, que envolveu a análise da qualidade e da estrutura das séries temporais extraídas do SNIRH. A fase de pré-processamento compreendeu subfases como coleta, limpeza, criação de variáveis temporais e defasagens. A modelagem abrangeu o treinamento dos modelos e a retroalimentação diária, enquanto a avaliação se baseou em métricas quantitativas. A fase de operação é representada pela implementação sequencial dessas etapas, com a retroalimentação contínua dos modelos. Além disso, utilizou-se inteligência artificial generativa para aprimorar a redação científica e refatorar o código, tornando o desenvolvimento mais claro e eficiente.

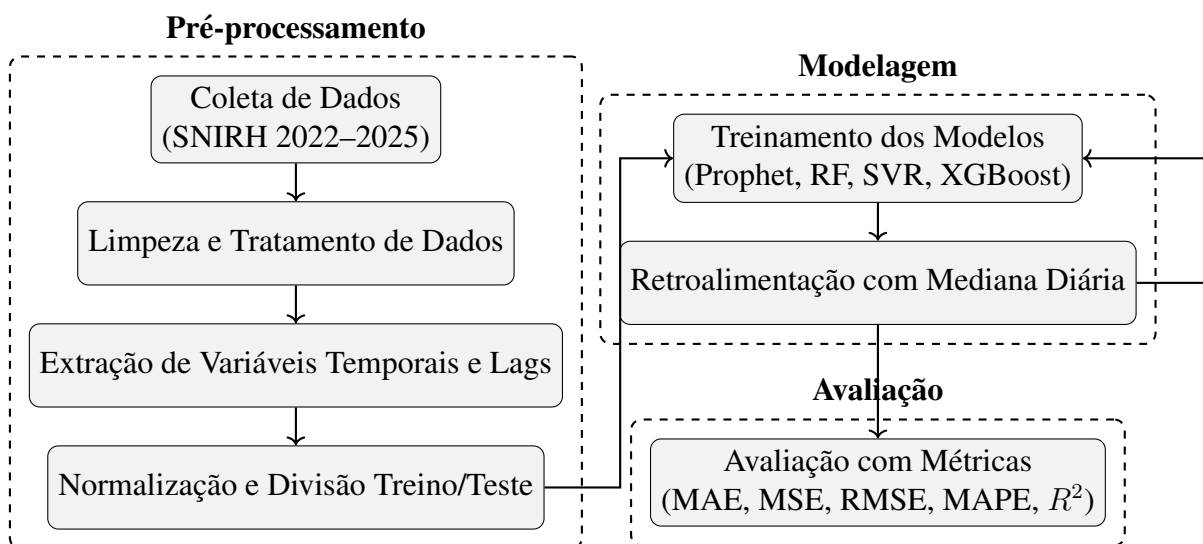


Figura 1. Fluxograma da metodologia adotada para previsão hidrométrica.

3.1. Pré-processamento dos Dados

O pré-processamento dos dados seguiu uma rotina padronizada para garantir qualidade, integridade e adequação das séries temporais ao treinamento dos modelos preditivos. Os dados horários de nível hidrométrico foram obtidos do SNIRH [SNIRH 2021] e abrangearam o período de 2022 a 2025. Inicialmente, as datas foram convertidas para o formato `datetime`, assegurando ordenação cronológica. Em seguida, foram tratados valores ausentes e leituras inconsistentes, por meio da conversão de entradas inválidas para `NaN`, remoção de linhas nulas e descarte de registros com valor zero, conforme orientações da ANA [ANA 2016], esses podem indicar falhas de sensor ou ausência de transmissão. Para melhorar os valores de entradas, foram extraídas variáveis temporais hora, dia, mês e ano, e adicionadas defasagens históricas (*lags*) de até seis instantes anteriores, permitindo aos modelos capturar padrões sazonais e de dependência temporal.

A normalização dos dados foi realizada com o `StandardScaler`, da biblioteca

`scikit-learn`¹, padronizando as variáveis de entrada para evitar distorções no treinamento dos modelos. A divisão entre conjuntos de treino e teste seguiu uma estratégia temporal (sem embaralhamento), com 80% dos dados utilizados para treinamento e 20% para validação. Todas as etapas foram implementadas utilizando bibliotecas amplamente reconhecidas na comunidade científica, como `pandas` para manipulação de dados tabulares e temporais, `NumPy` para operações numéricas e `scikit-learn` para pré-processamento e avaliação dos modelos. O desenvolvimento e execução dos experimentos foram realizados no ambiente colaborativo *Google Colab*², que oferece infraestrutura adequada para tarefas de ciência de dados.

3.2. Modelagem e Treinamento

A modelagem dos dados foi realizada de forma padronizada para os quatro modelos testados, com variações específicas conforme a estrutura exigida por cada técnica.

No Prophet, os dados foram agregados previamente em mediana diária e a coluna de nível foi renomeada para `y`, e a de data para `ds`, conforme exigido pela biblioteca. A cada iteração, foi gerada uma previsão para o próximo dia, e o valor real observado era incorporado à base para a etapa seguinte, promovendo uma retroalimentação progressiva com dados reais. Esse processo foi realizado sem ajuste automático de parâmetros.

No modelo RF, os dados foram normalizados com `StandardScaler` e divididos em 80% para o treinamento e 20% para teste (20%), sem embaralhamento, preservando a ordem temporal. A cada nova mediana diária observada, uma nova linha era construída com as variáveis temporais e as defasagens obtidas a partir dos últimos valores reais. Após a predição, o modelo era retreinado incrementalmente com essa nova amostra, incluindo a retroalimentação com os dados reais, como sugerido por [Barzegar et al. 2021].

O modelo SVR seguiu os mesmos passos do RF, ou seja, normalização com `StandardScaler`, entrada composta por variáveis temporais e *lags*, e retroalimentação diária com novos dados reais após cada previsão. O modelo foi retreinado a cada ciclo com a adição de uma nova amostra à base de treino, sem redefinir os dados anteriores. Já o XGBoost recebeu os dados normalizados e com defasagens, sendo retroalimentado diariamente com a mediana observada e reentrenado por completo com o `XGBRegressor`, mantendo os hiperparâmetros fixos. Embora essa estratégia seja viável para bases moderadas, em sistemas reais com grandes volumes de dados, pode-se adotar abordagens mais eficientes, como janelas deslizantes. Além do ganho técnico, essa retroalimentação contínua tem valor prático: permite detectar padrões incomuns que podem indicar falhas em sensores, apoiando decisões mais precisas na gestão hidrológica.

Em todos os casos, o valor real da mediana diária de nível foi utilizado para avaliar o erro de previsão, além de alimentar o modelo no passo seguinte, garantindo a consistência temporal da série durante o processo iterativo de predição. O uso de retroalimentação diária se baseou em estudos anteriores que demonstraram que a incorporação progressiva de dados reais melhora a precisão das previsões [Barzegar et al. 2021, Manandhar et al. 2024].

¹Biblioteca de aprendizado de máquina em Python. Disponível em: <https://scikit-learn.org>

²Ambiente gratuito baseado em nuvem para desenvolvimento de notebooks Jupyter. Disponível em: <https://colab.research.google.com>

3.3. Avaliação de Desempenho

Para avaliar a precisão dos modelos gerados, foram utilizadas métricas amplamente empregadas em previsões de séries temporais. Na Tabela 1 são apresentadas uma descrição das métricas utilizadas.

Tabela 1. Métricas de avaliação dos modelos

Métrica	Fórmula	Descrição
Erro Absoluto Médio (MAE)	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	Diferença média absoluta entre valores previstos e reais.
Erro Quadrático Médio (MSE)	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	Similar ao MAE, mas penaliza mais os erros maiores.
Raiz do Erro Quadrático Médio (RMSE)	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	Erro médio das previsões na mesma unidade dos dados.
Erro Percentual Absoluto Médio (MAPE)	$\frac{100}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right $	Erro médio expresso como porcentagem dos valores reais.
Coefficiente de Determinação (R^2)	$1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$	Grau de ajuste do modelo aos dados.

O comportamento dessas métricas permitiu avaliar qual modelo apresentou maior precisão na previsão do nível hidrométrico e sua viabilidade para aplicações futuras. Modelos com menores valores de MAE, MSE, RMSE e MAPE são preferíveis, pois indicam menor erro nas previsões. Já o coeficiente de determinação (R^2) deve ser o mais próximo possível de 1, pois valores negativos sugerem desempenho inferior à média histórica.

4. Resultados e Discussão

A avaliação dos modelos Prophet, RF, SVR e XGBoost foi realizada com base em séries temporais de nível d'água em três estações hidrológicas do Rio M, aqui denominadas de Rio Jusante, Rio Quilombo e Rio Casca. Para isso, foram utilizadas métricas clássicas de desempenho citadas — MAE, MSE, RMSE, MAPE e R^2 — aplicadas aos resultados suavizados de cada modelo. Os valores obtidos estão apresentados na Tabela 2.

Os resultados mostram que os modelos baseados em aprendizado de máquina apresentaram desempenho superior na previsão de níveis hidrométricos, destacando-se pela capacidade de capturar oscilações locais e eventos transitórios com maior precisão. Entre eles, RF, SVR e XGBoost apresentaram maior sensibilidade a variações rápidas nos dados, sendo mais indicados para aplicações em monitoramento preditivo. O modelo Prophet, apesar de resultados inferiores nas métricas quantitativas, pode complementar essas abordagens por sua habilidade em representar tendências de longo prazo e padrões sazonais.

O XGBoost foi o modelo com melhor desempenho geral entre todas as estações analisadas. Na estação Rio Jusante, obteve RMSE de 8.90 e R^2 de 0.94; em Rio Quilombo, os valores foram ainda mais expressivos, com RMSE de 5.37 e R^2 de 0.93. Na estação Rio Casca, também se destacou com RMSE de 7.54 e R^2 de 0.92. Esses resultados indicam alta precisão na previsão de eventos transitórios e variabilidade de curto prazo.

O modelo RF apresentou bom desempenho, especialmente nas estações Rio Quilombo e Rio Casca, alcançando R^2 de 0.91 e 0.92, respectivamente. Na estação Rio Jusante, obteve R^2 de 0.84, superando os demais modelos, exceto o XGBoost. Já o SVR, apesar de resultados um pouco inferiores, ainda mostrou desempenho satisfatório em Rio Quilombo ($R^2 = 0.84$), sendo uma opção viável para aplicações com menor exigência computacional.

Tabela 2. Métricas de avaliação dos modelos para as estações do Rio M

Estação	Modelo	MAE	MSE	RMSE	MAPE	R^2
Rio Jusante	Prophet	26.89	1606.77	40.08	7.44%	-0.31
	RF	10.66	195.38	13.98	3.14%	0.84
	XGBoost	5.91	79.25	8.90	1.71%	0.94
	SVR	10.11	321.19	17.92	2.72%	0.74
Rio Quilombo	Prophet	11.80	465.16	21.57	6.51%	-0.18
	RF	4.04	33.86	5.82	2.52%	0.91
	XGBoost	3.73	28.88	5.37	2.27%	0.93
	SVR	4.19	62.05	7.88	2.35%	0.84
Rio Casca	Prophet	11.75	989.15	31.45	7.49%	-0.36
	RF	5.89	56.62	7.52	2.62%	0.92
	XGBoost	5.40	56.78	7.54	2.36%	0.92
	SVR	7.76	216.80	14.72	3.0%	0.70

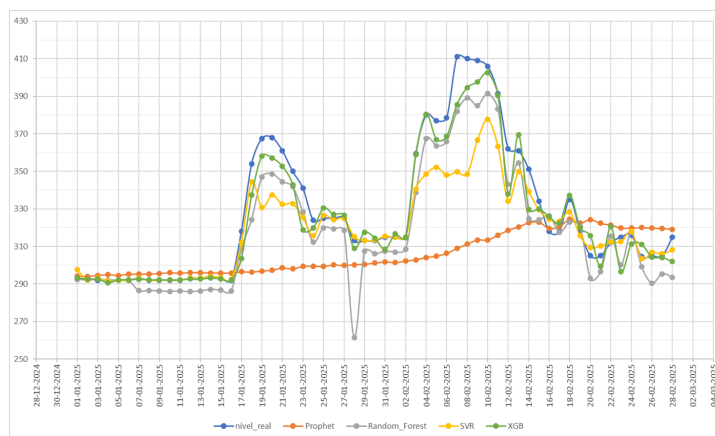


Figura 2. Previsão X Dados Reais: Rio Jusante (Prophet, RF, SVR, XGB).

Nas Figuras 2, 3 e 4 são apresentadas as curvas de previsão e os dados reais para cada estação Rio Jusante, Rio Quilombo, Rio Casca, respectivamente, ilustrando visualmente o desempenho dos modelos. Nota-se que o Prophet tende a suavizar excessivamente as variações, enquanto os modelos baseados em aprendizado de máquina se adaptam melhor às oscilações rápidas e eventos locais.

Notadamente, o XGBoost obteve os melhores desempenhos gerais, com valores de RMSE variando entre 5,37 e 8,90 e R^2 acima de 0,90 nas três estações avaliadas. O RF também apresentou resultados robustos, com RMSE entre 5,82 e 13,98 e R^2 consistentemente acima de 0,84. Já o SVR apresentou desempenho competitivo em alguns cenários,

enquanto o Prophet obteve resultados menos expressivos, com R^2 negativos e RMSE mais elevados, acima de 21, o que reforça que pode ser mais adequado para o acompanhamento de tendência do que para previsões de curto prazo.

Esses resultados reforçam a eficácia dos modelos de aprendizado de máquina — especialmente o XGBoost — para previsão de séries hidrológicas com comportamento dinâmico. O Prophet é útil como ferramenta complementar de acompanhamento de tendência. Por fim, os resultados servem de base para a etapa seguinte do projeto, que é a integração dessas previsões a um sistema de detecção de falhas em sensores do SNIRH por meio da análise de padrões e anomalias.

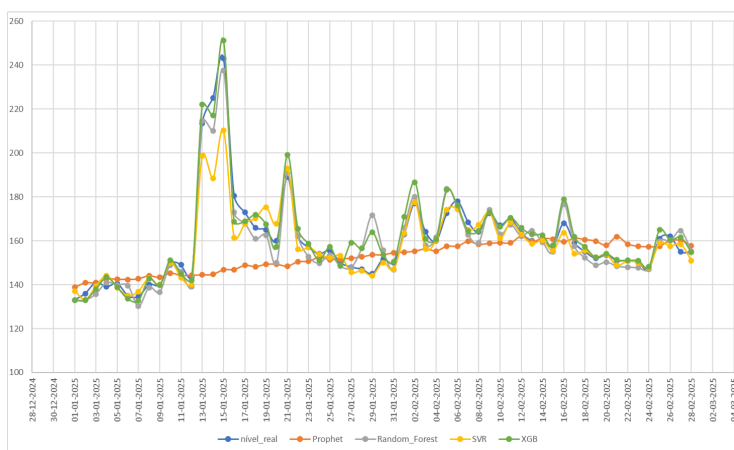


Figura 3. Previsão X Dados Reais: Rio Quilombo (Prophet, RF, SVR, XGB).

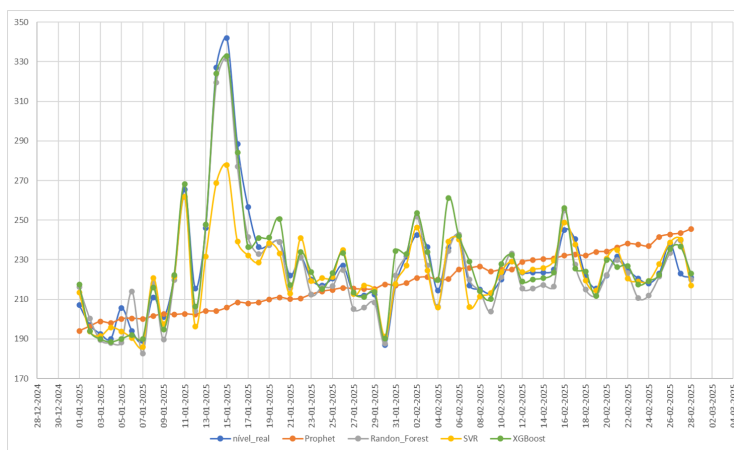


Figura 4. Previsão X Dados Reais: Rio Casca (Prophet, RF, SVR, XGB).

5. Conclusão e Trabalhos Futuros

Neste trabalho, foram avaliadas quatro abordagens de modelagem — Prophet, Random Forest, SVR e XGBoost — aplicadas a séries temporais de nível d'água em três estações hidrológicas do Rio M. Para o processamento usando esses dados, os resultados mostraram que, enquanto o modelo Prophet é mais eficiente na captura de tendências amplas e variações sazonais, os modelos baseados em aprendizado de máquina, especialmente

Random Forest e XGBoost, se destacam na previsão de eventos pontuais e comportamentos transitórios.

É necessário realizar novos processamentos, com mais dados e parâmetros, para abranger a série temporal dos dados. Porém, é possível perceber que a adoção desses modelos pode trazer benefícios significativos para a manutenção preditiva em sistemas de monitoramento hidrológico, permitindo intervenções mais rápidas e assertivas em casos de falhas de sensores, desvios inesperados ou padrões atípicos de comportamento.

Como continuidade deste trabalho, sugere-se a integração dos modelos de previsão com técnicas de detecção de anomalias — como Autoencoders, Isolation Forest e análise por entropia — aliadas a métodos de identificação de comportamentos incomuns, como períodos de baixa oscilação, valores fora de faixas históricas de cheia e seca, e outros indicadores de funcionamento anormal. A combinação dessas técnicas visa aprimorar a ferramenta de detecção de falhas em sensores de nível, elevando sua confiabilidade e capacidade de operação autônoma. Além disso, a implantação prática do sistema em tempo real, com *dashboards* inteligentes e alertas automáticos, consolidam os próximos passos rumo a uma solução completa de apoio à gestão hidrológica e monitoramento preventivo.

Referências

- ANA (2016). *Manual de Procedimentos para Estações Hidrológicas*. ANA, Brasília, Brasil.
- ANA (2021). *Relatório Conjuntura dos Recursos Hídricos - 2021*. ANA, Brasília, Brasil.
- ANA (2022). *Manual de Instalação de Estações Fluviométricas*. Agência Nacional de Águas e Saneamento Básico.
- Barzegar, R., Aalami, M. T., and Adamowski, J. (2021). Coupling a hybrid cnn-lstm deep learning model with a boundary corrected maximal overlap discrete wavelet transform for multiscale lake water level forecasting. *Journal of Hydrology*.
- EPAGRI/CIRAM-ANA (2024). *Manual de Operação da Sala de Situação*. EPA-GRI/CIRAM e ANA.
- Flyrank, A. R. T. (2023). How to use random forest for regression. Online.
- Jaimes Campos, D. L. and López Zúñiga, E. (2021). Modelo de forecast para predecir la demanda semanal de alimentos y bebidas de consumo masivo.
- Manandhar, P., Rafiq, H., Rodriguez-Ubinas, E., and Palpanas, T. (2024). New forecasting metrics evaluated in prophet, random forest, and long short-term memory models for load forecasting. *Energies*.
- Niazkar, M., Shamshirband, S., Kisi, O., and Choubin, B. (2024). Applications of xgboost in water resources engineering: A systematic literature review (dec 2018–may 2023). *Environmental Modelling and Software*, 174:105971.
- Sharma, B. and Goel, N. K. (2024). Streamflow prediction using support vector regression machine learning model for tehri dam. *Applied Water Science*, 14(99).
- Shearer, C. (2000). The crisp-dm model: The new blueprint for data mining.
- SNIRH (2021). *Manual de Operação do Sistema de Telemetria*. Agência Nacional de Águas e Saneamento Básico.