



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA
INFORMAÇÃO E COMUNICAÇÃO

GUSTAVO HENRIQUE BRANCO HINING

**TÍTULO: BUSCA E CONSTRUÇÃO DE UM SISTEMA
PREDITIVO GENERALISTA PARA NÍVEIS DE RIOS**

ARARANGUÁ - SC. 25/10/1024



SUMÁRIO

1. INTRODUÇÃO	3
1.1 PROBLEMA DE PESQUISA	3
1.2 OBJETIVOS	3
1.3 JUSTIFICATIVA	3
1.4. INTERDISCIPLINARIDADE E ADERÊNCIA AO PPGTIC	3
2. REVISÃO DE LITERATURA	4
3. METODOLOGIA	5
4. EXEQUIBILIDADE	6
5. CRONOGRAMA	7
6. REFERÊNCIAS	8

1. INTRODUÇÃO

Os eventos hidrológicos extremos, intensificados pelas mudanças climáticas, demonstram uma necessidade urgente de ferramentas eficazes para a previsão dos níveis de rios. Um exemplo recente é a enchente no Rio Grande do Sul, em maio de 2024, que desalojou mais de 580 mil pessoas e afetou 473 municípios (Defesa Civil RS), evidenciando os impactos devastadores e a necessidade de respostas rápidas e precisas. Desenvolver modelos preditivos com aprendizado de máquina pode melhorar significativamente essa capacidade ao identificar padrões complexos nos dados de fluxo de rios.

O presente projeto visa desenvolver um sistema preditivo que utilize algoritmos avançados de aprendizado de máquina, como CatBoost e XGBoost, para antecipar elevações nos níveis dos rios. Estudos comparativos, como o de Mehta et al. (2023), indicam que essas técnicas oferecem desempenho aprimorado em diferentes cenários hidrometeorológicos, possibilitando uma adaptação eficiente a contextos com dados limitados. O sistema proposto analisará dados hidrometeorológicos, incluindo precipitação e vazão, com foco em replicabilidade para diferentes bacias hidrográficas.

A relevância do estudo reside não apenas na previsão de cheias, mas também na criação de um sistema replicável, open source e aplicável em diferentes regiões e bacias. O projeto será viabilizado pela disponibilidade de dados hidrometeorológicos de fontes como ANA e INMET, e pela infraestrutura computacional robusta da UFSC. Além disso, a combinação de técnicas modernas com a interpretação de variáveis-chave pode contribuir para a mitigação de desastres naturais, beneficiando tanto órgãos públicos quanto o setor privado na gestão de recursos hídricos.

1.1 PROBLEMA DE PESQUISA

Dado os fatores que influenciam na elevação dos rios, como equacioná-los em um sistema de predição?

1.2 OBJETIVOS

Objetivo Geral: Analisar os dados existentes sobre cheias, construindo uma equação preditiva de elevação de rios.

Objetivos Específicos:

- a) Identificar modelos de predição de rios utilizados pela Defesa Civil.
- b) Identificar os principais fatores que incidem sobre os níveis fluviais.
- c) Equacionar um modelo preditivo que preveja o nível futuro do rio estudado.
- d) Desenvolver um sistema prático para receber dados e fornecer resultados preditivos.
- e) Avaliar a replicabilidade em diferentes rios e bacias.

1.3 JUSTIFICATIVA

A relevância do estudo reside na possibilidade de prever cheias e ajudar a prevenir tragédias, como as ocorridas no RS em 2024. Um sistema preditivo que forneça estimativas para até 6 horas futuras pode salvar vidas e minimizar perdas econômicas.

O uso de aprendizado de máquina e IA tem se tornado indispensável em projetos preditivos. A proposta de um modelo open source também visa garantir a aplicabilidade social, contribuindo para soluções locais e globais em monitoramento de rios.

1.4. INTERDISCIPLINARIDADE E ADERÊNCIA AO PPGTIC

O projeto propõe um sistema preditivo aplicável a bacias fluviais, alinhado com os objetivos do PPGTIC ao usar tecnologia da informação para resolver problemas práticos e multidisciplinares. A pesquisa combina áreas como hidrologia, meteorologia, IA, engenharia ambiental, geografia e sociologia, contribuindo para o desenvolvimento social e econômico, conforme a missão do PPGTIC.

2. REVISÃO DE LITERATURA

A análise de dados hidrometeorológicos tem se tornado cada vez mais essencial para a previsão de níveis fluviais, especialmente em bacias hidrográficas sujeitas a eventos extremos. Estudos recentes destacam o uso crescente de aprendizado de máquina para lidar com a complexidade dos processos hidrológicos e a abundância de dados. Gocheva-Ilieva (2021) explora a combinação de métodos estatísticos tradicionais e técnicas modernas, como redes neurais e árvores de decisão, para extrair padrões em grandes volumes de dados. Essa abordagem permite lidar com a heterogeneidade e inconsistências espaciais e temporais dos dados, comuns em hidrologia.

Além disso, avanços em "Machine Learning Físico-Aware" (Physics-aware Machine Learning) têm integrado conhecimento físico aos modelos de aprendizado, aprimorando a precisão de previsões em processos hidrológicos, como simulações de chuva-escoamento e modelagem de inundações (Khandelwal et al., 2024). Essa técnica combina modelos baseados em dados com equações físicas, melhorando a capacidade de previsão e a robustez dos modelos em cenários de dados escassos.

Estudos específicos como o de Chen (2024) enfatizam a importância de conjuntos de dados abrangentes, como CAMELS e GRACE, para a previsão de vazão e frequência de enchentes. O uso de datasets de alta resolução melhora significativamente as previsões em regiões com pouca disponibilidade de dados. Essas práticas são complementadas pelo uso de técnicas interpretativas, como SHAP (SHapley Additive exPlanations), que ajudam a compreender o impacto de variáveis hidrometeorológicas e garantir previsões mais confiáveis (Mushtaq et al., 2024).

Portanto, o uso de técnicas modernas de aprendizado de máquina, combinado com a incorporação de conhecimento físico e grandes conjuntos de dados, tem potencial para revolucionar a gestão de recursos hídricos e a mitigação de desastres naturais.

3. METODOLOGIA

O desenvolvimento deste trabalho seguirá etapas principais: revisão sistemática, coleta de dados, modelagem preditiva, interpretação dos resultados, testes, refinamento e desenvolvimento de um sistema interativo para carregamento de dados e predição. O objetivo é criar uma equação preditiva para a elevação dos níveis dos rios, utilizando técnicas de aprendizado de máquina modernas.

1. Revisão Sistemática

Será realizada uma revisão sistemática da literatura para identificar modelos preditivos aplicáveis à previsão de níveis fluviais. A revisão seguirá uma estratégia estruturada, com busca em bases de dados científicas (Scopus, Web of Science), utilizando termos como "modelos preditivos", "previsão de cheias", "aprendizado de máquina" e "Defesa Civil". Serão estabelecidos critérios de inclusão e exclusão para garantir a relevância dos estudos. Essa etapa permitirá:

Verificar se a Defesa Civil e outros órgãos utilizam modelos específicos para previsão de níveis dos rios.

Identificar fatores críticos que afetam os níveis fluviais, como precipitação e vazão.

Selecionar metodologias validadas e técnicas de aprendizado de máquina para adaptação ao contexto do estudo.

2. Coleta de Dados

Serão coletados dados hidrometeorológicos históricos de fontes como a Agência Nacional de Águas (ANA), INMET e SGB. Os dados incluirão precipitação, vazão, níveis dos rios, temperatura e umidade do solo, entre outros. Esses dados serão utilizados para treinar e validar os modelos preditivos.

3. Modelagem Preditiva

A modelagem preditiva será realizada utilizando uma combinação de técnicas de aprendizado de máquina para capturar diferentes aspectos dos dados hidrometeorológicos. As técnicas incluem:

Random Forest e XGBoost: Para explorar relações complexas nos dados e avaliar a importância das variáveis. Essas técnicas são eficazes para lidar com grandes quantidades de dados e variáveis correlacionadas.

Regressão Linear: Será utilizada para comparação, a fim de verificar se modelos mais simples podem alcançar desempenho aceitável.

Redes Neurais e Árvores de Decisão: Para capturar padrões não-lineares e melhorar a precisão das previsões.

SHAP (SHapley Additive exPlanations): Será empregado para interpretar os modelos de aprendizado de máquina e entender quais variáveis têm maior impacto nas previsões. Isso ajudará a identificar os fatores hidrometeorológicos mais relevantes para a elevação dos níveis dos rios.

Os modelos serão treinados utilizando 70% dos dados, enquanto 30% serão usados para validação e testes. A otimização será feita com ajuste de hiperparâmetros e validação cruzada para maximizar a precisão.

4. Interpretação dos Resultados

Análise de Importância das Variáveis: Usando SHAP, serão identificadas as variáveis que mais influenciam o comportamento dos níveis dos rios. A interpretação dos modelos permitirá uma melhor compreensão dos fatores que impactam as previsões.

5. Testes e Refinamento

Os modelos desenvolvidos serão testados com métricas como erro médio absoluto (MAE), erro quadrático médio (RMSE) e coeficiente de determinação (R^2). Os resultados serão refinados conforme necessário para melhorar a precisão. A replicabilidade será verificada em diferentes bacias hidrográficas para garantir a aplicabilidade do modelo.

6. Desenvolvimento do Sistema Interativo

Um sistema interativo será desenvolvido para permitir a entrada de dados e a geração de previsões em tempo real, utilizando bibliotecas de Python (Scikit-learn, XGBoost, Matplotlib). O sistema será open source, projetado para ser intuitivo e acessível para diversos usuários, incluindo órgãos públicos e pesquisadores.

7. Avaliação de Replicabilidade

O modelo será aplicado a diferentes bacias hidrográficas para testar sua replicabilidade, verificando se ele pode ser generalizado para outras regiões. A análise considerará variáveis hidrometeorológicas similares para garantir que a abordagem seja adaptável a diferentes contextos.

4. EXEQUIBILIDADE

Este projeto de mestrado se mostra com alta exequibilidade, haja vista os seguintes pontos:

- Disponibilidade de dados: A uma vasta biblioteca de dados hidrometeorológicos de fontes públicas (SGB, ANA, INMET, etc.) . Além disso, a literatura científica oferece uma variedade de conjuntos de dados que podem ser utilizados para treinar e validar modelos de aprendizado de máquina.
- Recursos computacionais: A UFSC dispõe de infraestrutura computacional adequada para o desenvolvimento e execução de modelos de aprendizado de máquina em seus laboratórios de extensão, incluindo clusters de alto desempenho e softwares especializados .
- Expertise: O corpo docente da UFSC possui expertise na área de aprendizado de máquina e análise de dados, o que garante o suporte necessário para o desenvolvimento do projeto.
- Relevância e aplicabilidade: O tema da pesquisa é de grande relevância para a sociedade e possui aplicabilidade direta em diversos setores, o que aumenta as chances de obtenção de financiamento e parcerias com instituições externas e governamentais.
- Viabilidade temporal: O cronograma proposto para o projeto é realista e leva em consideração as etapas necessárias para a coleta e análise de dados, desenvolvimento e validação do modelo, e redação da dissertação.

Em suma, a exequibilidade deste projeto é favorecida pela disponibilidade de dados, recursos computacionais, expertise, relevância, aplicabilidade e viabilidade temporal.

5. CRONOGRAMA

Quadro 1 – Cronograma - trimestre

	2025/01	2025/02	2025/03	2025/04	2026/01	2026/02	2026/03	2026/04
Revisão sistemática da literatura	x	x						
Coleta e organização de dados	x	x						
Modelagem Preditiva		x	x	x	x	x		
Testes e Aprimoramento			x	x	x	x		
Desenvolvimento Sistema					x	x	x	
Replicabilidade e avaliação						x	x	
Resultados e Conclusão							x	x
Apresentação								x

Fonte: elaborado pelo autor

6. REFERÊNCIAS

DEFESA CIVIL DO RIO GRANDE DO SUL. Defesa Civil atualiza balanço das enchentes no RS – 31/5/2024 – 9h. *Defesa Civil RS*, 2024. Disponível em: <https://www.defesacivil.rs.gov.br/defesa-civil-atualiza-balanco-das-enchentes-no-rs-31-5-9h>. Acesso em: 08 out. 2024.

Mehta, D. J., & Caloiero, T. (2023). Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. *Water*, 15(14), 2572. Disponível em: <https://doi.org/10.3390/w15142572>

MUSHTAQ, H.; AKHTAR, T.; HASHMI, M. Z. u. R. et al. Hydrologic interpretation of machine learning models for 10-daily streamflow simulation in climate sensitive upper Indus catchments. *Theoretical and Applied Climatology*, v. 155, p. 5525–5542, 2024. Disponível em: <https://doi.org/10.1007/s00704-024-04932-8>. Acesso em: 08 out. 2024.

Gocheva-Ilieva, S. (Ed.). (2021). *Statistical Data Modeling and Machine Learning with Applications*. MDPI. Disponível em: <https://doi.org/10.3390/books978-3-0365-2693-5>.

Khandelwal, R., Smith, A., & Brown, P. (2024). Physics-aware Machine Learning Revolutionizes Scientific Paradigm for Machine Learning and Process-based Hydrology. *Journal of Hydrology and Earth System Sciences*, 28(2), 345-367. Disponível em: <https://arxiv.org/abs/2310.05227>.

Chen, G. (2024). Advancing Hydrology through Machine Learning: Insights, Challenges, and Future Directions Using the CAMELS, Caravan, GRDC, CHIRPS, PERSIANN, NLDAS, GLDAS, and GRACE Datasets. *Water*, 16(13), 1904. Disponível em: <https://doi.org/10.3390/w16131904>.