

## SEMINÁRIO

### 1. Introdução

Neste seminário, abordaremos a análise de séries temporais aplicada ao estudo hidrológico, utilizando dados referente a parâmetros de qualidade da água e ambientais. Essa base de dados proporciona uma oportunidade de investigar as variações e padrões sazonais, bem como as mudanças e comportamentos recorrentes de parâmetros hidrológicos e climáticos ao longo de mais de quatro anos. Essa análise busca identificar tendências significativas e fornecer uma compreensão mais profunda dos fenômenos da região.

### 2. Objetivos

O objetivo deste trabalho é testar uma linha de base (baseline) composta por diversos modelos de séries temporais para prever o nível de turbidez da água, apesar da base de dados possuir mais atributos. Assim, teremos um problema de séries temporais univariadas.

### 3. Metodologia

Na metodologia proposta para o seminário, o primeiro passo é investigar valores faltantes e *outliers* no conjunto de dados. Devem ser discutidas e aplicadas técnicas apropriadas para lidar com esses problemas, como imputação de dados e remoção de outliers. Durante a análise exploratória, também deverão identificadas tendências, sazonalidades, autocorrelação, dentre outros aspectos inerentes à series temporais, permitindo a aplicação de tratamentos que otimizem o desempenho dos algoritmos escolhidos.

No desenvolvimento do modelo, cada grupo deve escolher uma abordagem específica, implementar e avaliar seu modelo, discutir os resultados e desafios enfrentados, utilizando ferramentas como GluonTS, Sktime, PyTorch Forecasting, Darts, AutoTS, FB Prophet, dentre outros. É desejável que os experimentos sejam documentados na ferramenta MLFlow.

A avaliação deverá ser apoiada por meio de três métricas importantes:

- **MAPE** (*Mean Absolute Percentage Error*): é calculado tomando a média das diferenças absolutas entre os valores previstos e os valores reais, quanto menor o valor do MAPE, melhor será a precisão do modelo.
- **MSE** (*Mean Squared Error*): calcula a média dos erros quadrados entre os valores previstos e os valores reais, penalizando de forma mais significativa os erros maiores. Quanto menor o valor, maior a precisão do modelo.
- **RMSE** (*Root Mean Squared Error*): é a raiz quadrada do MSE, fornece uma medida do desvio padrão dos erros de previsão e é útil para interpretar a precisão do modelo em unidades reais dos dados.

### 4. Cronograma do Projeto do Seminário

- **SEMANA 07 (07/11) – Ambientação nas Ferramentas:** Introdução e familiarização com a ferramenta escolhida (GluonTS, Sktime, PyTorch Forecasting, Darts, AutoTS, FB Prophet, dentre outros). Configuração do

ambiente de trabalho e revisão dos principais recursos de cada biblioteca.  
Utilizar datasets de séries temporais abertas.

- **SEMANA 08 (14/11) – Análise das Séries Temporais:** Coleta e preparação dos dados. Análise exploratória das séries temporais, identificação de padrões e compreensão das características das séries (tendências, sazonalidades, etc.). Será disponibilizado um dataset único.
- **SEMANA 09 (21/11) – Desenvolvimento dos Modelos – Parte 1:** Início do desenvolvimento de modelos básicos usando as ferramentas mencionadas, com foco na implementação e ajustes iniciais.
- **SEMANA 10 (28/11) – Desenvolvimento dos Modelos – Parte 2:** Refinamento dos modelos, implementação de abordagens mais avançadas, e comparações de desempenho entre diferentes ferramentas e técnicas.
- **SEMANA 11 (05/12) – Validação e Ajustes Finais:** Avaliação dos resultados obtidos, ajustes finais nos modelos, e preparação para a apresentação. Revisão dos pontos fortes e limitações da ferramenta usada.
- **SEMANA 12 (12/12) – Apresentação do Seminário:** Realização do seminário, com apresentação das análises, modelos desenvolvidos, resultados e conclusões, seguida de uma sessão de perguntas e respostas.