



Workshop de Medições do Comitê Técnico de Monitoramento da RNP 2023

Identificação de Problemas e Correção de Dados de Monitoramento no MonIpê para Predição de Desempenho

01/11/2023

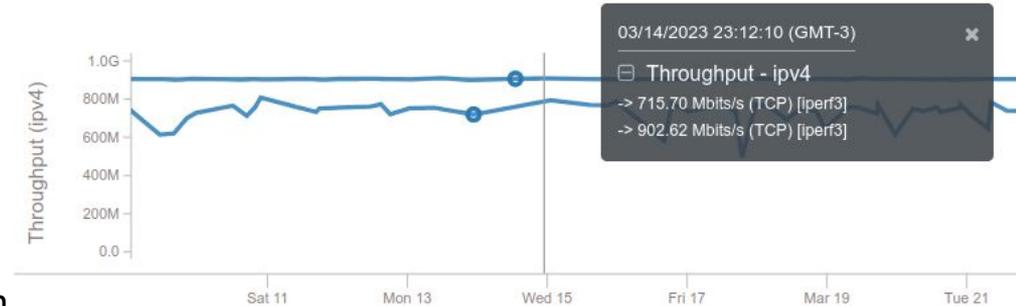
Problema

- **Falhas em Medições**

- Algumas medições não ocorrem.
 - Principalmente de Vazão: “esforço” alto para a rede.
 - Motivo: Restrição de Recursos ? Problemas de Agendamento ? Falha de Conexão ? Outros ?
- Resultam em lacunas nos dados de monitoramento.

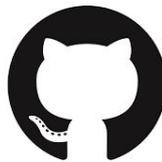
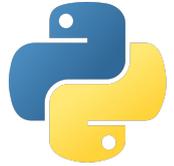
- **Impacto nas ações de P&D**

- Soluções da análise de dados são comprometidas.
- Série temporal “quebrada”.
- Falta de informações em determinados períodos.



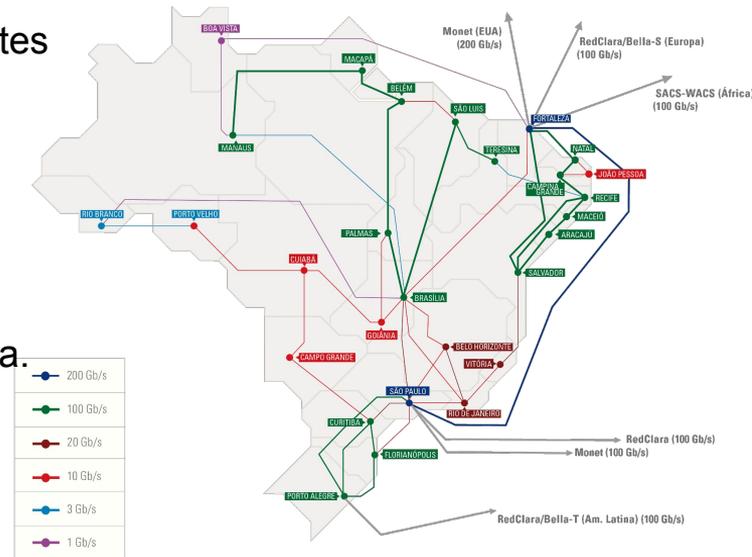
Solução Proposta

- **Identificação de Falhas de Medição e Correção de Dados**
 - Coletar os dados (via API Esmond).
 - Buscar lacunas de medição.
 - Realizar a Imputação de dados nas lacunas.
 - Predição de desempenho com dados imputados.
- **Resultados Esperado:**
 - Identificar a melhor abordagem de imputação para a predição de desempenho.
- **Tecnologias**
 - Python, Tensorflow, Jupyter, Github e Trello.



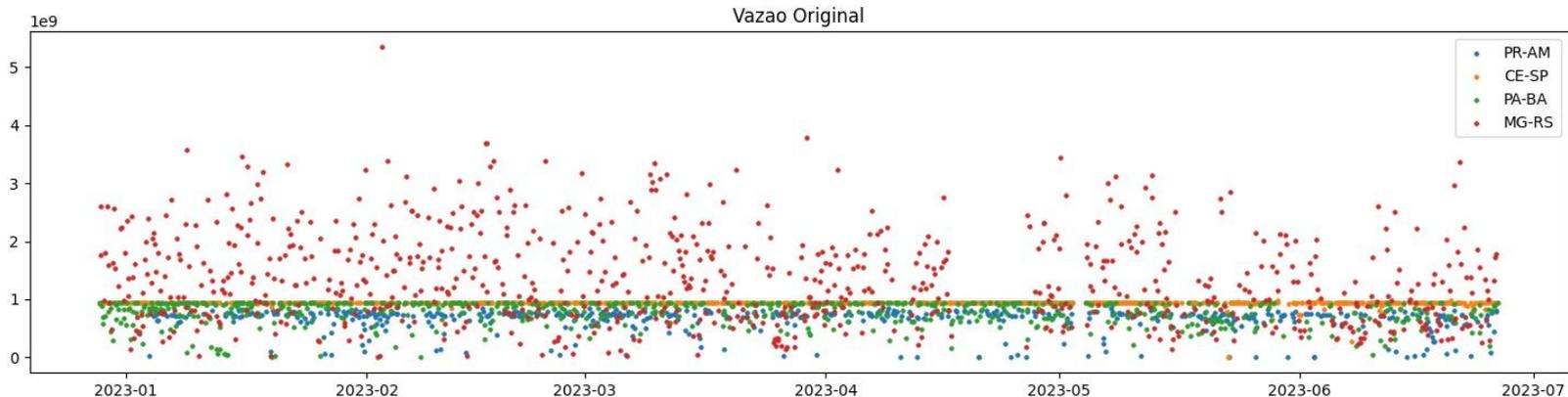
Coleta de Dados

- Coleta de Dados via API Esmond (PerfSonar): Janeiro a Junho de 2023 (150 dias)
 - Vazão BBR (Bottleneck Bandwidth and Round-trip propagation time)
 - Responde prioritariamente a perda de pacotes
 - Próprio conjunto de regras
 - Não exclusivamente a perda de pacotes
 - RTT, largura de banda, etc.
 - Vazão CUBIC
 - Função cúbica
 - Janela de congestionamento.
 - Resposta mais assertivamente a perdas:
 - Sem restringir o crescimento da janela.



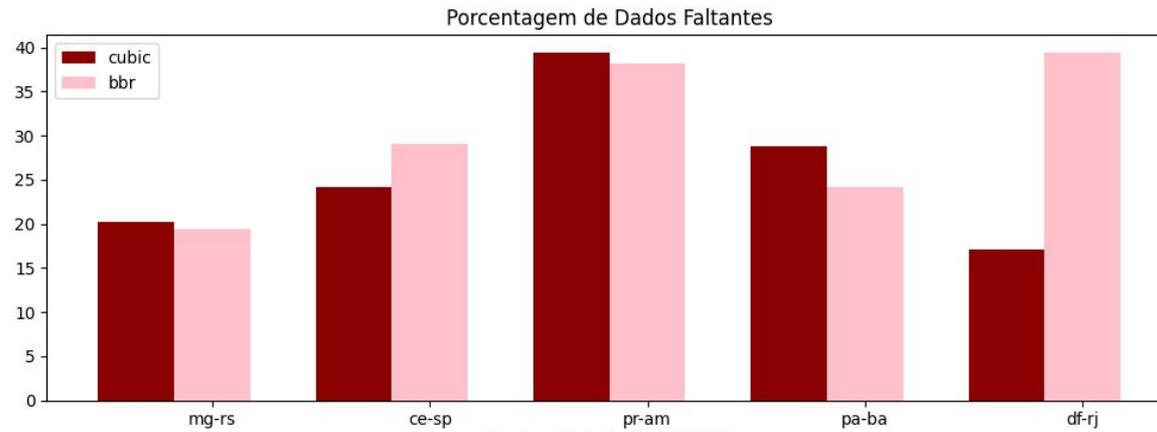
Coleta de Dados

- Coleta de Dados via API Esmond (PerfSonar): Janeiro a Junho de 2023 (150 dias)
 - Pontos de Comunicação selecionados:
 - Trabalhar com PoPs com perfis diferentes de comunicação (distâncias, número de links usados, capacidade dos links, etc)
 - Pontos de Comunicação: CE-SP, PR-AM, MG-RS e PA-BA.

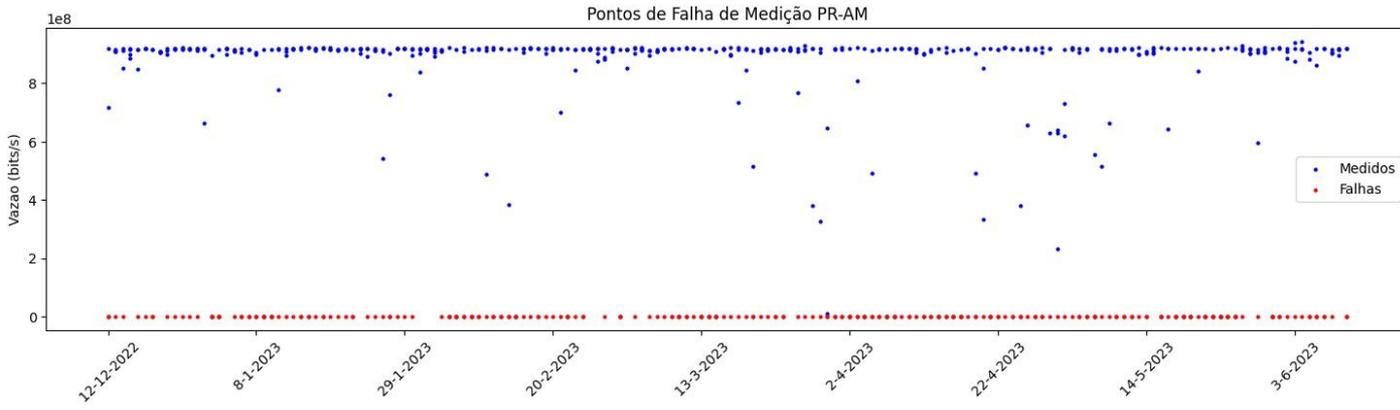
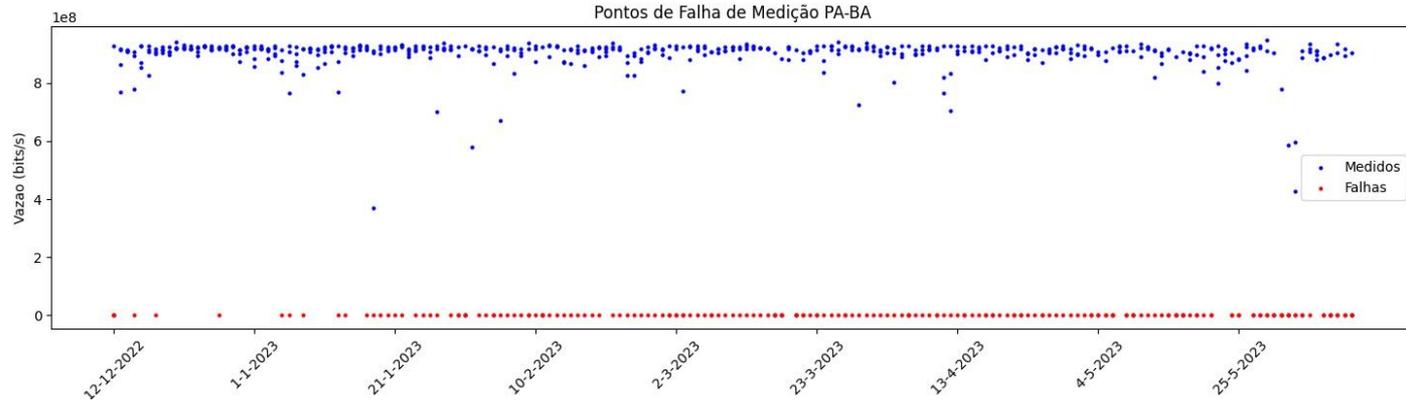


Identificação de Falhas de Medição

- Busca por falhas de medição
 - As medições de vazão ocorrem a cada 4h
 - Teoria: 6 medições por dia.
 - Prática: Em média 30% das medições não ocorrem.
 - Casos raros de dias seguidos sem falhas.
- Organização por períodos do dia: Madrugada, Manhã, Tarde e Noite

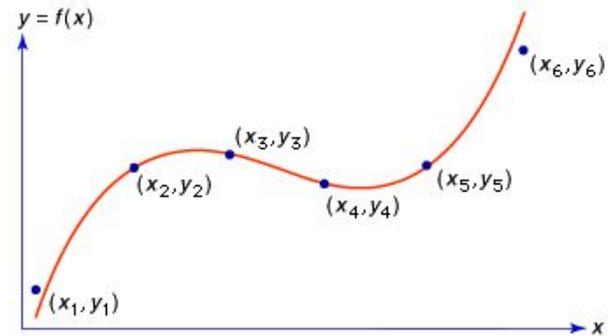


Identificação de Falhas de Medição



Métodos de Imputação de Dados

- Interpolação
 - Interpolação Linear
 - Os valores são calculados assumindo uma relação linear entre os dados.
 - Interpolação Ponderada no Tempo
 - São aplicados pesos aos valores de acordo com a distância deles da brecha.

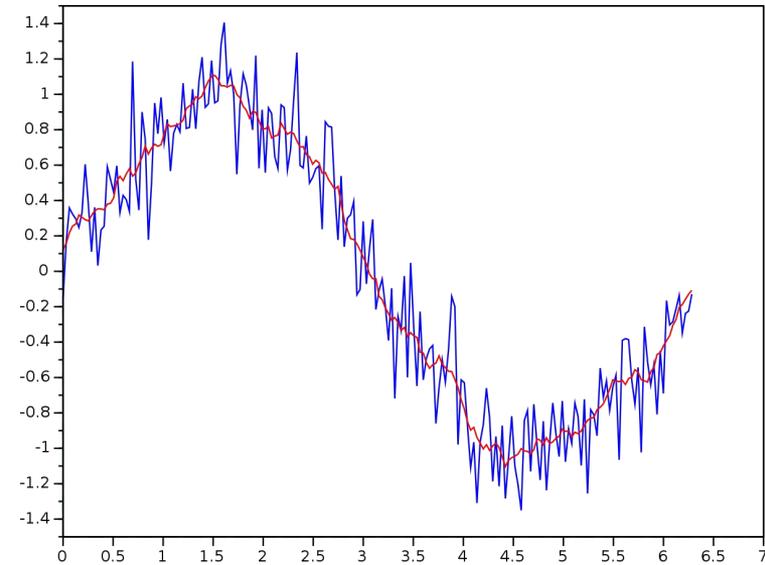


© 2003 Encyclopædia Britannica, Inc.

$$Y = Y_1 + \frac{(Y_2 - Y_1)}{(X_2 - X_1)} * (X - X_1)$$

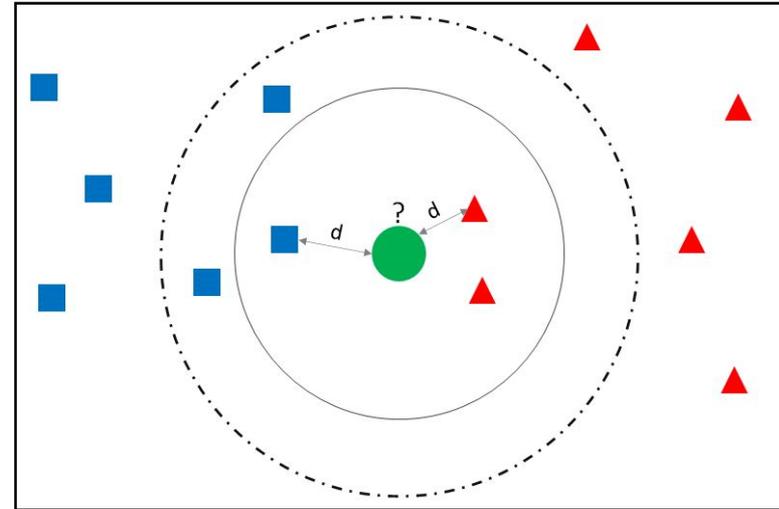
Métodos de Imputação de Dados

- Média e Mediana Móveis
 - O valor da lacuna é substituído por um valor calculado levando em conta uma janela de valores para calcular a média e a mediana.
 - A Mediana Móvel se mostrou menos sensível a outliers e performou melhor que a média.
 - Foram testados diversos valores na janela, mas o RMSE aumentou à medida que os valor da janela aumentava, por conta da grande quantidade de valores faltantes.



Métodos de Imputação de Dados

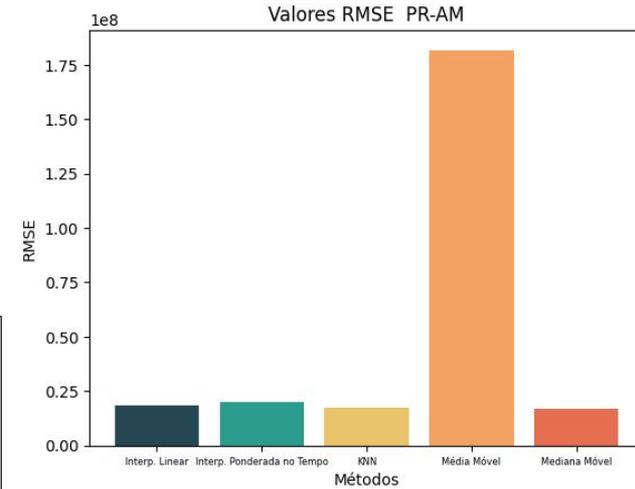
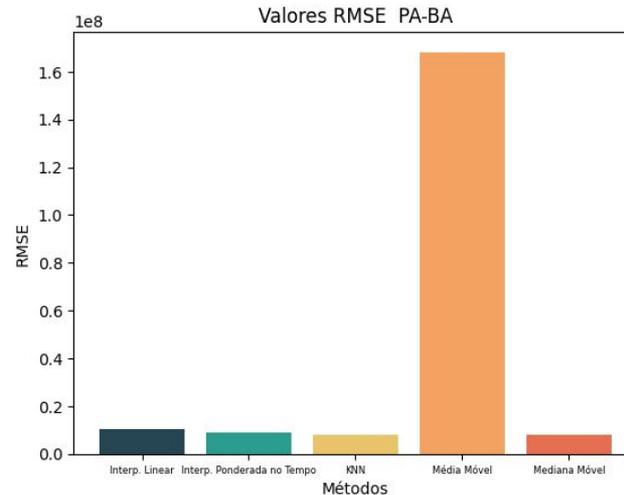
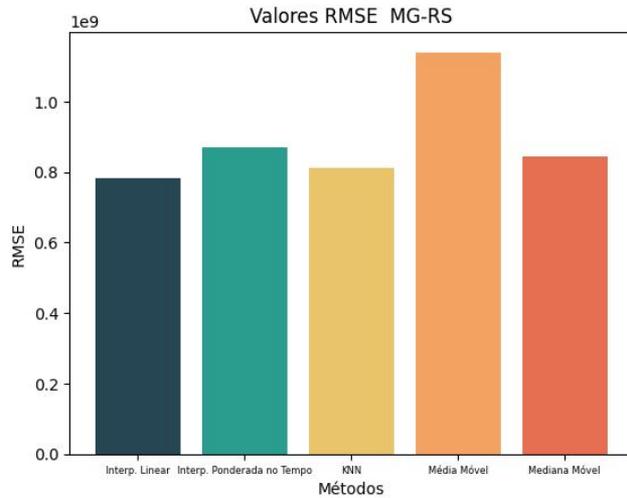
- KNN
 - Preenche as lacunas nos dados pelo valor estimado entre os pontos circundantes na série temporal.
 - Foram testadas diversos valores na janela, mas performaram de forma parecida.



Resultados de Imputação

- Avaliação das Técnicas de Imputação
 - RMSE (Root Mean Squared Error)
 - Medida de erro absoluto que eleva os desvios ao quadrado para impedir que os desvios positivos e negativos se cancelem.
 - Acurácia de imputação dentro da faixa da medição.
 - Definição de intervalos.
 - Verificação se o valor imputado está dentro da faixa do dado real.
 - Visualização dos Dados Imputados
 - Comparação empírica dos dados imputados em relação as medições reais.
 - Impacto na Predição
 - Usando GRU e LTSM.

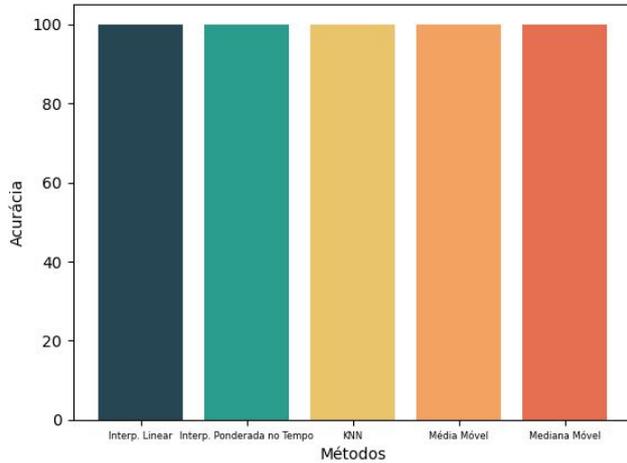
RMSE Métodos de Imputação



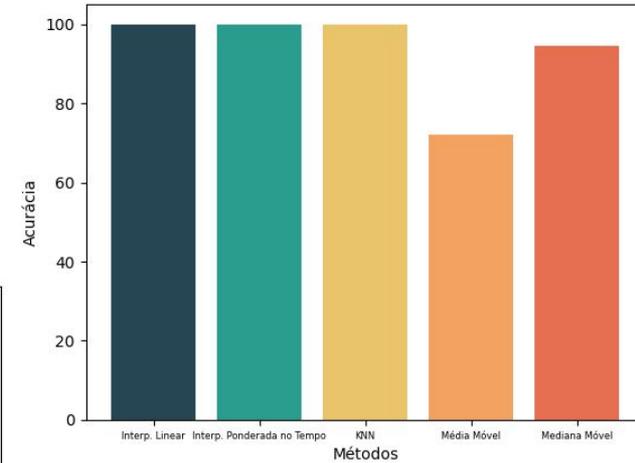
Acurácia Métodos de Imputação



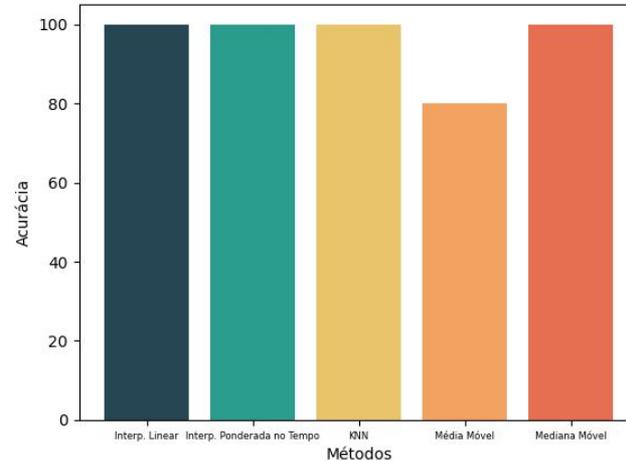
Valores Acurácia MG-RS



Valores Acurácia PR-AM

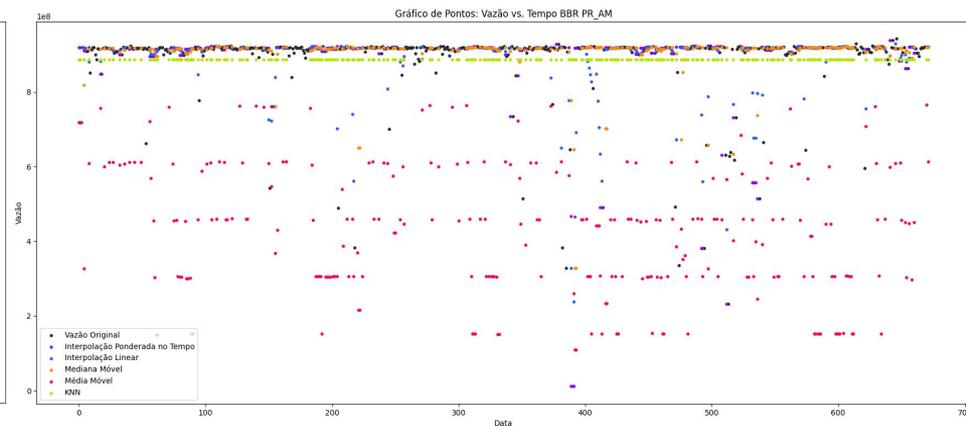
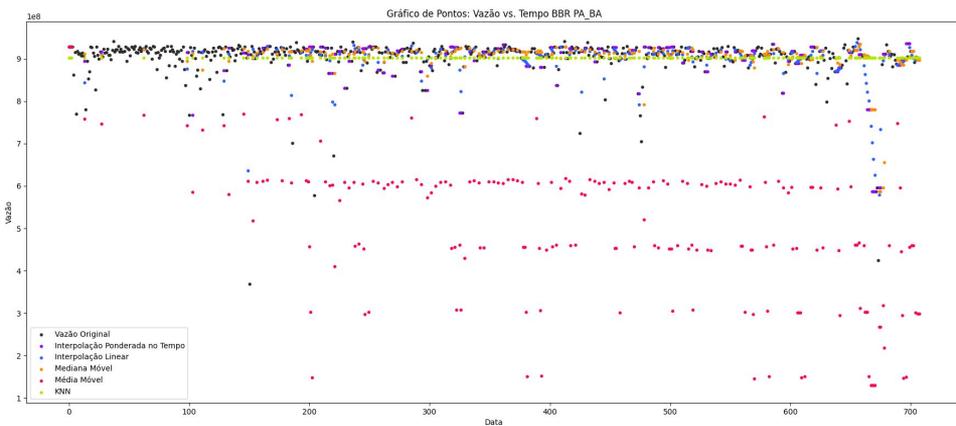
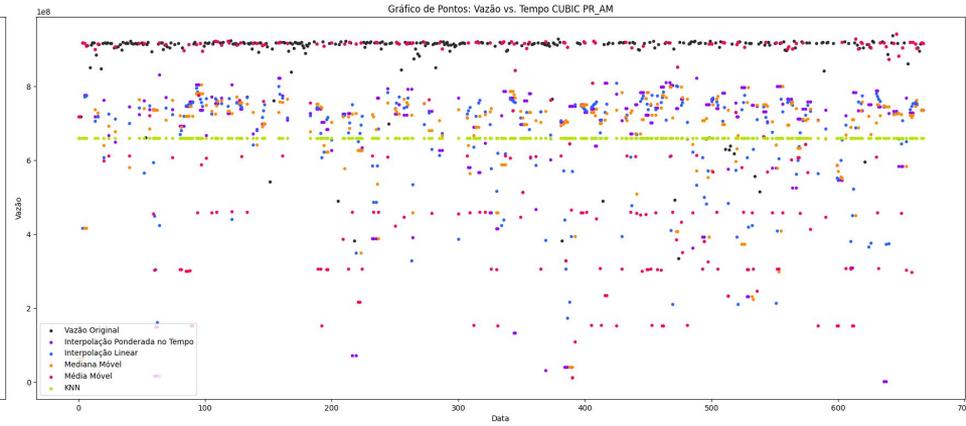
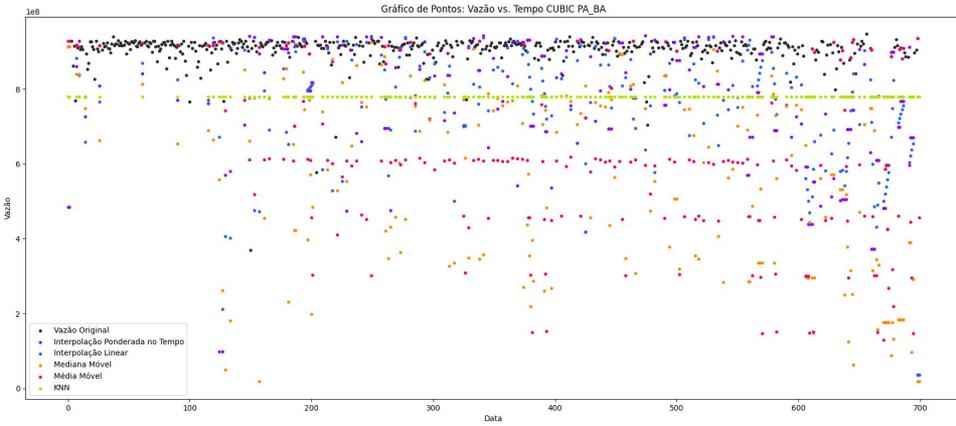


Valores Acurácia PA-BA

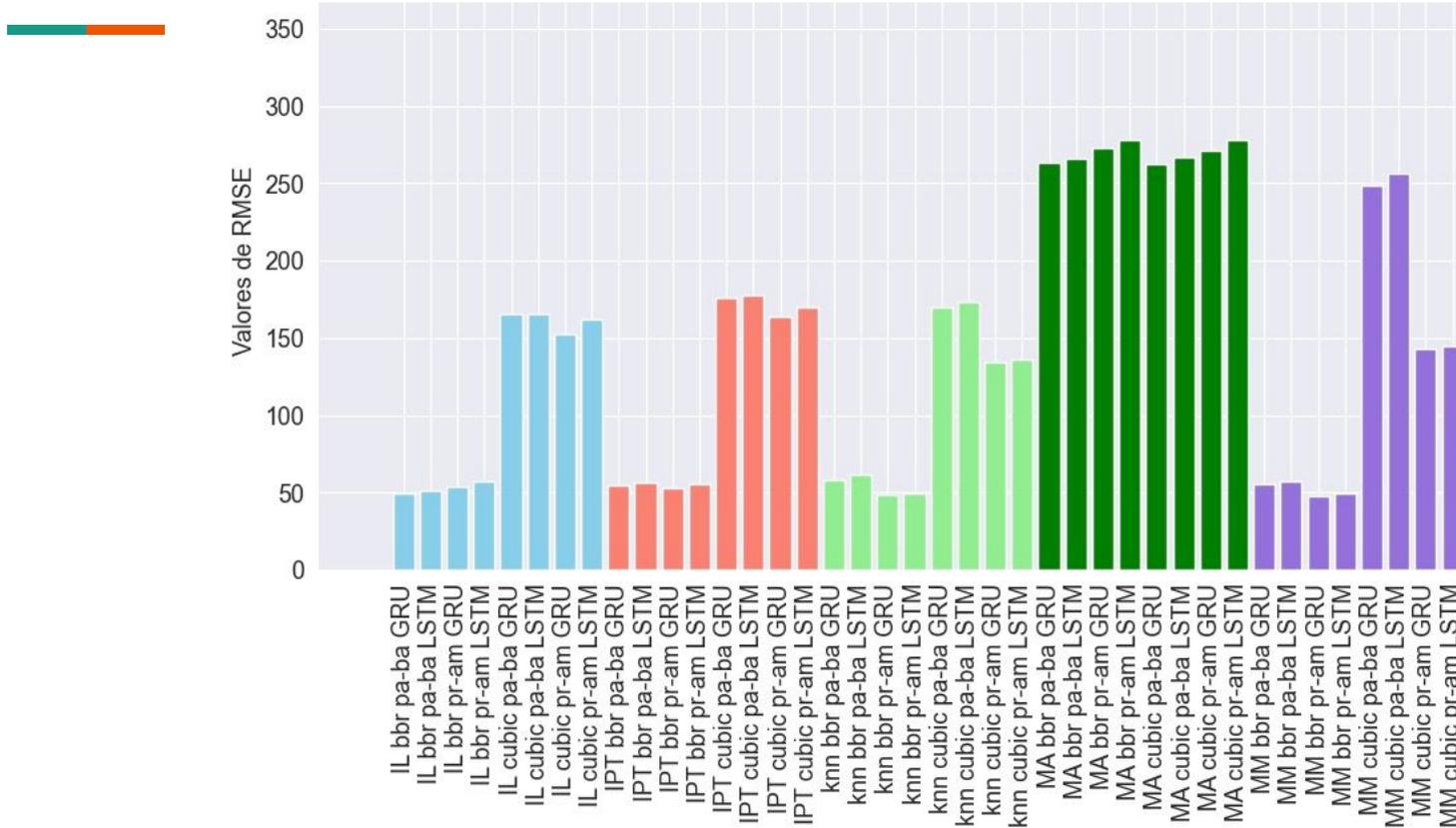


- Vazão Original
- Interpolação Ponderada no Tempo
- Interpolação Linear
- Mediana Móvel
- Média Móvel
- KNN

Visualização dos Dados Imputados

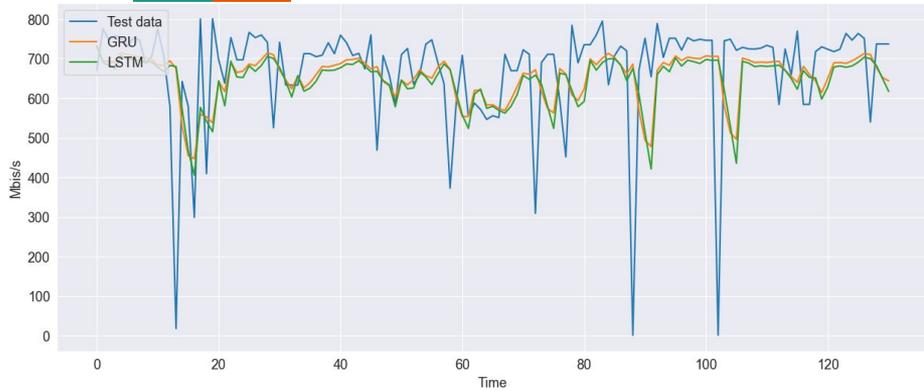


RMSE das Predições

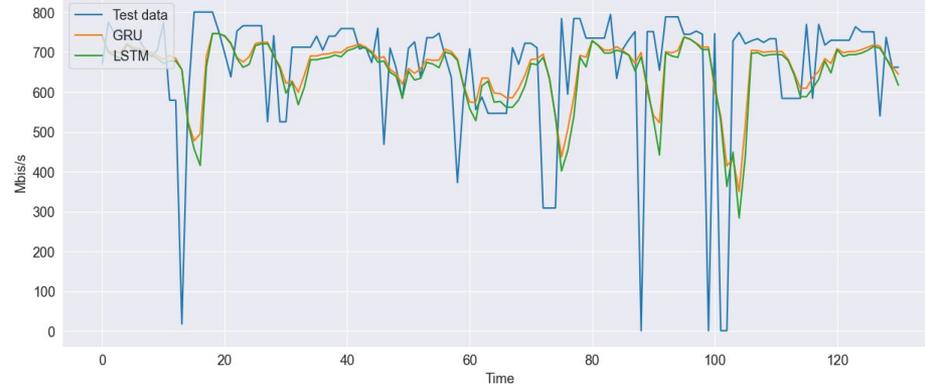


Exemplos de Predição

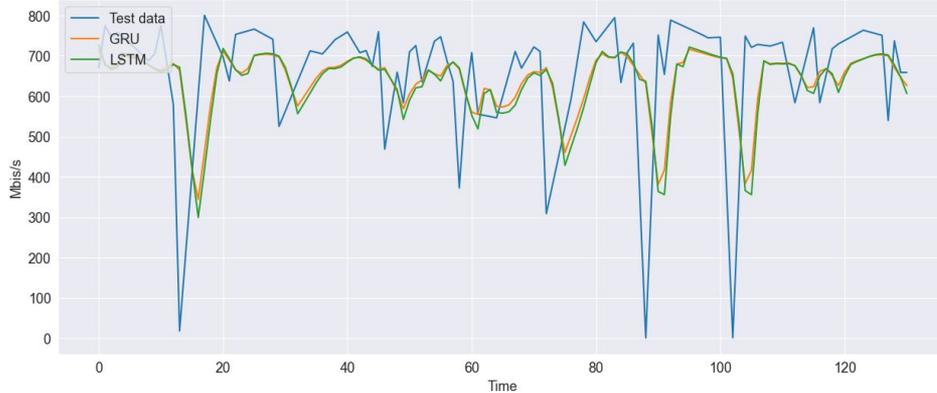
Mediana Móvel - CUBIC - PR-AM



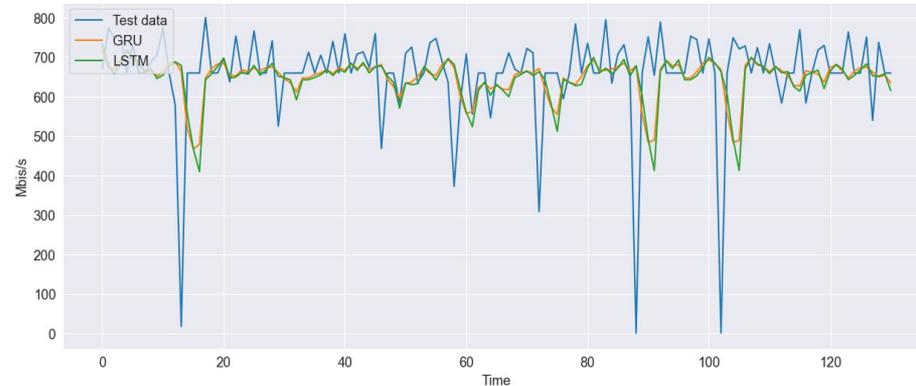
Interpolação Ponderada - CUBIC - PR-AM



Interpolação Linear - Cubic - PR-AM

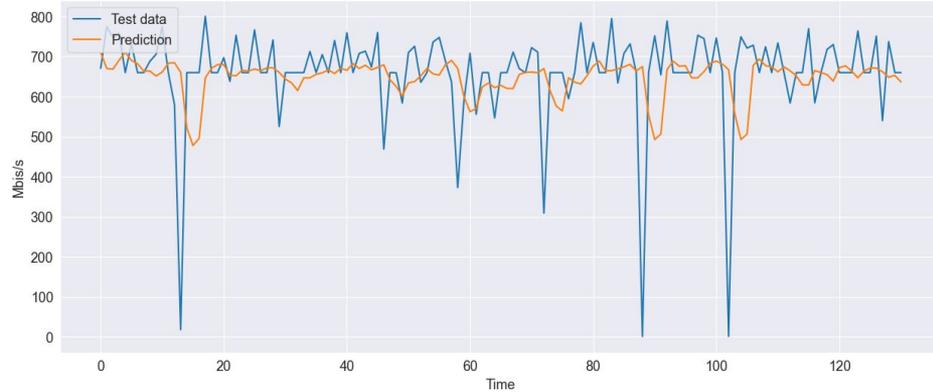


KNN - Cubic - PR-AM

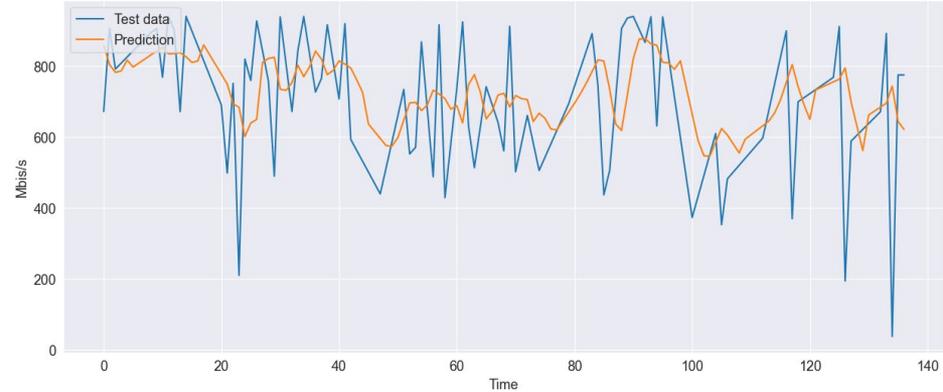


Exemplos de Predição

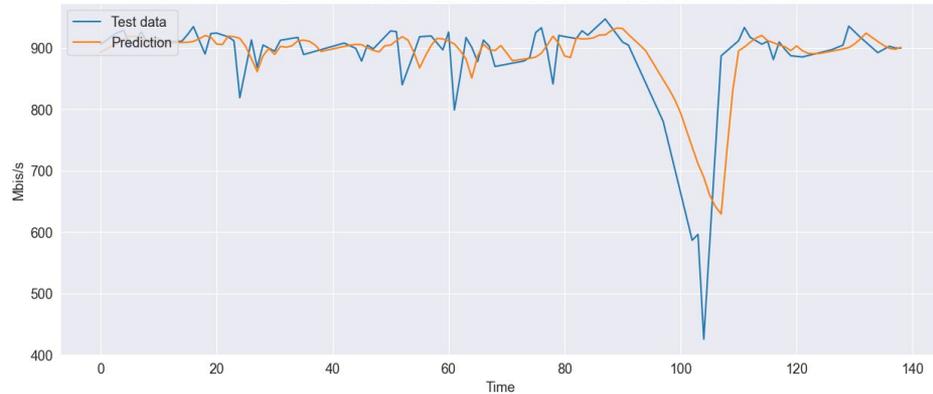
GRU - KNN - CUBIC - PR-AM



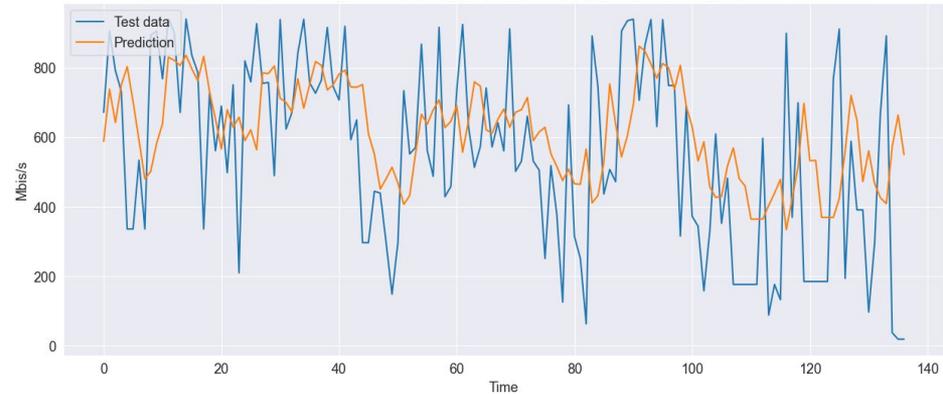
GRU - Interpolação Linear - CUBIC - PA-BA



LSTM - Inter. Linear - BBR - PA-BA



LSTM - Med. Mov. - CUBIC - PA-BA



Próximos Passos

- Buscar por mais Técnicas de Imputação
 - Técnicas baseadas em Inteligência Artificial.
 - Outros métodos Estatísticos: Spline.
 - Propor uma nova abordagem de imputação para o contexto de monitoramento de rede.
- Predição com Dados Imputados
 - Realizar mais experimentos de Predição,
 - Analisando hiperparâmetros: Grid Search.
 - Em conjunto com outras técnicas de imputação.
 - Analisar outros pontos de comunicação.
- Escrita de Artigo e Relatório Final
 - SBRC 2024.
 - Relatório do PMON.

Equipe



Rafael L.



Thelmo P.



Silvio R.



Maria M.



Obrigado!

Email: larcés@uece.br