

Previsão de demanda de água potável em Cidades Inteligentes por meio do algoritmo de modelagem de séries temporais PROPHET

Tobias Barreto^{1, 2}, Flavia Bernardini¹, Daniel de Oliveira¹

¹Instituto de Computação
Universidade Federal Fluminense (UFF) - Niterói, RJ - Brasil

²Coordenação de Gestão de Informações
Gerência de Planejamento, Gestão de Projetos e Informações (PGPI)
Diretoria de Operações - Grupo Águas do Brasil (GAB) - Niterói, RJ - Brasil

tobiasbarreto@id.uff.br, fcbernardini@ic.uff.br, danielcemo@ic.uff.br

Abstract. *Smart Cities (SC) aim at improving services provision through the implementation of new technologies in cities planning, development, operation, and governance. Indicators associated to Basic Sanitation, such as sewage volume index, average per capita water consumption and losses in water distribution, may compose one of the dimensions used for evaluating SCs. In this context, to improve citizens life quality, forecasting the volume of treated water to be distributed is of great importance in cities management. The goal of this paper is to carry out an empirical evaluation of the statistical modeling algorithm PROPHET for time to forecast the volume of water to be distributed in Niterói, RJ. Real data was collected and processed from an organization based in Niterói, RJ.*

Resumo. *As Cidades Inteligentes (CI) buscam melhorar a prestação de serviços por meio da implementação de novas tecnologias com foco no planejamento, desenvolvimento, operação e governança das cidades. Indicadores referentes ao Saneamento Básico, como índice de volume de esgoto, consumo médio per capita de água e perdas na distribuição de água, compõem uma das dimensões de avaliação de uma CI. Nesse contexto, para evolução da qualidade de vida dos cidadãos, a previsão de volume de água tratada a ser distribuída é de grande importância no planejamento estratégico das organizações responsáveis de saneamento básico nas cidades. O objetivo desse artigo é realizar uma avaliação empírica do modelo de series temporais PROPHET para realizar a previsão do volume de água a ser distribuído em Niterói, RJ. Os dados reais foram coletados e tratados de uma organização sediada em Niterói, RJ.*

1. Introdução

Com a crescente urbanização, as cidades vêm buscando melhorar a prestação de serviços por meio da implementação de novas tecnologias no planejamento, desenvolvimento, operação e governança das cidades, criando assim as chamadas Cidades Inteligentes (CI), também conhecidas como Smart Cities [Cruz et al. 2022]. O conceito de Cidades Inteligentes ganhou relevância, em especial na última década [Banni et al. 2022]. Um aspecto

essencial das cidades inteligentes é a utilização de dados e tecnologias de comunicação para monitorar e gerenciar diversos aspectos da vida urbana, incluindo transporte, energia, meio ambiente, segurança pública e governança. As cidades inteligentes são caracterizadas pela aplicação de tecnologias de informação e comunicação em todas as áreas da vida urbana, promovendo uma governança mais eficaz, uma economia mais dinâmica e uma maior qualidade de vida para os habitantes [Giffinger et al. 2007].

As CI são caracterizadas pela co-criação de soluções urbanas entre os diversos atores da cidade, incluindo governos, empresas, organizações da sociedade civil e cidadãos. Isso possibilita uma abordagem mais colaborativa e democrática para a gestão urbana, onde os cidadãos têm um papel ativo na definição das políticas e na resolução dos problemas locais [Anthopoulos 2017]. Os programas e projetos de CI buscam desenvolver soluções inovadoras para minimizar problemas urbanos, e, conseqüentemente, proporcionar uma melhor qualidade de vida para o cidadão e a sociedade. Os centros urbanos apresentam continuamente problemas estruturais que demandam manutenção [Zhang 2003]. Os problemas podem ser apontados pelos seus cidadãos, seja por diversos meios de comunicação ou por participação em conselhos e fóruns municipais que indicam ações para o executivo. Como exemplo do primeiro caso, podemos citar o Colab.re¹, que tem como objetivo centralizar uma gama de serviços para o cidadão e ser uma plataforma ativa para decisões da cidade. Em tal plataforma, ou qualquer outro serviço digital que uma prefeitura possa vir a apresentar para seus cidadãos, as pessoas podem comunicar os problemas encontrados na cidade de forma individualizada. No segundo caso, para que haja a participação dos cidadãos em conselhos e fóruns municipais, o acesso a dados abertos e indicadores deve ser disponibilizados, de forma a permitir uma participação mais ativa em tais espaços. Nessa linha, os dados abertos têm o potencial de serem importantes aliados para o planejamento e gestão de CI, por meio de indicadores tornados públicos.

No contexto de CI o saneamento básico atua em diferentes áreas relacionadas a infraestrutura das cidades, envolvendo serviços de água tratada, esgotamento sanitário, tratamento de esgotos e disposição de resíduos sólidos. O saneamento básico fica intrinsecamente relacionado com as condições de saúde da população e o meio ambiente onde ela está inserida. As ações de Estado e contribuições de empresas privadas no setor de saneamento, contudo, devem ser precedidas de estudos que qualifiquem e quantifiquem as especificidades do saneamento básico na sociedade, especialmente sua relação com outras áreas do conhecimento. Tais estudos permitem a identificação dos impactos, benefícios e custos das ações de saneamento. O fato de diversas doenças serem influenciadas pelas condições inadequadas desses sistemas, tais como diarreias, febres entéricas, hepatite A, dentre outras, reforçam a importância de considerar o saneamento básico como serviço essencial de um município [Camara et al. 2019].

Nesse aspecto, reforçando a questão da análise da qualidade do saneamento básico de municípios, há diversos instrumentos que indicam o que uma cidade pode realizar que vá ao encontro do conceito de CI. Um desses instrumentos é a norma internacional ISO 37122. Tal instrumento apresenta indicadores que permitem um melhor entendimento do que pode ser implementado como serviços com o uso de tecnologias na gestão das cidades

¹Disponível em <https://www.colab.com.br/sou-cidadao/>. Acessado em 23/03/2024.

[Abreu 2023]. Além disso, há também instrumentos, como o *intelli.gente*², sob responsabilidade do Ministério da Ciência, Tecnologia e Inovação (MCTI) com participação do centro de pesquisa IARA (Inteligência Artificial Recriando Ambientes), que mede o nível de maturidade de uma CI em múltiplas dimensões. Uma dessas dimensões é o meio ambiente, que possui indicadores específicos de água e esgoto, dentre outros. Tais instrumentos denotam a importância dos dados abertos e indicadores para melhor entender as perdas e a demanda de consumo de água tratada em um município [de Assis Cunha et al. 2019].

Para análise da demanda de água tratada, as organizações responsáveis pelo tratamento e distribuição medem o volume disponibilizado de água para o município, que é composto pela quantidade de água disponibilizada pelas estações de tratamento de água e reservatórios ao longo do sistema de distribuição. Além da análise dos dados passados, prever o volume disponibilizado de água tem um alto impacto financeiro e ambiental. Para realizar a previsão desse volume de água a ser disponibilizado, será utilizado um modelo de série temporal. Uma série temporal é uma sequência de observações, de mesma natureza, coletados em intervalos de tempo consistentes. Para realizar uma previsão, é necessário receber um valor de entrada (variável original), esse valor é processado dentro de um modelo preditivo, e por fim, o resultado é o valor da previsão (variável predita), sendo necessário observar métricas para estudar os erros dessa previsão e avaliar sua precisão [Chen 2016].

Nessa linha, o modelo PROPHET para previsão de séries temporais tem sido utilizado na literatura em diversos domínios de aplicação. No caso de previsão de demanda de água, suas variações para aprendizado de máquina têm sido aplicadas. Por exemplo, podemos citar o trabalho de Sajjanshetty et al (2023). Esse trabalho apresentou o uso de ensembles, que são uma combinação de modelos de aprendizado de máquina e modelos estatísticos para previsão de consumo de água de uma comunidade [Sajjanshetty et al. 2023]. Os autores utilizaram os modelos treinados pelos algoritmos SARIMAX, PROPHET e XGBoost, sendo os dois primeiros algoritmos de modelagem estatística e o último, um algoritmo de Aprendizado de Máquina. Também há trabalhos que apresentaram um método de previsão de demanda de água urbana utilizando o algoritmo de modelagem estatística PROPHET e algoritmos para construção de redes neurais profundas convolucionais com mecanismo de atenção e recorrentes do tipo LSTM [Liu et al. 2024]. Assim, como primeiro passo é importante analisar a capacidade do algoritmo PROPHET para previsão de demanda de água nas cidades brasileiras para, a partir daí, evoluir métodos de previsão tendo tal algoritmo como base de comparação.

O objetivo deste artigo é apresentar uma análise experimental realizada para previsão de demanda de água potável nos próximos 30 dias com periodicidade de 1 hora. Para tanto, foram utilizados dados reais fornecidos por uma empresa privada de saneamento básico que atua em –Omitido–, bem como o algoritmo estatístico para séries temporais. Para atingir tal objetivo, este trabalho está dividido em cinco seções além da Introdução. A Seção 2 apresenta, brevemente, o PROPHET. A Seção 3 apresenta o método da pesquisa. A Seção 4 os resultados e discussões. Finalmente, A Seção 5 apresenta as conclusões e trabalhos futuros.

²Disponível em <https://inteligente.mcti.gov.br/>. Acessado em 23/03/2024.

2. PROPHET: Modelagem Estatística

PROPHET é um procedimento para previsão de dados de séries temporais que foi criado pelo Facebook Core Data Science. O seu objetivo é ser capaz de prever “em escala”, ou seja, o PROPHET é uma ferramenta de previsão automatizada desde a sua concepção, proporcionando mais facilidade de uso no ajuste de métodos de séries temporais. O PROPHET possui melhor desempenho com séries temporais que apresentam efeitos sazonais e várias temporadas de dados históricos sendo robusto a valores discrepantes e mudanças na tendência [Taylor and Letham 2018]. Nessa linha, o PROPHET é usado para prever dados de séries temporais de médio e longo prazo. Os três principais componentes decompostos dos modelos de séries temporais são tendências, sazonalidade e feriados. Esses componentes podem ser combinados conforme especificado como: $y(t) = g(t) + s(t) + h(t) + et$, onde:

- **g(t)**: mudanças não periódicas nos dados de séries temporais são modeladas usando uma curva de crescimento.
- **s(t)**: são modeladas alterações periódicas nos dados (por exemplo, sazonalidade semanal/anual).
- **h(t)**: captura efeitos de feriados ou datas eventos sinalizadas no parâmetro.
- **et**: são consideradas quaisquer alterações anormais acomodadas pelo modelo.

No PROPHET, existem parâmetros que podem ser ajustados para personalizar o modelo de previsão de acordo com as características dos dados e as necessidades específicas do problema. Para esse estudo experimental, será utilizado 2 opções de tipo de sazonalidade são elas: aditiva e multiplicativa. A sazonalidade aditiva ocorre quando a variação sazonal em uma série temporal permanece constante ao longo do tempo, ou seja, a magnitude da variação sazonal é independente do nível médio da série. A sazonalidade aditiva implica que as oscilações sazonais têm um efeito constante sobre a série temporal, adicionando uma quantidade fixa de variação a cada período sazonal. Já a sazonalidade multiplicativa, presente no PROPHET B, implica que a variação sazonal em uma série temporal é proporcional ao nível médio da série. Isso significa que a magnitude da variação sazonal aumenta ou diminui em relação ao nível médio da série.

É necessário preparar o conjunto de dados e criar um quadro de dados com duas colunas: ‘ds’ como carimbo de data (no formato data e hora) e ‘y’ sendo a medida de previsão que deve estar em valores numéricos. Em seguida, é preciso criar um objeto da classe PROPHET(), onde é possível escolher o período desejado a ser previsto e então prosseguir com o processo de previsão. O resultado da previsão será um conjunto de dados com diversas colunas, nas quais as mais relevantes para esse estudo, são as colunas ‘ds’ e ‘yhat’, sendo ‘yhat’ a coluna que contém os resultados preditos de ‘y’.

3. Método da Pesquisa

Para esse trabalho foi utilizado um conjunto de dados reais, extraídos de uma base de dados de uma organização responsável pelos serviços de saneamento básico de um município. O dataset é composto por uma variável independente (tempo) e uma variável dependente (volume de água distribuída). Para extrair o dataset, foi selecionado um Distrito de Medição e Controle (DMC). DMC é um termo utilizado para uma unidade operacional fundamental em sistemas de abastecimento de água, responsável por monitorar, controlar e otimizar o fluxo de água em uma determinada área geográfica. Um DMC garante a

eficiência e a qualidade do fornecimento de água que compõe a rede de distribuição de água [Schwetschenau et al. 2019]. Cada linha do conjunto de dados representa o volume de água que passou em um determinado DMC em uma janela de tempo de 5 minutos. O Volume de água se encontra representado em metros cúbicos (m³). O conjunto de dados cobre o período de 09/07/2023 às 06:30:00 até 15/12/2023 às 01:20:00, totalizando 27.401 registros. Dessa forma, foi aplicada uma metodologia de análise definida em 3 etapas:

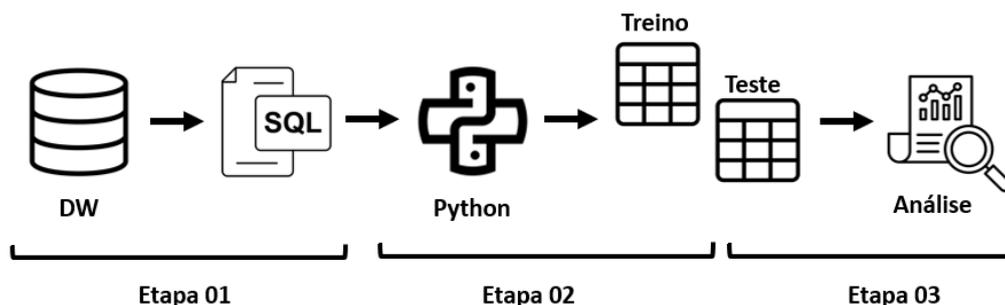


Figura 1. Fluxo do Processo de obtenção de dados

Na Etapa 01, etapa de obtenção dos dados, o conjunto é gerado diretamente no Data Warehouse (DW) de produção de uma empresa privada de saneamento, localizada no estado do Rio de Janeiro. Um DW é um sistema centralizado de armazenamento de dados projetado para facilitar a análise e a tomada de decisões. O DW integra dados de diversas fontes em um único local, fornecendo uma visão unificada e consistente dos dados para os usuários finais.[Kimball and Ross 2013] Por ser tratar de um DW empresarial, a arquitetura do banco e a consulta SQL realizada não serão abordados nesse artigo. Para garantir a reprodutibilidade dessa pesquisa, o conjunto de dados resultante da consulta ao DW pode ser obtido em: https://github.com/tobiasrbarreto/tobias_serie_temporal.

Na Etapa 02 foi aplicado o script desenvolvido em Python 3.11. Esse script realiza todo o processo dessa etapa em diante. A seguir é apresentado as técnicas de tratamento dos dados utilizadas nesse estudo.

- **Eliminação Manual de Registros:** Devido à natureza do problema estudado, diversos fatores podem contribuir para obter dados incoerentes, como desligamentos de bombas hidráulicas para manutenção preventiva ou corretiva. Com apoio de um profissional da área, foi realizado a eliminação de registros do conjunto de dados que apresentam valores menores de 2500 m³ a cada hora ou mais de 4000 m³ a cada hora, com objetivo de eliminar dados incoerentes.
- **Limpeza dos dados com erros de telemetria:** Antes de criar o conjunto de dados, foi realizado um filtro na consulta ao DW, onde apenas dados que foram capturados com boa qualidade de sinal na telemetria foram buscados. Dessa forma, grandes ruídos, redundâncias e zeros foram limpos do conjunto de dados.

O script também é responsável por importar os dados coletados, transformar e agregar os dados em intervalos de tempo adequados para análise, separar a base de treino e de teste, treinar o modelo e avaliar seus resultados. Nesse estudo, com objetivo de realizar uma análise experimental, a mesma base de treino foi submetida a 2 modelos PROPHET:

- **PROPHET A:** Utilizado o PROPHET na base de treino com os parâmetros: growth='linear', seasonalitymode='additive'
- **PROPHET B:** Utilizado o PROPHET na base de treino com os parâmetros: growth='linear', seasonalitymode='multiplicative'

No modelo A, denominado "PROPHET A", os parâmetros foram definidos como growth='linear' e seasonalitymode='additive'. O parâmetro 'growth' especifica o modelo de crescimento da tendência temporal, onde 'linear' indica um crescimento linear ao longo do tempo. Já o parâmetro 'seasonalitymode' determina como a sazonalidade é tratada pelo modelo, com 'additive' indicando que a sazonalidade é adicionada ao modelo de forma linear. Por outro lado, no modelo B, denominado "PROPHET B", os parâmetros foram definidos como growth='linear' e seasonalitymode='multiplicative'. Assim como no PROPHET A, o parâmetro 'growth' especifica um crescimento linear da tendência temporal. No entanto, o parâmetro 'seasonalitymode' foi configurado como 'multiplicative', implicando na sazonalidade modelada de forma multiplicativa, ou seja, a amplitude da sazonalidade varia de acordo com a tendência dos dados. Essas configurações influenciam a maneira como o modelo PROPHET captura e incorpora a tendência e a sazonalidade nos dados de séries temporais, resultando em diferentes comportamentos e performances de previsão.

Na Etapa 03, avaliação dos modelos, ocorre após a separação das bases de treino e teste. Adotou-se a estratégia de divisão dos dados em duas partes distintas. Para a base de treino, foi aplicado um corte com 2/3 dos dados, destinando-se essa porção de dados ao treinamento dos modelos PROPHET A e PROPHET B. Por sua vez, a base de teste foi composta pelos 1/3 restantes do dataset, reservada para a avaliação do desempenho dos modelos e subsequente cálculo das métricas de avaliação. Nessa etapa, foram computadas as seguintes métricas de avaliação:

Tabela 1. Métricas de Avaliação de Modelos de Séries Temporais

Métrica	Nome da Métrica	Fórmula
MAE	Erro Absoluto Médio	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
MAPE	Erro Percentual Absoluto Médio	$\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right) \times 100\%$

Na Tabela 1, as variáveis das métricas apresentadas, são definidas como:

- n é o número de observações
- y_i é o valor observado no tempo i
- \hat{y}_i é o valor previsto no tempo i

Para as métricas de avaliação utilizadas, o Erro Médio Absoluto (MAE) é uma métrica estatística utilizada para avaliar a precisão de um modelo de previsão ou de estimativa. Ele calcula a diferença absoluta entre os valores observados e os valores previstos pelo modelo, em seguida, calcula a média dessas diferenças. O MAE é uma medida simples e direta de erro, indicando o quão próximo as previsões estão dos valores reais, independentemente de sua direção.

Por outro lado, o Erro Percentual Absoluto Médio (MAPE) é uma medida de erro relativo frequentemente usada em previsão de séries temporais e em outras aplicações de previsão. O MAPE calcula a média percentual das diferenças absolutas entre os valores

observados e os valores previstos, expressando o erro como uma porcentagem do valor real. Essa métrica é útil para entender a precisão de um modelo em termos percentuais, fornecendo uma visão relativa do desempenho do modelo em diferentes cenários.

4. Resultado e Discussão

Com o objetivo de analisar o comportamento da série temporal de volume disponibilizado de água tratada no município estudado, utilizou-se os dados diários do período de 07/09/2023 a 15/12/2023, formando uma série de aproximadamente 4 meses, com 27.401 observações correspondentes a um período de 5 minutos cada. Para uma análise prévia gerou-se, inicialmente, um gráfico da série original. Essa primeira observação é importante para identificar possíveis discrepâncias ou erros que possam existir na extração dos dados, bem como entender, superficialmente, como é o comportamento da curva em relação ao tempo estudado.

Tabela 2. Dataset Original - a cada 5 minutos

ds (5 minutos)	y (m³)
2023-09-07 06:30:00	336.0
2023-09-07 01:25:00	336.0
2023-09-07 03:00:00	280.0
2023-09-07 00:55:00	320.0
2023-09-07 02:50:00	328.0
...	...
2023-12-15 01:00:01	328.0
2023-12-15 01:10:00	288.0
2023-12-15 00:50:00	280.0
2023-12-15 00:30:00	368.0
2023-12-15 01:20:00	328.0

Observando a Tabela 3, para a coluna 'ds', temos o carimbo de data e hora, na coluna 'y' temos o valor do volume de água disponibilizada, em metros cúbicos, a cada 5 minutos. A série assume um comportamento estacionário, pois não apresenta uma tendência bem definida e picos muito elevados. Como existe uma variação muito baixa dos dados com uma frequência de 5 mim, foi realizado um agrupamento no conjunto de dados. Nessa linha, foi realizada uma transformação, obtendo uma nova série. Pela natureza do problema, dados a cada 5 minutos possuem uma granularidade baixa. Devido a isso, foi realizado um agrupamento dos dados para que a série fique com período de tempo de 1 hora (série horária). Após a transformação da frequência, a série passou a apresentar uma distribuição de dados um pouco maior, sendo possível observar alguns comportamentos. Os dados da série foram novamente observados, de acordo com a Tabela 3.

Observa-se que a serie horária (Tabela 3) possui uma média aritmética de 3443,89 m³ de água a cada 1 hora, com um desvio padrão de 1025,35. Os valores mínimos e máximos foram fixados conforme a metodologia. Para prosseguir com a análise, o conjunto de dados, com registros a cada hora, foi dividido em dois subconjuntos distintos. O primeiro subconjunto, designado como conjunto de treinamento, compreende 2/3

Tabela 3. Dataset da série horaria - a cada 1 hora

ds (1 hora)	y (m ³)
2023-09-07 00:00:00	3648.0
2023-09-07 01:00:00	3720.0
2023-09-07 02:00:00	3680.0
2023-09-07 03:00:00	3704.0
2023-09-07 04:00:00	3736.0
...	...
2023-12-15 09:00:00	3464.0
2023-12-15 10:00:00	3352.0
2023-12-15 11:00:00	3328.0
2023-12-15 12:00:00	3360.0
2023-12-15 13:00:00	3304.0

das observações disponíveis totalizando 1535 registros, Já o segundo subconjunto, representando 1/3 restantes, sendo composto por 766 registros, será destinado aos testes e validações subsequentes, conforme descrito na seção de métodos da pesquisa. Essa estratégia de divisão permite que o modelo seja treinado em uma parcela dos dados e, em seguida, testado em um conjunto independente.

Tabela 4. Tabela de Treino

n	ds	y
0	2023-09-07 00:00:00	3648.0
1	2023-09-07 01:00:00	3720.0
2	2023-09-07 02:00:00	3680.0
3	2023-09-07 03:00:00	3704.0
4	2023-09-07 04:00:00	3736.0
...
1531	2023-11-12 03:00:00	3504.0
1532	2023-11-12 04:00:00	3552.0
1533	2023-11-12 05:00:00	3528.0
1534	2023-11-12 06:00:00	3696.0
1535	2023-11-12 07:00:00	3544.0

Tabela 5. Tabela de Teste

n	ds	y
1536	2023-11-12 08:00:00	3504.0
1537	2023-11-12 09:00:00	3488.0
1538	2023-11-12 10:00:00	3408.0
1539	2023-11-12 11:00:00	3376.0
1540	2023-11-12 12:00:00	3360.0
...
2298	2023-12-15 09:00:00	3464.0
2299	2023-12-15 10:00:00	3352.0
2300	2023-12-15 11:00:00	3328.0
2301	2023-12-15 12:00:00	3360.0
2302	2023-12-15 13:00:00	3304.0

A Tabela 4 apresenta os dados de treinamento, os quais consistem em observações horárias de uma série temporal. Cada linha representa uma observação, com as colunas indicando o índice da observação (n), a data e hora da observação (ds), e o valor observado (y). Esses dados são utilizados para treinar o modelo de previsão e são essenciais para ajustar os parâmetros do modelo com base nos padrões históricos da série temporal. Por outro lado, a Tabela 5 mostra os dados de teste, que são usados para avaliar o desempenho do modelo treinado. Assim como na tabela de treinamento, cada linha representa uma observação horária, com as mesmas colunas de n , ds e y . No entanto, esses dados não são usados durante o treinamento do modelo; em vez disso, são utilizados para verificar o quão bem o modelo é capaz de generalizar para novos dados não vistos. Essa etapa é crucial para avaliar a capacidade do modelo de fazer previsões precisas e úteis em

situações do mundo real.

Tabela 6. Resultados das previsões dos modelos PROPHET A e PROPHET B.

ds	y	yhat_PROPHET_A	yhat_PROPHET_B
2023-11-12 08:00:00	3504.0	3493.9	3491.8
2023-11-12 09:00:00	3488.0	3413.3	3410.4
2023-11-12 10:00:00	3408.0	3348.7	3345.1
2023-11-12 11:00:00	3376.0	3320.9	3317.1
2023-11-12 12:00:00	3360.0	3323.4	3319.7
...
2023-12-15 09:00:00	3464.0	3771.1	3768.5
2023-12-15 10:00:00	3352.0	3703.9	3698.1
2023-12-15 11:00:00	3328.0	3672.4	3665.2
2023-12-15 12:00:00	3360.0	3670.2	3663.0
2023-12-15 13:00:00	3304.0	3674.9	3668.0

A tabela 6 apresenta os resultados das previsões dos modelos PROPHET A e PROPHET B em comparação com os valores reais do conjunto de teste (y). Ao analisar os dados, observamos que ambas as previsões $yhat_{PROPHETA}$ e $yhat_{PROPHETB}$ estão próximas dos valores reais em algumas instâncias, mas também mostram discrepâncias significativas em outras. Além disso, podemos observar que, em geral, os valores previstos pelo modelo PROPHET A ($yhat_{PROPHET A}$) parecem estar mais próximos dos valores reais do que os valores previstos pelo modelo PROPHET B ($yhat_{PROPHET B}$). Isso sugere que o modelo PROPHET A pode ter um desempenho ligeiramente melhor na previsão dos dados do que o modelo PROPHET B. No entanto, é importante ressaltar que a avaliação completa do desempenho desses modelos requer a análise de uma variedade de métricas de avaliação, como MAE, RMSE, MAPE, entre outras. Portanto, esses resultados são apenas uma parte do processo de avaliação e devem ser interpretados juntamente com outras métricas para uma avaliação abrangente da capacidade de previsão dos modelos.

Tabela 7. Métricas de Avaliação dos PROPHET A e PROPHET B

Métricas de Avaliação	PROPHET A	PROPHET B
MAE (Erro Absoluto Médio)	133.40	131.97
MAPE (%) (Erro Médio Percentual Absoluto)	3.86	3.81

Os resultados das métricas (Tabela 7) de avaliação do PROPHET A e do PROPHET B. Ao analisar as métricas de avaliação dos modelos, podemos entender seu desempenho na previsão dos dados. O Erro Absoluto Médio (MAE) de 133.40 para o PROPHET A e 131.97 para o PROPHET B indica que, em média, as previsões desses modelos estão desviadas em aproximadamente 133.40 e 131.97 unidades, respectivamente, dos valores reais. Esses valores fornecem uma medida direta do erro absoluto das previsões em relação aos valores reais. Quanto ao Erro Percentual Absoluto Médio (MAPE), uma métrica que expressa o erro como uma porcentagem do valor real, observamos que o PROPHET A tem um valor de 3.86 e o PROPHET B tem um valor ligeiramente menor, de 3.81. Isso significa que, em média, as previsões do PROPHET A apresentam um erro

percentual absoluto de aproximadamente 3.86 em relação aos valores reais, enquanto as previsões do PROPHET B apresentam um erro ligeiramente menor de cerca de 3.81. Para melhor visualizar ambos os modelos, o perfil das séries é apresentado na Figura 2.

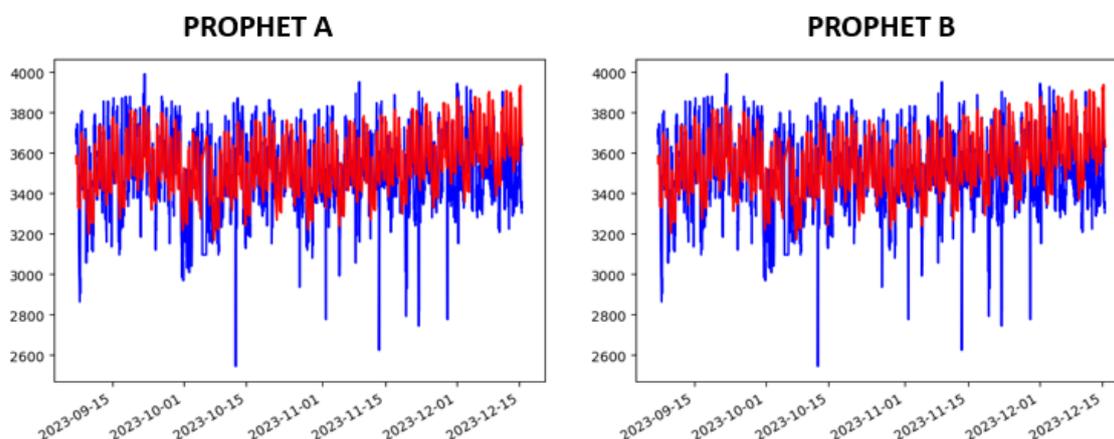


Figura 2. Resultados do PROPHET A e PROPHET B

Na Figura 2, no eixo Y é apresentado o volume distribuído em m^3 , no eixo X é apresentado o tempo em horas. Em azul, é apresentado os valores reais e em vermelho, temos o volume distribuído predito. Observa-se que há alguns pontos onde a série vermelha (predita) fica significativamente distante da série em azul, esses erros visivelmente maiores podem ser reflexo dos problemas de qualidade de sinal apresentados pela telemetria do equipamento, sendo necessário um estudo mais aprofundado para tratar essa questão.

Com base nesses resultados, é possível prever a demanda de água a ser distribuída para a população da cidade em estudo. Essa análise oferece uma alternativa aos gestores responsáveis por esse serviço para gerir melhor os recursos necessários na captação da água bruta e seu posterior tratamento. Nesse processo, recursos como energia elétrica, produtos químicos e pessoas podem ser melhor dimensionados. Um especialista no assunto poderá ser beneficiado com maiores informações para tomada de decisão na operação das estações de tratamento bem como nos reservatórios de água tratada.

Já para o cidadão, ter acesso a esse tipo de previsão pode fornecer maior transparência no acompanhamento dos serviços dessa natureza. Aplicando esse mesmo método de pesquisa, com foco nos DMC's o cidadão poderá analisar o melhor momento de poupar água, encher sua caixa d'água e/ou reservatórios bem como prevenir perdas de água em seu dia a dia. Outra possibilidade é a automatização dessa análise, fazendo com que previsões da próxima hora esteja sempre disponível para profissionais gerir melhor seus ativos responsáveis pelo saneamento básico da cidade.

5. Conclusão

Nesse trabalho, o objetivo foi apresentar uma análise experimental realizada para previsão de demanda de água potável. A partir de uma série temporal criada com base em dados reais, esse objetivo foi alcançado para uma previsão dos próximos 30 dias a cada 1 hora, para isso foi utilizada uma comparação de 2 modelos série temporal, cada um com

seu parâmetro, horária conforme descrito no método da pesquisa, e, posteriormente foi realizado análise das métricas de avaliação das previsões.

Embora as diferenças entre os modelos sejam sutis em algumas métricas, o PROPHET B demonstra consistentemente um desempenho ligeiramente superior em termos de precisão, conforme indicado pelo MAE e MAPE. Além disso, o PROPHET B exibe uma estabilidade semelhante ao PROPHET A, tornando-o uma escolha preferencial devido à sua capacidade de prever com precisão os valores futuros apresentando menor erro médio absoluto. A contribuição desse estudo beneficia tanto uma organização que atua na prestação de serviços de saneamento básico, quanto para o cidadão.

No contexto de CI, previsões de demanda de água podem beneficiar a cidade com as análise de dados e apoiar possíveis implementações de sistemas inteligentes de monitoramento. Por exemplo, sensores instalados em redes de distribuição de água podem fornecer dados contínuos sobre o consumo e a pressão da água, permitindo ajustes rápidos e precisos na operação das estações de tratamento. Com esses dados, modelos preditivos de água disponibilizada podem ser cada vez otimizados. Para uma organização, análises de previsão de demanda de água podem ser utilizadas como ferramenta de suporte e auxílio a tomada de decisão durante a operação das estações de tratamento de água. Além disso, esse tipo de previsão possibilita uma melhor alocação dos produtos químicos utilizados para o tratamento da água e melhor uso de energia elétrica nessas instalações, uma vez que a organização saberá, antecipadamente, o volume de água que deverá ser tratado. Já para o cidadão, ter acesso a esse tipo de previsão pode fornecer maior transparência no acompanhamento dos serviços dessa natureza. Sobre o PROPHET, após aplicação do modelo proposto, a vantagem que foi observada em sua aplicação foi que o modelo de previsão possui parâmetros facilmente interpretáveis que podem ser alterados para impor suposições à previsão.

Para trabalhos futuros, recomenda-se realizar comparação com outros modelos de série temporal e algoritmos de aprendizagem de máquina como, por exemplo, SARIMAX, PROPHET e XGBoost. Outro ponto a ser trabalhado é buscar uma volumetria de dados ainda maior do que o aplicado nesse estudo, quanto maior o conjunto de dados, maior será a probabilidade de captura de possíveis tendências necessárias para criar uma previsão com maior qualidade. Dessa forma, outras agregações de frequências também podem ser observadas, como frequência diária e semanal. Tais frequências não são abordadas nesse trabalho por ainda apresentarem baixo volume de dados.

Referências

- Abreu, João Paulo Maciel Marchiori, F. F. (2023). Ferramentas de avaliação de desempenho de cidades inteligentes: uma análise da norma iso 37122:2019. *PARC Pesquisa em Arquitetura e Construção*, 14(00).
- Anthopoulos, L. (2017). Smart utopia vs smart reality: Learning by experience from 10 smart city cases. *Cities*, 63:128–148.
- Banni, M., Rosseti, I., and de Oliveira, D. (2022). Hurricane: um serviço para gerência de dados de aplicações de cidades inteligentes. In *Anais do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 151–163. SBC.

- Camara, L. R. A., da Silva, D. D. S., Sales, L. L. N., Silva, D. W. S., and Pinheiro, E. M. (2019). Qualidade de vida e percepção ambiental dos moradores de comunidades rurais em são luís (ma). *Revista Brasileira de Educação Ambiental (RevBEA)*, 14(1):263–274.
- Chen, Tianqi, G. C. (2016). Xgboost: a scalable tree boosting system proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016: 785-794. *ACM, New York, NY*.
- Cruz, M. M., Bernardini, F. C., Nunes, V. T., and Bastos, C. A. (2022). Uma análise empírica dos planejamentos estratégicos das capitais brasileiras com foco na construção de cidades inteligentes. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 157–167. SBC.
- de Assis Cunha, I. B., Baracho, R. M. A., et al. (2019). Dados abertos e suas aplicações em cidades inteligentes. *Liinc em Revista*, 15(2).
- Giffinger, R., Fertner, C., Kramar, H., Meijers, E., et al. (2007). City-ranking of european medium-sized cities. *Cent. Reg. Sci. Vienna UT*, 9(1):1–12.
- Kimball, R. and Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley, Indianapolis, IN, 3 edition.
- Liu, C., Liu, Z., Yuan, J., Wang, D., and Liu, X. (2024). Urban water demand prediction based on attention mechanism graph convolutional network-long short-term memory. *Water*, 16(6):831.
- Sajjanshetty, A. S., Jayanth, V., Mohan, R., Pahari, S., and Deepti, C. (2023). Estimation of community water consumption using multivariate ensemble approach. In *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, volume 1, pages 1–5. IEEE.
- Schwetschenau, S. E., VanBriesen, J. M., and Cohon, J. L. (2019). Integrated multiobjective optimization and simulation model applied to drinking water treatment placement in the context of existing infrastructure. *Journal of Water Resources Planning and Management*, 145(11):04019048.
- Taylor, S. J. and Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1):37–45.
- Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.