

Processos Gaussianos na Predição da Movimentação de Carga no Porto de Santos

Gaussian Processes in Predicting Cargo Movement at the Port of Santos

SANTOS, Dyanna Cruz dos ¹
CAMPOS, Renato Gomes de²
CAMBLOR, Guilherme Rodrigo³
OLIVEIRA, Alexandre Garcia de⁴

RESUMO

O Porto de Santos, em São Paulo, Brasil, é essencial para o comércio marítimo internacional. Este estudo examina a aplicação de processos gaussianos na análise preditiva da movimentação de carga no porto, utilizando dados históricos e conhecimento especializado. A pesquisa comparou a eficácia dos processos gaussianos com métodos tradicionais como médias móveis e regressão linear simples, destacando as vantagens dos processos gaussianos em reduzir incertezas e lidar com padrões sazonais complexos.

Foram analisados dados de 2005 a 2022 para prever as tendências de movimentação de carga para 2023. Os resultados mostram que os processos gaussianos oferecem previsões robustas, no entanto, a regressão linear em cima dos meses agrupados superou os resultados novos obtidos. Esses achados contribuem para a otimização das operações portuárias e a tomada de decisões estratégicas baseadas em dados.

PALAVRAS-CHAVE: Análise preditiva, processos gaussianos, modelos probabilísticos, Porto de Santos e movimentação de carga.

ABSTRACT

The Port of Santos, in São Paulo, Brazil, is essential for international maritime trade. This study examines the application of Gaussian processes in the predictive analysis of cargo movement at the port, using historical data and specialized knowledge. The research compared the effectiveness of Gaussian processes with traditional methods such as moving averages and simple linear regression, highlighting the advantages of Gaussian processes in reducing uncertainties and handling complex seasonal patterns.

Data from 2005 to 2022 were analyzed to predict cargo movement trends for 2023. The results show that Gaussian processes offer robust predictions; however, linear regression on grouped months outperformed the new results obtained. These findings contribute to the optimization of port operations and data-based strategic decision-making.

KEYWORDS: Predictive analysis, Gaussian processes, probabilistic models, Port of Santos, and cargo movement.

¹ Eng. Civil e Cientista de Dados, FATEC Rubens Lara, dyanna.santos@fatec.sp.gov.br

² Eng. Mecânico e Cientista de Dados, FATEC Rubens Lara, renato.campos3@fatec.sp.gov.br.

³ Dev. Full Stack e Cientista de Dados, FATEC Rubens Lara, guilherme.santos304@fatec.sp.gov

⁴ PhD, Ciência da Computação, Professor Associado, FATEC Rubens Lara, alexandre.oliveira@fatec.sp.gov.br

1 INTRODUÇÃO

O Porto de Santos, localizado no estado de São Paulo, Brasil, é uma peça fundamental no cenário logístico e econômico nacional, desempenhando um papel crucial no comércio marítimo internacional. Sua relevância estratégica como o principal porto da América Latina impulsiona a constante busca por métodos inovadores que otimizem suas operações e antecipem cenários futuros.

Nesse contexto, a análise preditiva por meio de processos gaussianos se destaca, proporcionando uma abordagem probabilística robusta para a previsão de eventos futuros a partir de dados históricos e conhecimento especializado (Rasmussen & Williams, 2006). Estudos recentes apontam sua eficácia, destacando os processos gaussianos como uma técnica cada vez mais reconhecida e aplicada com sucesso em diversos campos (Roberts et al., 2013; Williams et al., 2017).

Este artigo examinará a aplicação de processos gaussianos no contexto portuário, como uma continuidade de estudos anteriores que utilizaram médias móveis e regressão linear simples. Será realizada uma comparação detalhada para avaliar os benefícios adicionais proporcionados pelos processos gaussianos (Rasmussen & Williams, 2006).

Embora o processo gaussiano ofereça vantagens, como a capacidade de focar na redução da incerteza de maneira localizada ao longo da distribuição de dados, também apresenta desafios. Entre eles, destacam-se a complexidade do modelo e a necessidade de escolher uma função de covariância (ou kernel) que mais se adeque ao problema (Rasmussen & Williams, 2006).

1.1 JUSTIFICATIVA

A implementação de análise preditiva com processos gaussianos na movimentação de carga no Porto de Santos justifica-se pela sua relevância estratégica e operacional. Essa técnica possibilita a identificação de melhorias significativas, ao mesmo tempo em que enriquece substancialmente o conhecimento em probabilidade e estatística. Com essa análise, espera-se promover aprimoramentos nas operações portuárias e oferecer suporte para decisões mais fundamentadas e estratégicas no contexto portuário.

1.2 PROBLEMA DA PESQUISA

A pesquisa consiste em identificar as tendências de movimentação de carga no Porto de Santos, analisando os picos sazonais de 2005 a 2022 e projetando as tendências para o ano de 2023.

2 OBJETIVOS DA PESQUISA

Este estudo baseia-se no artigo intitulado “Porto de Santos: Análise da Movimentação de Carga e Perspectivas Futuras”, empregando o mesmo conjunto de dados que cobre o período de 2005 a 2022. O objetivo é aproveitar os insights anteriormente desenvolvidos para projetar as futuras movimentações de carga utilizando processos gaussianos. Os objetivos da pesquisa são delineados como segue:

2.1 Objetivos Gerais

Realizar uma análise abrangente das tendências de movimentação de carga no Porto de Santos ao longo do período mencionado, com o intuito de identificar padrões recorrentes e fazer projeções para o ano subsequente.

2.2 Objetivos Específicos

- I. Prever as tendências de movimentação de carga para o ano de 2023, utilizando modelos probabilísticos avançados baseados em processos gaussianos;
- II. Avaliar o desempenho da predição por processos gaussianos em comparação com os resultados do estudo anterior, como médias móveis e regressão linear, analisando qual abordagem produz resultados mais precisos;
- III. Analisar a eficácia dos processos gaussianos em relação às abordagens tradicionais, particularmente em termos de sua capacidade de capturar e prever variações sazonais e tendências a longo prazo.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Análise de Séries Temporais

A análise de séries temporais é uma técnica estatística utilizada para estudar

padrões, tendências e comportamentos em conjuntos de dados ao longo do tempo. Ela examina sequências de pontos de dados indexados temporalmente, permitindo a identificação de consistências, padrões e inferências sobre possíveis tendências futuras. Essa abordagem é essencial para compreender fenômenos que variam ao longo do tempo, como comportamentos sazonais, ciclos e tendências emergentes. Ao aplicar a análise de séries temporais, é possível decompor os dados em componentes sazonais, de tendência e cíclicos, facilitando uma análise mais detalhada e a previsão de eventos futuros (Hamilton, 1994).

No contexto de operações portuárias, como a movimentação de cargas no Porto de Santos, essa metodologia se torna crucial para o planejamento eficaz e a gestão estratégica, fornecendo insights valiosos sobre as variações temporais e ajudando na tomada de decisões baseadas em dados.

3.2 Métodos Tradicionais

3.2.1 Médias móveis

As médias móveis são técnicas estatísticas usadas para suavizar as flutuações de curto prazo e ressaltar as tendências de longo prazo nos dados. A média móvel simples calcula a média aritmética dos dados dentro de uma janela temporal específica, realocando-se ao longo da série temporal. Em contrapartida, a média móvel exponencial atribui maior peso aos dados mais recentes, reagindo mais rapidamente às mudanças. Segundo Chatfield (2004), estas ferramentas são indispensáveis para análise de dados temporais porque facilitam a visualização de tendências ao reduzir o ruído aleatório presente nas séries temporais.

3.2.2 Regressão Linear

A regressão linear é um modelo estatístico que estabelece uma relação de dependência linear entre uma variável dependente e uma variável independente. Este modelo é frequentemente empregado em análises de séries temporais para identificar e quantificar tendências lineares nos dados. No entanto, como aponta, a regressão linear simples pode encontrar limitações ao lidar com séries temporais que exibem padrões sazonais com alta variabilidade, o que pode resultar em previsões menos precisas em contextos mais complexos como a movimentação de cargas em portos (Weisberg, 2014).

3.3 Métodos Robustos

3.3.1 Modelos Bayesianos

Os modelos Bayesianos, conforme referenciado por Gelman et al. (2013), atualizam probabilidades com base em novas evidências de forma contínua, aplicando o Teorema de Bayes. Em termos matemáticos, se $P(\theta|X)$ é a probabilidade *posterior*, $P(X|\theta)$ é a verossimilhança, $P(\theta)$ é a probabilidade *prior* e $P(X)$ é a probabilidade dos dados observados (Box, Jenkins, e Reinsel, 2015). O Teorema de Bayes é dado por:

$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{P(X)}$$

3.3.2 Processos Gaussianos

Os processos gaussianos, descritos por (Rasmussen e Williams, 2006), utilizam uma abordagem estatística probabilística onde cada conjunto de observações de uma série temporal é assumido como uma realização de um processo estocástico multivariado normal. Esta metodologia é útil para incorporar a incerteza nos dados e gerar previsões detalhadas.

Matematicamente, os processos gaussianos são descritos por uma função média $m(x)$ e uma função de covariância $k(x, x')$, onde cada ponto na série é normalmente distribuído. A função de cada ponto é descrita por:

i. Função média

$$m(x) = E[f(x)]$$

Onde $E[f(x)]$ denota o valor esperado da função f no ponto x .

ii. Função de covariância

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]$$

Onde $f(x)$ e $f(x')$ são os valores da função no ponto x e x' , e $m(x)$, $m(x')$ são os valores esperados (médias) nesses pontos, respectivamente.

3.4 Modelo Prophet

O modelo *Prophet*, desenvolvido pelo Facebook, é uma ferramenta poderosa para previsão de séries temporais. Ele é especialmente útil para dados com padrões sazonais fortes em dados históricos. O *Prophet* é robusto a valores faltantes e mudanças na tendência, e geralmente lida bem com outliers (Taylor e Letham, 2018).

Esse modelo é baseado em um modelo aditivo, onde as séries temporais $y(t)$

são decompostas em três componentes principais:

- i. Tendência $g(t)$: Modela mudanças não periódicas na série temporal.
- ii. Sazonalidade $s(t)$: Captura padrões periódicos, como diário, semanal e anual.
- iii. Feriados $h(t)$: Acomoda efeitos de feriados ou eventos específicos.

A formulação básica do modelo é demonstrada como:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Onde ϵ_t representa o erro que não é explicado pelo modelo, assumido como ruído gaussiano.

3.4.1 Componente da tendência

O componente de tendência $g(t)$ é frequentemente modelado como uma função linear ou logística, para capturar o crescimento saturante, conforme descrito abaixo:

- i. Linear

$$g(t) = k \cdot t + m$$

- ii. Logística

$$g(t) = \frac{C}{1 + e^{-k(t-m)}}$$

Onde:

k é a taxa de crescimento (inclinação da linha indicando quanto $g(t)$ aumenta por unidade de tempo);

m é o ponto de mudança *offset* (parâmetro que ajusta a localização da curva ao longo do eixo horizontal);

C é a capacidade de carga (usada para crescimento saturante, à medida que t aumenta, $g(t)$ se aproxima assintoticamente de C).

3.4.2 Componente Sazonal

No modelo *Prophet* modificado com *Transformada de Fourier*, descrito por Taylor e Letham (2018), a transformada é utilizada para decompor a componente sazonal em uma soma de funções senoidais. Isso permite ao modelo adaptar-se a padrões sazonais complexos e de múltiplas frequências. A expressão matemática para um componente sazonal modelado por Fourier é:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

Onde:

$s(t)$ representa o componente sazonal da série temporal;

N é o número de termos harmônicos na série de *Fourier*, determinando a complexidade e a granularidade dos padrões sazonais capturados;

a_n e b_n são os coeficientes para as funções cosseno e seno, respectivamente;

P é o período da sazonalidade (por exemplo, $P = 365.25$);

t é o tempo.

Esta fórmula permite que o modelo *Prophet* se adapte a padrões sazonais que podem ter múltiplas frequências, tornando-o bastante flexível e capaz de lidar com complexidades em dados de séries temporais (Taylor & Letham, 2018).

3.4.3 Componente de Feriados $h(t)$

O $h(t)$ modela impactos de feriados ou eventos específicos usando indicadores, como:

$$h(t) = \sum_i \delta_i I(t \in D_i)$$

onde:

D_i são as datas de i -ésimo feriados ou eventos;

δ_i é o efeito do i -ésimo feriado ou evento;

$I(t \in D_i)$ é uma função indicadora, sendo 1 no dia do evento e 0 caso contrário.

O modelo *Prophet* implementa técnicas de otimização numérica, com destaque para o método do gradiente descendente, a fim de afinar os parâmetros do modelo em relação aos dados históricos (Taylor & Letham, 2018). Essa abordagem envolve a estimativa dos coeficientes a_n , b_n , k , e m , entre outros, com o objetivo principal de minimizar uma função de custo, tipicamente o erro quadrático. A flexibilidade do modelo é uma de suas características mais robustas, permitindo ajustes na sensibilidade do componente de tendência conforme as mudanças observadas, o que o torna adequado para aplicação diversificada em séries temporais.

3.5 Erro absoluto médio

No âmbito dos processos gaussianos, o erro absoluto médio (MAE) é uma métrica utilizada para avaliar a precisão das previsões. A escolha apropriada do valor

de MAE dependerá da escala dos dados e do contexto específico da aplicação. De modo geral, o MAE representa a média das diferenças absolutas entre as previsões e os valores reais. Não há um “valor apropriado” universal existente para o MAE; ele deverá ser comparado com outras métricas e modelos para entender o desempenho relativo de seu valor. A interpretação do MAE deverá considerar a escala dos dados. Por exemplo, um MAE de valor 10 pode ser considerado pequeno em um contexto em que os valores variam de 0 a 1000, mas é considerado grande em um contexto em que os valores variam de 0 a 10. Além do MAE, outras métricas como o erro quadrático médio (RMSE) ou o coeficiente de determinação (R^2) podem fornecer uma visão mais completa do desempenho do modelo (Hyndman e Athanasopoulos, 2018).

3.6 Modelagem de Séries Temporais

No modelo *Prophet*, a série temporal é modelada com componentes de tendência e sazonalidade, que podem ser aditivos ou multiplicativos, dependendo da relação da sazonalidade com a magnitude da tendência. A abordagem aditiva é aplicada quando a sazonalidade é independente da tendência, e a fórmula usada é $\mu = \text{tendência} + \text{sazonalidade}$. A abordagem multiplicativa é usada quando a sazonalidade varia com a tendência, utilizando a fórmula $\mu = \text{tendência} \times (1 + \text{sazonalidade})$. Esses modelos podem ser implementados usando PyMC3 para ajustar a série temporal conforme necessário (Facebook, 2024).

4 PROCEDIMENTOS METODOLÓGICOS

Para realizar uma análise preditiva sobre a movimentação de cargas no Porto de Santos empregando processos gaussianos, adotou-se uma abordagem metodológica rigorosa, delineada em várias etapas essenciais.

Inicialmente, os dados foram obtidos através do estudo anterior que compilou registros históricos da movimentação de cargas, fornecidos pelos relatórios estatísticos mensais da Autoridade Portuária de Santos. Posteriormente, esses dados foram submetidos a um processo de pré-processamento utilizando um notebook ETL desenvolvido especificamente para esta pesquisa. Durante esta fase, utilizou-se um algoritmo para organizar, limpar e normalizar os dados, removendo inconsistências, valores ausentes e outliers, além de uniformizar as variáveis, assegurando assim a

integridade e a comparabilidade dos dados.

Uma vez preparados, os dados foram submetidos a uma análise exploratória detalhada para identificar padrões iniciais, tendências de longo prazo e sazonalidades, fornecendo insights fundamentais para a modelagem subsequente.

Os dados foram divididos em dois conjuntos: um conjunto de treinamento, que abrange o período de 2005 a 2019, e um conjunto de teste, de 2020 a 2022. Essa divisão foi estrategicamente realizada para validar a eficácia preditiva dos modelos em um contexto temporal realista.

Implementaram-se modelos baseados em processos gaussianos e comparou-se sua performance com técnicas tradicionais, como médias móveis e regressão linear simples. A acurácia das previsões foi meticulosamente avaliada através da métrica do erro absoluto médio (MAE) e do Erro Quadrático Médio (RMSE), permitindo uma análise quantitativa da precisão dos modelos.

Com os modelos ajustados, realizaram-se projeções para estimar a movimentação de cargas no Porto de Santos para o ano de 2023. Estas projeções foram comparadas entre si para avaliar não apenas a precisão, mas também a robustez dos processos gaussianos frente aos métodos mais convencionais.

Finalizou-se o estudo com uma análise crítica dos resultados, identificando desafios e oportunidades na aplicação de processos gaussianos. A eficácia desses processos foi destacada, particularmente em sua capacidade de integrar incertezas e adaptar-se a padrões sazonais complexos.

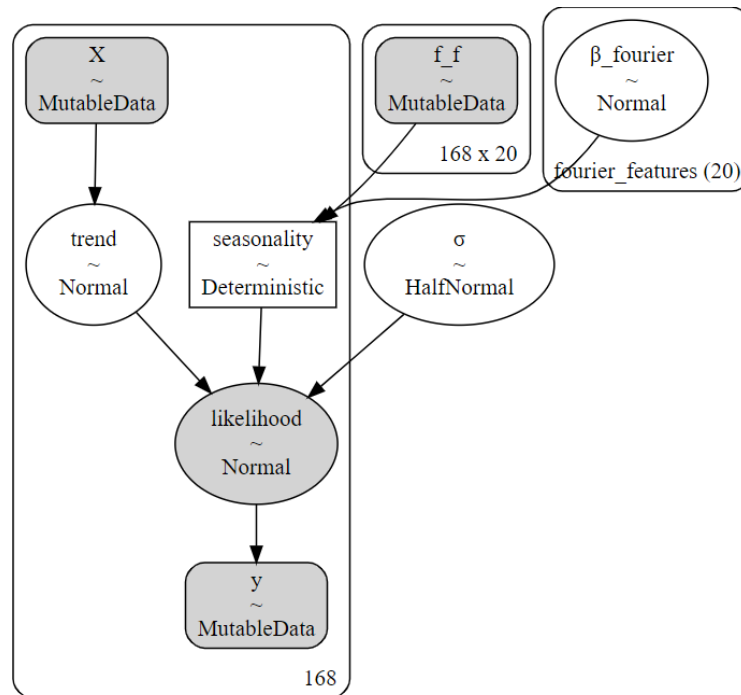
Esta etapa final enfatiza as contribuições práticas e teóricas do estudo, explorando como os insights derivados podem melhorar as operações portuárias e influenciar práticas futuras no campo da logística e gestão de portos.

4.1 Ferramentas e Softwares Utilizados

Python foi a linguagem de programação utilizada para a implementação dos modelos e análise dos dados. As bibliotecas *pandas*, *numpy*, *scikit-learn*, *arviz*, *matplotlib* e *seaborn* foram empregadas para manipulação e análise de dados, operações numéricas, implementação dos modelos por Processos Gaussianos, e visualização de dados e resultados.

5 RESULTADOS E DISCUSSÕES

5.1 Modelo *Prophet* Aditivo

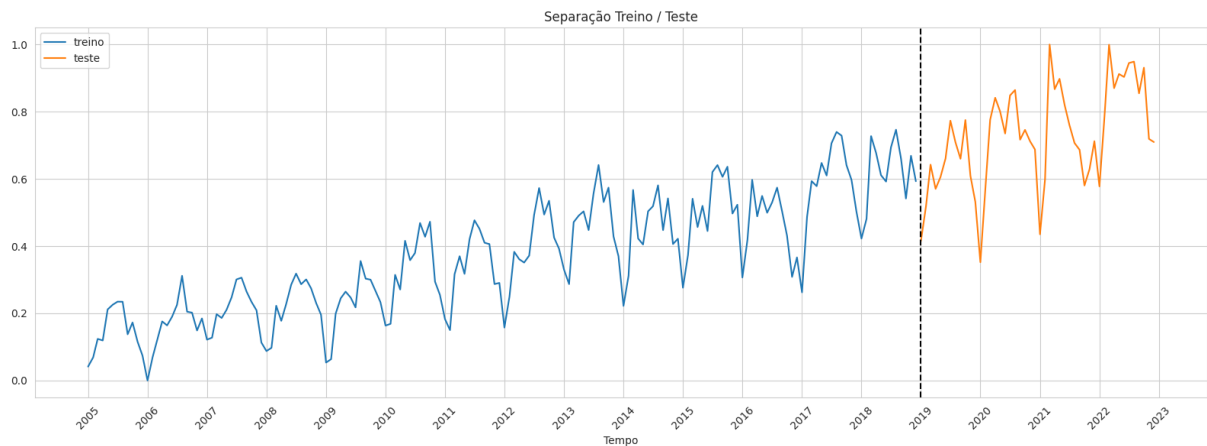


Fonte: Autores, 2024.

O gráfico abaixo apresenta a divisão dos dados em conjuntos de treino e teste ao longo do tempo, de 2005 a 2022, para a análise de movimentação de carga. A linha azul, representando o conjunto de treino, mostra flutuações na variável analisada de 2005 até 2019, indicando o período utilizado para o treinamento do modelo. A linha laranja, que representa o conjunto de teste, começa após uma linha vertical tracejada em preto, marcando o início de 2020 e continua até 2022. Este segmento é utilizado para validar a eficácia do modelo em dados não vistos anteriormente, mostrando também variações na mesma variável. O gráfico é útil para visualizar a temporalidade dos dados usados em cada fase do modelo, essencial para entender a performance do modelo em condições reais e temporais.

5.1.1 Análise Temporal da Separação da Movimentação de Carga

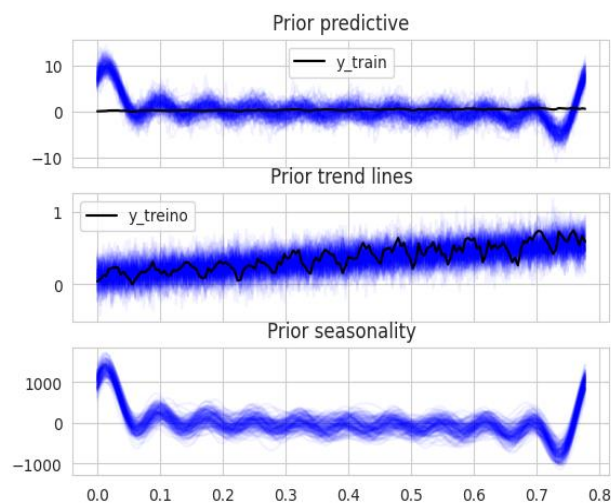
Figura 1: Análise Treino/Teste da Movimentação de Carga



Fonte: Autores, 2024.

5.1.2 Análise de Componentes de Tendência e Sazonalidade em Treino.

Figura 2: Análise Preditiva a Priori



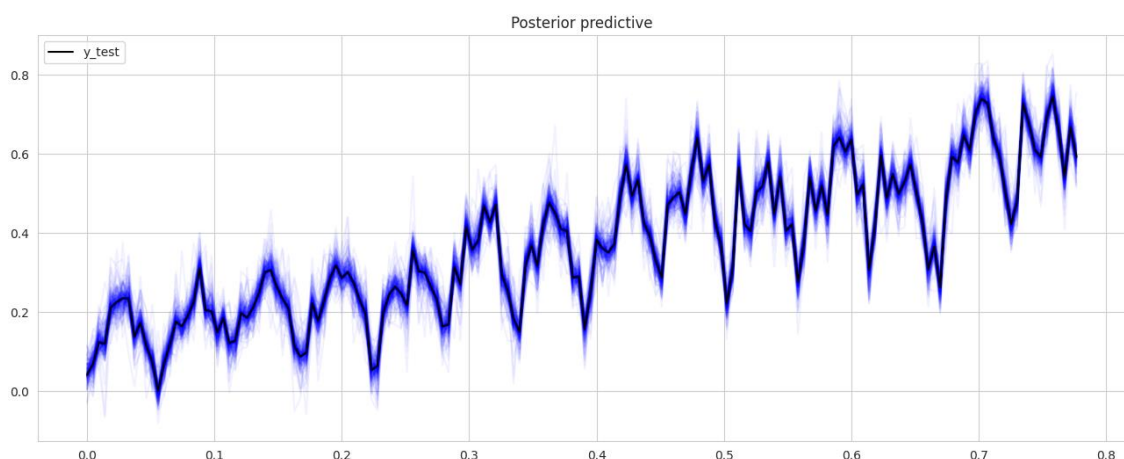
Fonte: Autores, 2024.

O gráfico acima apresenta três análises distintas dos dados de treinamento: o painel superior mostra a previsão preditiva anterior (prior) com os dados reais 'y_train', destacando a incerteza nas previsões com uma área azul. O painel do meio foca na tendência dos dados, exibindo variações da tendência através de múltiplas linhas azuis. O painel inferior ilustra os padrões sazonais, com variações sazonais indicadas pela sombra azul em torno da linha média. Cada seção do gráfico revela aspectos cruciais dos dados, abrangendo desde previsões globais até nuances detalhadas das tendências e variações sazonais.

O gráfico a seguir mostra a análise preditiva posterior para o conjunto de dados de teste `y_test`, representado pela linha preta sólida. A linha preta traça a tendência real dos dados ao longo de um período específico. Em torno desta linha, uma série de linhas azuis ilustra a distribuição das previsões geradas pelo modelo, indicando diferentes possíveis trajetórias com base nas informações aprendidas durante o treinamento. As áreas azuis mais claras destacam as regiões de maior incerteza, onde as previsões variam mais amplamente, oferecendo uma visualização clara de como o modelo avalia a variabilidade e a incerteza dos dados ao longo do tempo.

5.1.3 Análise Preditiva Posterior em Dados de Teste.

Figura 3: Análise Preditiva a Posteriori

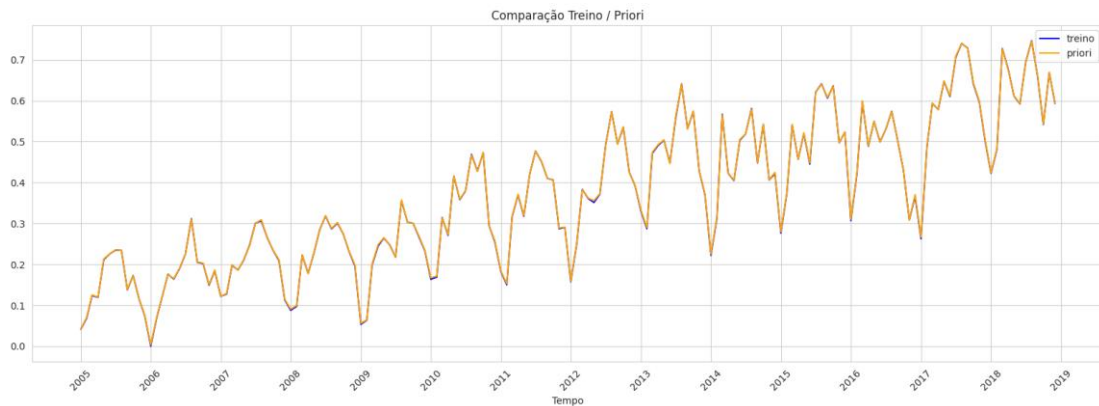


Fonte: Autores, 2024.

Este gráfico fornece uma visualização detalhada da comparação entre os dados de treino e as previsões priori, abrangendo o período de 2005 a 2019. A linha laranja, representando os dados de treino, exibe a variação real observada ao longo do tempo, incluindo picos e vales que caracterizam a dinâmica da série temporal analisada. Paralelamente, a linha azul, que representa as previsões priori geradas pelo modelo, acompanha de perto as flutuações dos dados reais. Esta sobreposição sugere que o modelo está bem ajustado e é capaz de replicar com precisão as tendências e padrões históricos.

5.1.4 Comparação das Séries Temporais de Treino e Priori.

Figura 4: Comparação Treino a Priori

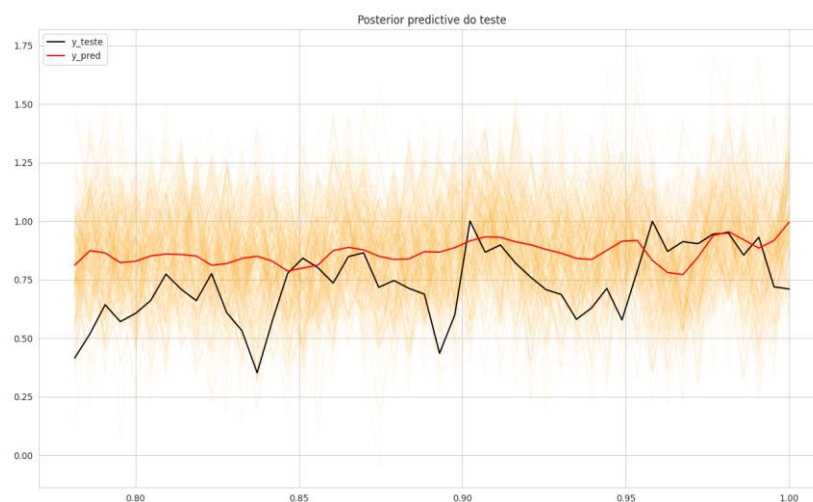


Fonte: Autores, 2024.

O gráfico a seguir exhibe a comparação entre os dados de teste reais e as previsões posteriores geradas por um modelo estatístico. A linha preta representa os valores observados no conjunto de teste, enquanto a linha vermelha ilustra as previsões do modelo para esses mesmos pontos de dados. A área em laranja, que exibe uma ampla gama de linhas, representa a distribuição das previsões simuladas, destacando a incerteza associada a cada ponto previsto.

5.1.5 Avaliação das Previsões Posteriores em Dados de Teste

Figura 5: Análise Preditiva a Posteriori do Teste



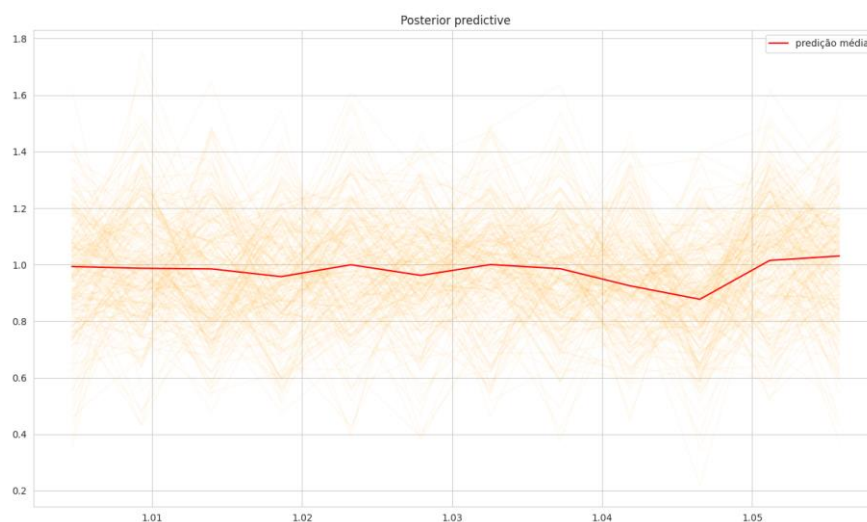
Fonte: Autores, 2024.

O gráfico abaixo mostra a previsão média e a incerteza associada para uma série temporal, focando na capacidade do modelo de gerar previsões probabilísticas. A

linha vermelha, denominada "previsão média", traça a média das previsões geradas pelo modelo ao longo de um intervalo de tempo específico, destacando a tendência central estimada pelo modelo. As linhas laranjas, que formam uma nuvem densa ao redor da previsão média, representam uma grande quantidade de trajetórias simuladas, ilustrando a distribuição das previsões possíveis e a incerteza inerente ao modelo.

5.1.6 Análise das Previsões Posteriores com Incerteza Modelada

Figura 6: Análise Preditiva a Posteriori



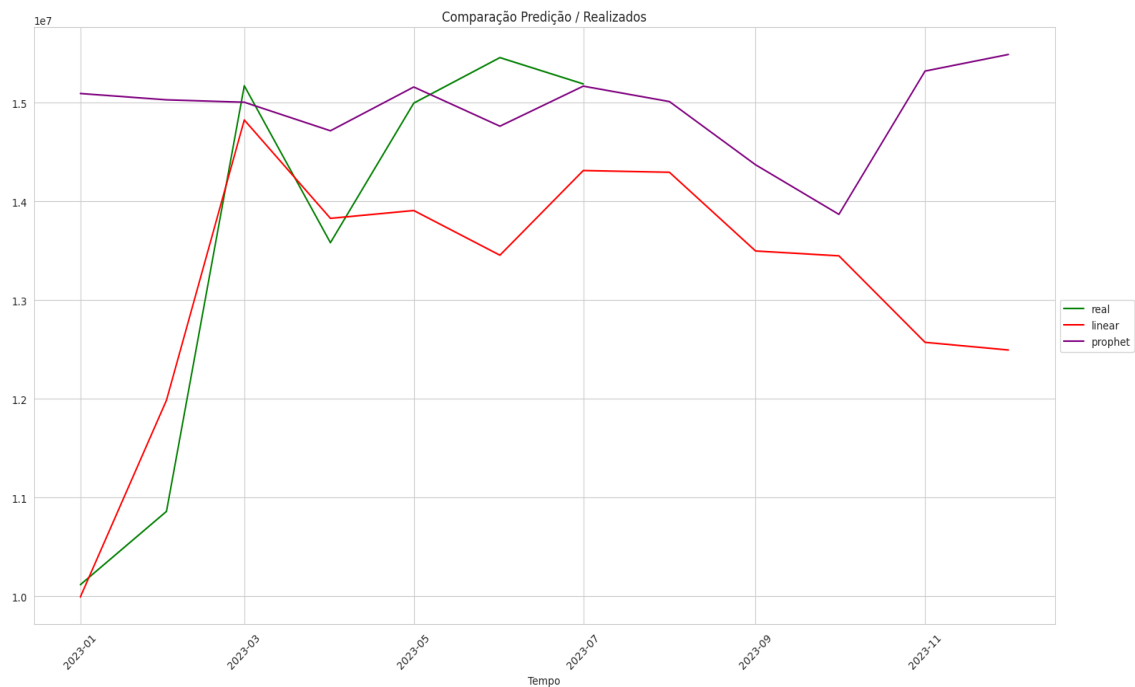
Fonte: Autores, 2024.

O gráfico abaixo apresenta uma comparação entre dados reais e as previsões de dois modelos diferentes durante o ano de 2023. A linha verde representa os valores reais observados, a linha vermelha mostra as previsões geradas por um modelo linear, e a linha roxa indica as previsões feitas pelo modelo *Prophet*.

As linhas ilustram claramente como cada modelo se comporta em relação aos dados reais. O modelo linear oferece uma abordagem mais estável, com pouca variação ao longo do tempo, enquanto o modelo *Prophet* apresenta maior flexibilidade, ajustando-se mais de perto aos picos e quedas nos dados reais, como pode ser visto nos movimentos ascendentes e descendentes mais acentuados. Esta visualização ajuda a avaliar a eficácia de cada modelo na previsão de tendências e variações temporais, fornecendo insights valiosos sobre qual modelo pode ser mais adequado dependendo das características dos dados e das necessidades de previsão.

5.1.7 Comparação de Modelos de Previsão com Dados Reais

Figura 7: Comparação da Predição com Realizados



Fonte: Autores, 2024.

5.1.8 Cálculo do MAE

MAE: 0.0013249907941508

5.1.9 Comparação do Modelo *Prophet* Aditivo

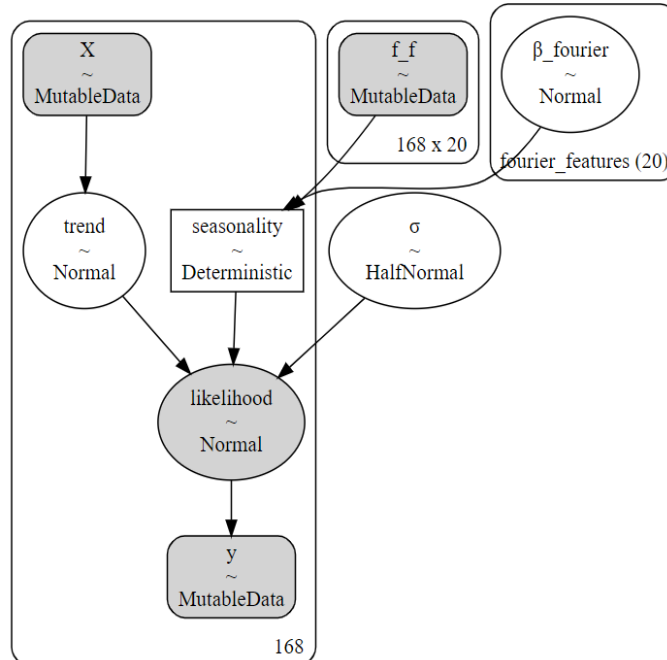
i. Erro do Modelo Linear

RMSE: 1030378.989730963 e MAE: 830471.2857142857

ii. Erro do Modelo *Prophet* FFT Aditivo

RMSE: 2447474.2479497315 e MAE: 1596268.221357349

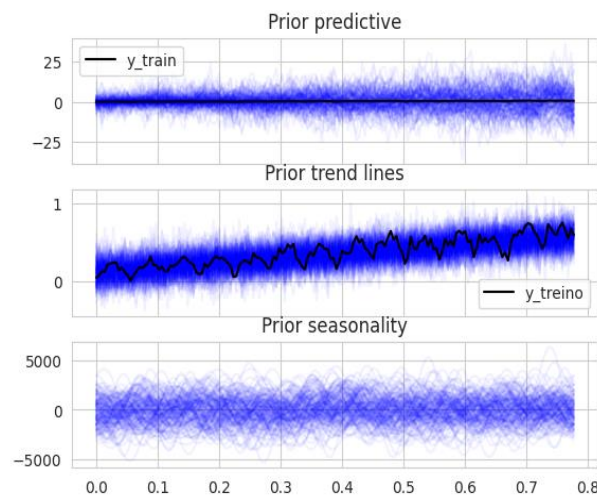
5.2 Modelo Prophet Multiplicativo



Fonte: Autores, 2024.

5.2.1 Figura 8: Análise de Tendência e Sazonalidade.

Figura 8: Análise Preditiva a Priori



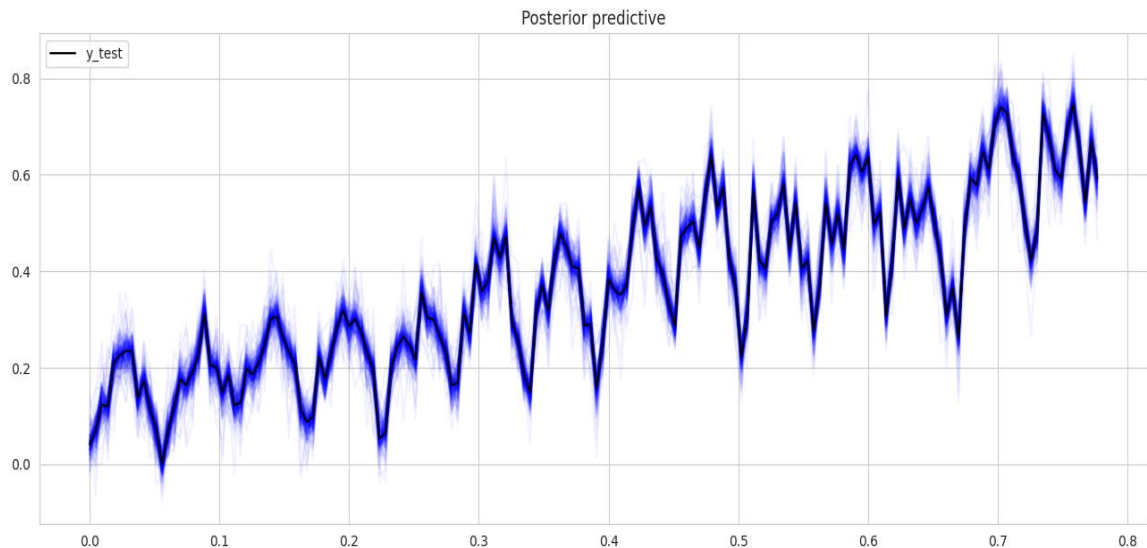
Fonte: Autores, 2024.

O gráfico acima exibe três camadas de análise de uma série temporal: a previsão preditiva com dados reais representados pela linha preta e incerteza mostrada pela nuvem azul, a tendência central marcada pela linha preta mais grossa com variações sugeridas pelas linhas azuis mais finas, e a sazonalidade, destacando os padrões periódicos através de uma densa nuvem de linhas azuis que capturam as

flutuações ao longo do tempo.

5.2.2 Figura 9: Análise de Previsão Posterior para Dados de Teste.

Figura 9: Análise Preditiva a Posteriori

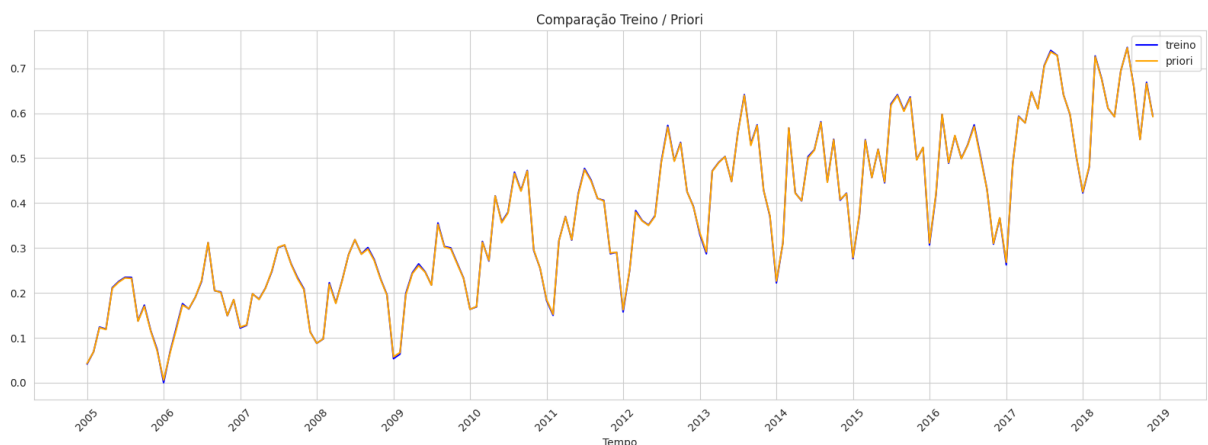


Fonte: Autores, 2024.

O gráfico acima exibe as previsões posteriores para uma série temporal comparadas aos dados reais de teste, representados pela linha preta. A linha azul e a área sombreada em torno dela mostram a média das previsões e a incerteza associada, respectivamente. As flutuações e a sobreposição indicam como o modelo se ajusta aos padrões observados, evidenciando a eficácia das previsões em capturar as variações ao longo do tempo.

5.2.3 Comparação das Previsões Priori com Dados de Treino.

Figura 10: Comparação entre Treino e a Priori



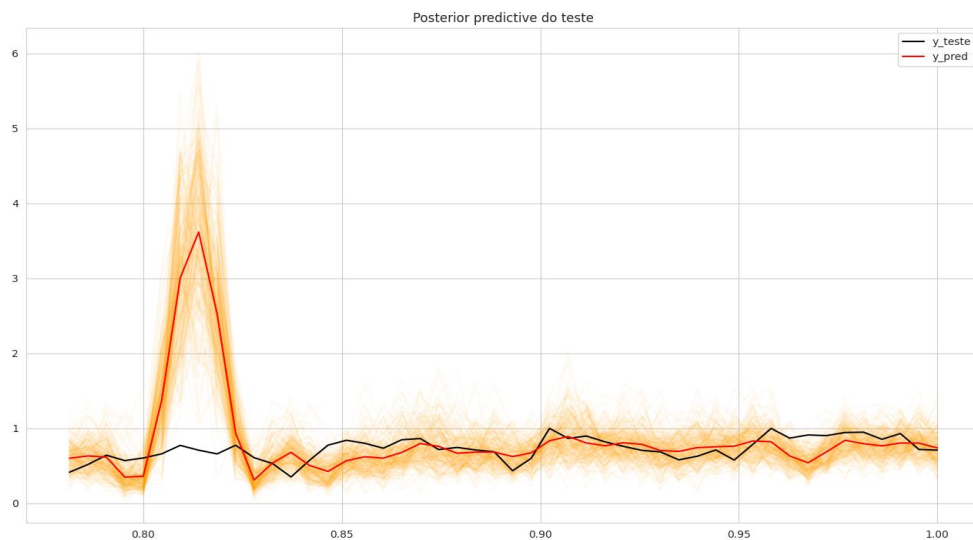
Fonte: Autores, 2024.

O gráfico acima ilustra a comparação entre os dados de treino e as previsões priori ao longo do período de 2005 a 2019. As duas linhas exibem uma correlação estreita, indicando que as previsões priori estão alinhadas com os padrões reais observados nos dados de treino. Esta consistência sugere que o modelo priori é eficaz.

5.2.4 Análise de Previsão Posterior e Incerteza em Dados de Teste.

O gráfico abaixo apresenta as previsões posteriores para um conjunto de dados de teste e suas respectivas previsões. A área laranja mostra uma ampla distribuição de possíveis trajetórias de previsão, destacando a incerteza associada a cada previsão. Um pico acentuado na linha vermelha e na nuvem de previsões em torno de 0.85 no eixo x sugere um evento ou observação anômala que o modelo tentou capturar, demonstrando como o modelo responde a variações súbitas e significativas nos dados.

Figura 11: Análise Preditiva a Posteriori

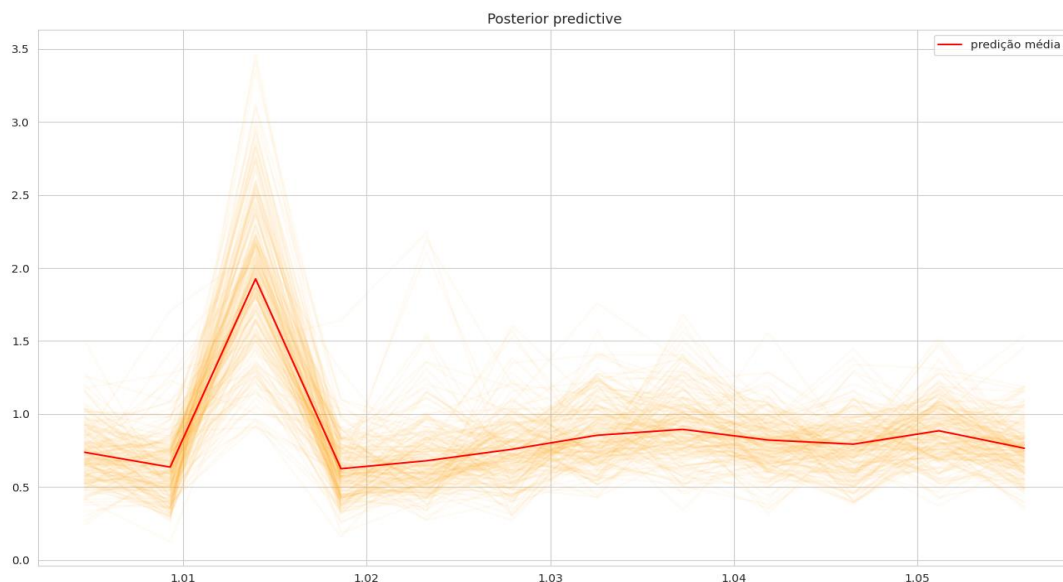


Fonte: Autores, 2024.

O gráfico a seguir mostra a previsão média e uma ampla distribuição de previsões, destacando um pico. Este pico, cercado por alta incerteza, indica uma resposta do modelo a uma possível anomalia nos dados.

5.2.5 Visualização de Previsões Posteriores e Incerteza.

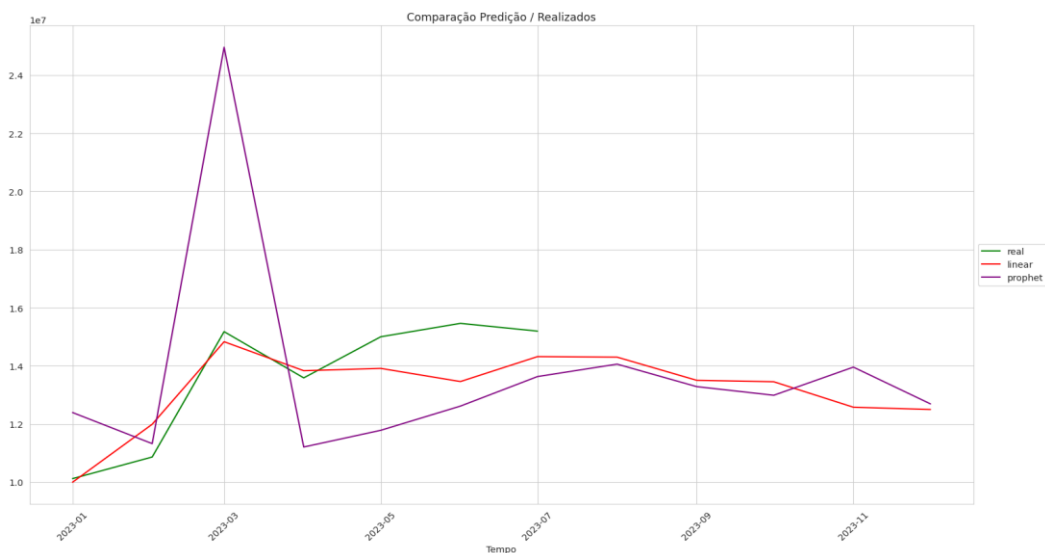
Figura 12: Análise Preditiva a Posteriori das Médias



Fonte: Autores, 2024.

5.2.6 Comparação de Modelos de Previsão com Dados Reais.

Figura 13: Comparação Previstos e Realizados



Fonte: Autores, 2024.

O gráfico acima compara os dados reais com as previsões geradas por um modelo linear e pelo modelo *Prophet* no ano de 2023. Observa-se que o modelo linear e o *Prophet* variam em sua capacidade de capturar o pico acentuado em meados do ano, com o modelo *Prophet* ajustando-se mais suavemente às variações, enquanto o

modelo linear apresenta maior divergência especialmente durante o pico.

5.2.7 Cálculo do MAE

MAE: 0.23822461987645857

5.2.8 Comparação do Modelo *Prophet* Multiplicativo

i. Erro do Modelo Linear

RMSE: 1030378.989730963 e MAE: 830471.2857142857

ii. Erro do Modelo *Prophet* FFT multiplicativo

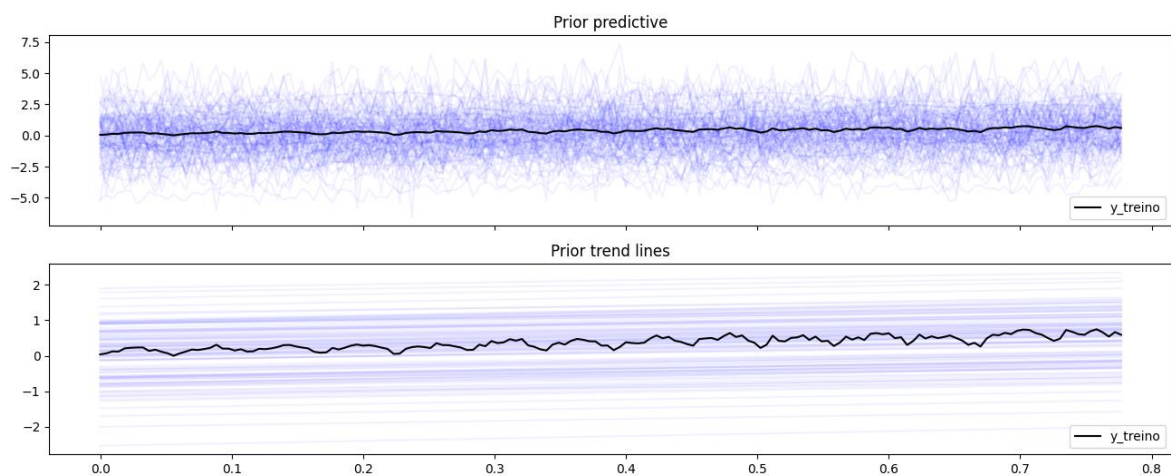
RMSE: 4892531.946507645 e MAE: 3592427.9181903144

Em construção...

5.3 Modelo *Prophet* Cosine

5.3.1 Análise Preditiva a Priori do Modelo Cosine

Figura 14: Análise Preditiva a Priori

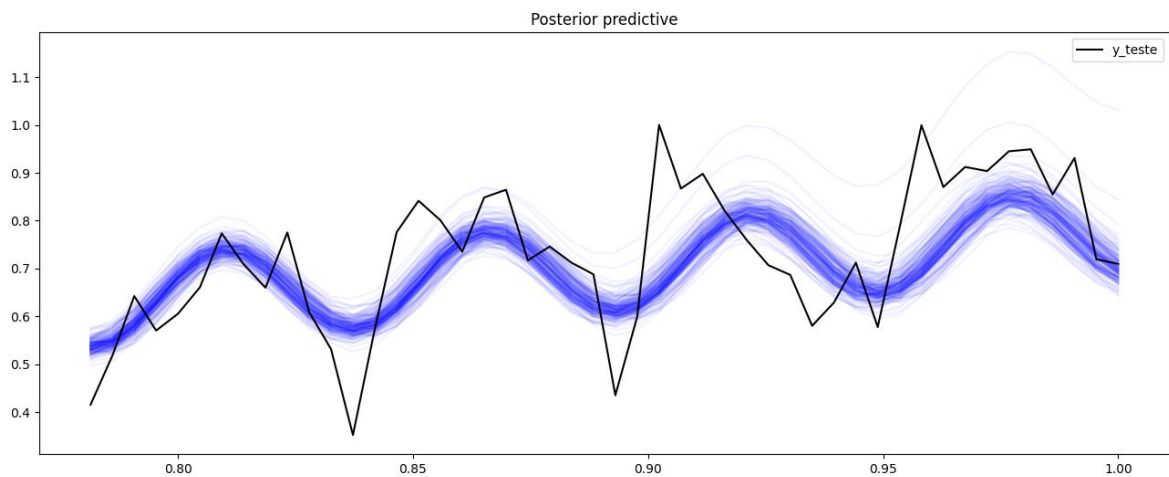


Fonte: Autores, 2024.

Escrever...

5.3.2 Análise Preditiva a Posteriori do Modelo Cosine

Figura 15: Análise Preditiva a Posteriori em Teste

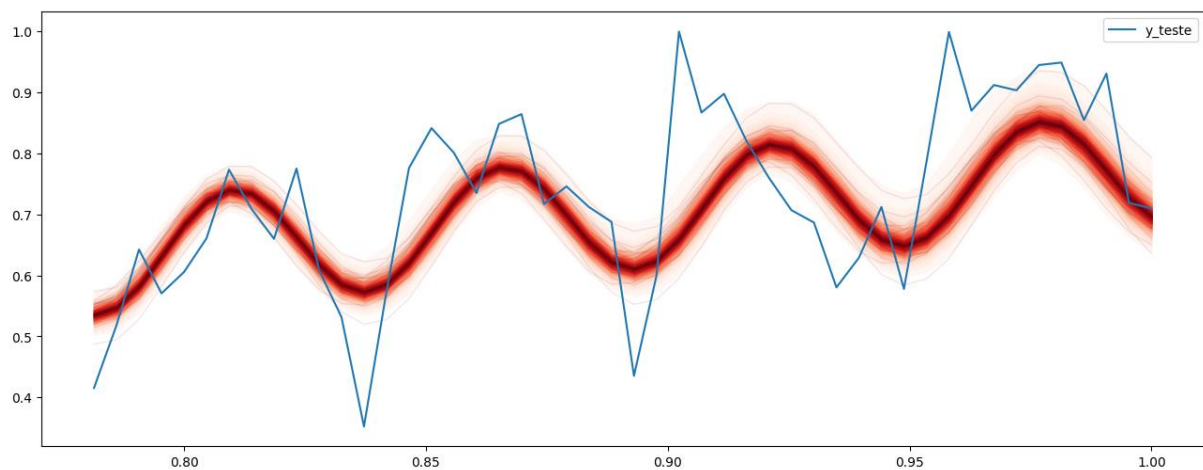


Fonte: Autores, 2024.

Escrever...

5.3.3 Análise do Conjunto de Teste

Figura 16: Análise de Teste



Fonte: Autores, 2024.

Escrever..

5.3.4 Cálculo do MAE

MAE: 0.08930173057475017

INCLUIR MODELO REGRESSION TREE (provavelmente similar a nossa regressão mensal)

INCLUIR MODELO BART

AJUSTAR MODELO COSINE E PLOTAR NOVAMENTE

6 CONCLUSÃO

Este estudo explorou uma gama abrangente de técnicas de modelagem preditiva, com o objetivo de prever e antecipar a complexa dinâmica da movimentação de cargas no Porto de Santos, o mais significativo porto da América Latina. Com uma análise detalhada de dados históricos que se estendem de 2005 a 2022, a pesquisa empregou uma abordagem comparativa, contrapondo metodologias tradicionais, como médias móveis e regressão linear simples, a métodos preditivos mais avançados, incluindo processos gaussianos. O ápice do estudo foi marcado pela implementação do modelo *Prophet*, uma estratégia inovadora destinada a otimizar a precisão das previsões futuras.

As investigações conduzidas revelaram que, embora os processos gaussianos demonstrem uma capacidade notável de fornecer previsões robustas e confiáveis, foi a regressão linear, aplicada aos dados mensalmente agrupados do estudo anterior, que superou os resultados mais recentemente obtidos. Este achado sublinha a importância crítica de uma seleção cuidadosa de métodos de análise, adaptados às características específicas dos dados e às exigências analíticas do projeto. Além disso, os resultados deste estudo são de imenso valor para a melhoria contínua das operações no Porto de Santos, oferecendo insights fundamentais que podem orientar decisões estratégicas de alto nível, baseadas em uma fundação de dados rigorosamente analisados e interpretados.

Sendo assim, este trabalho não apenas avança o campo da análise preditiva portuária, mas também serve como um marco referencial para futuras investigações e aplicações práticas no contexto da gestão portuária global.

Referências

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). John Wiley & Sons. ISBN 978-1118674925. Acesso em: 02 jun. 2024.

Chatfield, C. (2003). The Analysis of Time Series: An Introduction. CRC Press.

CRUZ DOS SANTOS, Dyanna; GOMES DE CAMPOS, Renato; SANTOS CORREA, Jobel. Porto de Santos: análise da movimentação de carga e perspectivas futuras. 2023. Semana Acadêmica, [S.l.], edição 000237, v. 11. Disponível em: <https://semanaacademica.org.br/artigo/porto-de-santos-analise-da-movimentacao-de-carga-e-perspectivas-futuras>. Acesso em: 02 jun. 2024.

Fabris, F., Magalhães, J.P.d. & Freitas, A.A. A review of supervised machine learning applied to ageing research. Biogerontology 18, 171–188 (2017). <https://doi.org/10.1007/s10522-017-9683-y>. Acesso em: 02 jun. 2024.

Facebook, Inc. Prophet: forecasting at scale. Disponível em: https://facebook.github.io/prophet/docs/quick_start.html. Acesso em: 11 jun. 2024.

Gaussian Processes for Time-Series Modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1984), 20110550.

GELMAN, A. et al. Bayesian Data Analysis. Chapman and Hall/CRC, 2013.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts. Disponível online em <https://otexts.com/fpp2/>. Acesso em: 10 jun. 2024.

Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press.

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. The American Statistician, 72(1), 37-45. DOI: 10.1080/00031305.2017.1380080.

Roberts, S., Osborne, M., Ebdon, M., Reece, S., Gibson, N., & Aigrain, S. (2013).

Williams, C. K. I., Barber, D., & Girolami, M. (2017). Machine Learning: Bayesian and Gaussian Processes. Cambridge University Press.

Weisberg, S. (2014). Applied Linear Regression (4th ed.). School of Statistics, University of Minnesota, Minneapolis, MN. John Wiley & Sons. ISBN 978-1-118-38608-8.