



De 11 a 14 - NOVEMBRO DE 2024

**Inteligência Artificial
na Gestão de Operações:
Limitações e possibilidades**



PROPOSTA DE MODELO DE PREVISÃO DE VENDA PARA UMA EMPRESA DO MERCADO DE GASOLINA PREMIUM NO BRASIL

PATRICK DA SILVA GOMES - patricksg.uerj@gmail.com
UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO - UERJ

DANIEL BOUZON NAGEM ASSAD – danielassad@eng.uerj.br
UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO - UERJ

THAÍS SPIEGEL – thais@eng.uerj.br
UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO – UERJ

PATRICIA GOMES FERREIRA DA COSTA – patricia.gomes@eng.uerj.br
UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO - UERJ

ÁREA: 3. PESQUISA OPERACIONAL
SUBÁREA: 3.6 – ANÁLISE DE DEMANDA

RESUMO: O presente trabalho tem como objetivo aplicar e mensurar modelos preditivos de séries temporais para uma empresa do mercado brasileiro de combustíveis premium baseado em séries históricas. Para atingir esse objetivo foi utilizado um conjunto de dados públicos acessados junto à Agência Nacional do Petróleo (ANP). Também se utilizou software Excel para aplicação de modelos mais simples, além da linguagem de programação Python na ferramenta/interface Jupyter utilizando bibliotecas relacionadas à ciência de dados como Pandas, Numpy, Matplotlib, Scikit-Learn, entre outras. Durante o processo de aplicação, diferentes tipos de modelos foram comparados e analisados.

PALAVRAS-CHAVES: Previsão; Combustíveis; Séries Temporais; Machine Learning.

1. INTRODUÇÃO

O mercado de combustíveis *premium* normalmente atende a um público de luxo, oferecendo produtos que não apenas garantem maior potência e eficiência de motores de alto desempenho, mas também promovem a longevidade e cuidados avançados com os veículos. Este segmento é atraído pelos benefícios dos combustíveis de alta octanagem e aditivos especializados que limpam e protegem componentes críticos do motor (Petrobras, 2024), e valoriza a percepção de status e qualidade conferida por esses combustíveis. Desta forma, o mercado de combustíveis *premium* não só atende a requisitos técnicos avançados, mas também atende aos desejos dos consumidores por exclusividade e luxo.

Segundo dados coletados pela ANP (2024), atualmente no Brasil, três grandes empresas dominam o mercado de distribuição de combustíveis *premium*. Essas empresas divulgam trimestralmente seus resultados aos investidores e, portanto, precisam estimar as vendas do próximo trimestre para encontrar estratégias que possibilitem melhorar os lucros e ganhar maior participação de mercado.

Com essa missão desafiadora, as empresas do setor buscam reduzir seus custos operacionais, aumentar a velocidade de entrega e melhorar a qualidade dos produtos ofertados. Por essas razões, um modelo de previsão mais preciso auxiliaria significativamente no planejamento estratégico das empresas, desde a compra de insumos até a entrega aos clientes e a redução dos preços dos seus produtos, especificamente a gasolina *premium*.

Diante desse contexto, este trabalho se propôs a atingir dois objetivos principais, quais sejam: (1) Encontrar modelos para previsão do mercado de gasolina *premium* no Brasil; e (2) Comparar os modelos encontrados e determinar o melhor tipo de modelo para previsão.

A proposta desse trabalho é aplicar as técnicas de previsão que proporcionem as respostas mais acuradas, contudo, uma das limitações é que os dados disponíveis para análise consistem apenas em informações sobre a comercialização de gasolina *premium* no Brasil, coletadas de 2021 até março de 2024.

Este estudo examina cinco modelos univariados de previsão para comparação: regressão linear, médias móveis, ARIMA e SVM (máquinas de vetores de suporte). Como os dados históricos são fornecidos pela Agência Nacional do Petróleo (ANP), logo, confiaremos a eles a veracidade das informações.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Mercado de Combustíveis no Brasil

Segundo dados coletados da ANP (2024), o mercado de combustíveis no Brasil tem

apresentado um crescimento constante com exceção ao ano de 2020 (devido à pandemia da COVID-19). A cadeia de suprimentos do mercado de combustíveis é fundamental para o funcionamento da economia global e possui características únicas devido à complexidade, volume e importância estratégica dos produtos envolvidos. Essa cadeia começa com a exploração e extração de petróleo bruto, passando pela sua refinaria em diferentes tipos de combustíveis, como gasolina, diesel e querosene. Após o refino, os combustíveis são transportados, geralmente por oleodutos, navios-tanque ou caminhões, até depósitos e distribuidoras. Finalmente, chegam aos postos de serviço, onde são disponibilizados ao consumidor final.

2.1.1 Combustíveis Claros

Os combustíveis claros são uma categoria crucial na matriz energética, especialmente para o transporte, devido à sua aparência mais limpa e propriedades menos densas em comparação com outros tipos de combustíveis pesados. Incluem a gasolina, amplamente utilizada em veículos de passeio e motocicletas pela sua eficiência em motores de combustão interna; o diesel, escolhido para veículos de transporte pesado como caminhões e ônibus, por sua maior densidade energética; o querosene, essencial no transporte aéreo e também utilizado em sistemas de aquecimento e iluminação; e o etanol, um biocombustível produzido a partir da cana-de-açúcar, que serve tanto como combustível puro quanto misturado à gasolina, promovendo sustentabilidade e reduzindo a dependência de fontes fósseis. Esses combustíveis são fundamentais para a mobilidade moderna, atendendo a diferentes necessidades e aplicações dentro da infraestrutura global de transporte, e são vitais para alcançar eficiência energética e reduzir emissões, alinhando-se com metas globais de sustentabilidade ambiental.

Dentro da categoria de combustíveis, diversas empresas desempenham o papel crucial de transportar os produtos das refinarias até os postos de venda, onde são comercializados para o consumidor final. Embora várias empresas atuem na distribuição desses produtos, apenas três delas detêm uma parcela significativa do mercado brasileiro.

2.1.2 Principais Produtos e Combustíveis Premium

Dentre os combustíveis claros mencionados, destacam-se as categorias especiais que oferecem benefícios adicionais em relação aos principais tipos de combustível: gasolina, etanol e diesel. Essas categorias incluem os combustíveis aditivados e os combustíveis *premium*.

Os combustíveis aditivados contêm aditivos especiais projetados para maximizar o desempenho dos motores, ajudando na limpeza dos componentes internos, como injetores e válvulas, melhorando a eficiência do combustível e reduzindo a emissão de poluentes. Eles

prolongam a vida útil do motor e mantêm uma performance mais consistente ao longo do tempo.

Os combustíveis *premium* oferecem vantagens ainda maiores, como uma octanagem mais alta, ideal para veículos de alto desempenho que requerem uma combustão mais eficiente para maximizar a potência e a resposta do motor. Eles também incluem aditivos que promovem a limpeza do motor e otimizam a queima de combustível, resultando em uma operação mais suave e eficiente.

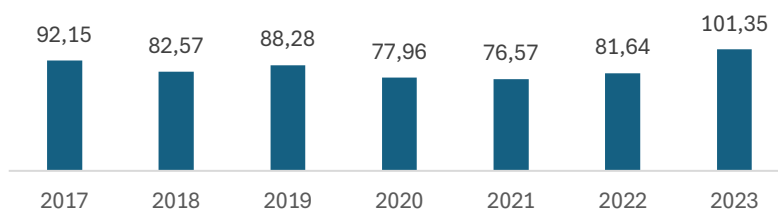
Essas categorias são formuladas para atender às necessidades de consumidores que buscam melhor rendimento e maior preocupação ambiental. No Brasil, conforme dados da ANP coletados entre 2017 e 2023, os produtos comuns representam 86,84% do mercado, os produtos aditivados compõem 13,09%, e os produtos *premium* representam apenas 0,07%.

No Brasil, os combustíveis *premium*, como a gasolina *premium* (Podium e Shell V-Power) e o diesel *premium*, são formulados com aditivos especiais que melhoram o desempenho e protegem os motores. A gasolina *premium* possui uma octanagem mais alta (95 a 100 octanas), o que melhora o desempenho, especialmente em veículos de alta performance. Este tipo de gasolina também contém aditivos detergentes que ajudam a limpar o sistema de combustão e melhorar a eficiência do motor. O diesel *premium* possui uma formulação que melhora a lubricidade, reduz o consumo de combustível e diminui as emissões de poluentes, indicado para veículos pesados como caminhões e ônibus.

Os combustíveis *premium* oferecem benefícios significativos em comparação aos combustíveis regulares, como a melhoria do desempenho do motor, redução da emissão de poluentes e aumento da eficiência energética. O crescimento do mercado de combustíveis *premium* no Brasil de 2017 a 2023 destaca-se na Figura 2.

FIGURA 1 - Volume anual de combustíveis premium no Brasil

Volume combustíveis Premium (Mil m³)



Fonte: elaborada pelos autores, baseada em dados adaptados da ANP(2024).

2.2 Modelos de Previsão

Modelos de previsão são ferramentas essenciais em áreas como finanças, economia, meteorologia, planejamento urbano e saúde. Eles utilizam dados históricos para prever eventos futuros. Neste estudo, foram selecionados 7 artigos de 70 encontrados no portal de Periódicos CAPES usando as palavras-chave “*Forecasting Models*”, “*Fuels*” e “*Time-Series*”.

DEY et al. (2023) utilizaram ARIMA, Regressão Linear, Regressão Exponencial, Regressão Logarítmica, Regressão Inversa e Regressão de Potência para prever a demanda de etanol na Índia, destacando ARIMA (1,1,0) como o mais preciso. KUMAR et al. (2023) usaram Regressão de Floresta Aleatória e SVM para prever a demanda de gasolina na Austrália, com SVM mostrando melhor desempenho. LI et al. (2023) combinaram Floresta Aleatória, SVM e Cointegração para prever preços da gasolina, concluindo que o modelo misto é mais preciso. WANG et al. (2023) propuseram um modelo de decomposição-ensemble para previsão do consumo de gasolina, que demonstrou maior precisão.

CHEN et al. (2023) utilizaram um modelo de Tendência Duplamente Amortecida Estrutural para prever a demanda de combustível no varejo na China, destacando sua robustez em mercados voláteis. SILVA et al. (2023) avaliaram o consumo de combustível leve no Brasil usando ARIMA, VECM, SVM e Floresta Aleatória, com SVM mostrando melhor desempenho em períodos de alta volatilidade. ASSAD, CARA e ORTEGA-MIER (2023) compararam modelos univariados e multivariados de previsão de séries temporais para surtos de doenças, com NNETAR obtendo o melhor desempenho.

Os métodos mais estudados foram ARIMA e SVM. Além disso, nesta pesquisa foram utilizadas técnicas de regressão linear e médias móveis, preenchendo lacunas na literatura. A seguir, serão apresentados os principais modelos de previsão utilizados, divididos em modelos estatísticos clássicos e de aprendizado de máquina.

2.2.1 MODELOS ESTATÍSTICOS CLÁSSICOS

Os modelos estatísticos clássicos são amplamente utilizados na previsão de séries

temporais devido à sua simplicidade e eficácia. Entre os métodos mais comuns estão a regressão linear, médias móveis, suavização exponencial e ARIMA. A suavização exponencial, introduzida por Holt (1957) e Winters (1960), é eficaz para dados com tendência e sazonalidade, atribuindo pesos maiores às observações mais recentes. O modelo ARIMA, popularizado por Box e Jenkins (1970), combina componentes autorregressivos e de médias móveis para modelar séries temporais estacionárias.

2.2.1.1 Regressão Linear

A regressão linear prevê o valor de uma variável usando outra variável, assumindo uma relação linear entre elas. A equação básica é:

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (1)$$

Onde: Y_t é a variável dependente, x_t é a variável independente, β_0 é o intercepto, β_1 é o coeficiente da variável independente e ε_t é o termo de erro.

2.2.1.2 Médias Móveis

As médias móveis são usadas para analisar tendências de dados ao longo do tempo e suavizar séries temporais. A média móvel de ordem m é descrita como:

$$\bar{Y}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j} \quad (2)$$

Onde $m = 2k + 1$. Ou seja, a estimativa de tendência no momento t é obtido calculando a média dos valores da série temporal dentro de k períodos de t . As observações próximas no tempo também provavelmente terão valores próximos. Portanto, a média elimina parte da aleatoriedade dos dados, deixando um componente de ciclo de tendência suave.

2.2.1.3 ARIMA

O modelo ARIMA (AutoRegressive Integrated Moving Average) é uma das técnicas mais utilizadas na análise e previsão de séries temporais. Este modelo combina três componentes principais: autorregressivo (AR), integração (I) e médias móveis (MA). A combinação desses componentes permite que o ARIMA capture padrões complexos nas séries temporais e forneça previsões precisas. De acordo com Hyndman e Athanasopoulos (2021), os modelos ARIMA são poderosos porque podem modelar tanto a estrutura de autocorrelação de uma série temporal quanto a não estacionaridade, através da diferenciação.

O componente autorregressivo (AR) envolve o uso de observações passadas da série temporal para modelar o valor atual. A integração (I) refere-se ao número de diferenciações

necessárias para tornar a série estacionária. O componente de médias móveis (MA) modela a dependência entre uma observação e os erros de previsão passados.

A equação geral do modelo ARIMA(p,d,q) é expressa como:

$$Y_t = c + \phi Y_{t-1} + \phi Y_{t-2} + \dots + \phi Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

Onde, Y_t é o valor da série temporal no tempo t , c é uma constante, ϕ_i são os coeficientes dos termos autorregressivos, θ_i são os coeficientes dos termos de médias móveis, ε_t é o erro no tempo t , p é a ordem do termo autorregressivo, d é a ordem da diferenciação e q é a ordem do termo de médias móveis.

2.2.1.4 Suavização Exponencial

Uma outra abordagem amplamente utilizada na previsão de séries temporais é o modelo de suavização exponencial, especialmente quando os dados incluem fatores de tendência e sazonalidade. Esse modelo foi apresentado no final dos anos 1950 (Brown, 1959; Holt, 1957; Winters, 1960). O modelo de amortecimento exponencial dá maior importância às observações mais recentes, atribuindo pesos decrescentes às observações anteriores. Essa característica torna o modelo particularmente eficaz em situações em que as alterações devem ser ajustadas rapidamente quando os dados mudam.

Holt (1957) desenvolveu a suavização exponencial para lidar com séries temporais que exibem tendência linear. Sua abordagem considera duas equações principais: uma para o nível da série e outra para a tendência.

Winters (1960) estendeu esse modelo para incluir componentes sazonais, resultando no modelo de suavização exponencial tripla, ou modelo de Holt-Winters. Este modelo é composto por três equações: uma para o nível, outra para a tendência e uma terceira para a sazonalidade, permitindo a previsão de séries temporais com padrões sazonais regulares. As equações são:

$$y_{t+h|t} = (l_t + hb_t) S_{t+h-m(k+1)} \quad (3)$$

$$l_t = \alpha \frac{y_t}{S_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (4)$$

$$b_t = \beta(l_t + l_{t-1}) + (1 - \beta)b_{t-1} \quad (5)$$

$$S_t = \gamma \frac{y_t}{(l_{t-1} + b_{t-1})} + (1 - \gamma)S_{t-m} \quad (6)$$

Onde, l_t é nível suavizado no tempo t , y_t é a observação no tempo t , α é o parâmetro de suavização ($0 < \alpha < 1$), b_t é a tendência no tempo t , β é o parâmetro de suavização para a tendência ($0 < \beta < 1$), S_t é o componente sazonal no tempo t , γ é o parâmetro de suavização para

a sazonalidade ($0 < \gamma < 1$) e m é o período sazonal.

Em um estudo recente revisado por Assad et al. (2023), o modelo de Holt-Winters foi utilizado para prever casos diários de COVID-19 em diferentes regiões, demonstrando sua utilidade em cenários de saúde pública. Embora existam limitações, como a necessidade de ajustar parâmetros de suavização e a dificuldade em lidar com mudanças abruptas ou não lineares nos dados, o modelo de amortecimento exponencial continua sendo uma ferramenta valiosa para analistas e pesquisadores que buscam previsões precisas e de fácil interpretação.

2.2.2 MODELOS DE APRENDIZADO DE MÁQUINAS

Aprendizado de Máquina é uma subcategoria da inteligência artificial que se concentra no desenvolvimento de algoritmos e técnicas que permitem que os computadores aprendam e melhorem o desempenho em tarefas específicas a partir de dados, sem a necessidade de programação explícita para cada ação.

A seguir veremos 3 dos mais comuns modelos de aprendizado de máquina encontrados na literatura:

Random Forest - Random Forest é usado para regressão e classificação, construindo múltiplas árvores de decisão e agregando os resultados. A predição final é dada pela média das predições das árvores no caso de regressão ou voto majoritário no caso de classificação.

LSTM - As redes LSTM (Long Short-Term Memory) são uma variação das redes neurais recorrentes, eficazes para modelar sequências de dados. Utilizam células de memória com três portões principais: entrada, esquecimento e saída, regulando o fluxo de informações.

SVM - As máquinas de vetores de suporte (SVM) são usadas para regressão e classificação, encontrando o hiperplano que maximiza a margem de separação entre as classes. A função objetivo é minimizar a perda de dados e a perda de regularização.

3. MÉTODOS DE PESQUISA

3.1 Definição do problema

A definição do problema é crucial para alcançar os objetivos propostos. Nesta etapa, foi definido o problema central do trabalho. O primeiro objetivo, encontrar modelos de previsão para uma empresa de distribuição de gasolina *premium*, foi alcançado. Os próximos objetivos consistem em prever as vendas de combustíveis *premium* para os próximos três meses e determinar quais previsões possuem a melhor acurácia. Utilizamos o RMSE (Root Mean Square Error) para comparar os métodos atuais (Regressão Linear e Médias Móveis) com novos

modelos, avaliando a melhoria proporcionada.

3.2 Obtenção de dados e configuração das ferramentas

Os dados utilizados são públicos e estão disponíveis no site da ANP, focando no Mercado Brasileiro de Combustíveis Líquidos. As principais variáveis são a data de cada transação e a quantidade de produtos (mil m³), filtrando apenas os produtos *premium*.

Serão utilizadas diferentes ferramentas e *softwares* conforme a complexidade dos modelos aplicados. O Excel será utilizado para modelos mais simples, como Regressão Linear e Médias Móveis. Para modelos mais avançados, como ARIMA e SVM, será utilizada a linguagem Python com bibliotecas especializadas como Pandas, Numpy, Matplotlib e Scikit-Learn, através da interface Jupyter Notebook.

3.3 Construção dos modelos

Com os dados tratados e as ferramentas configuradas, os modelos selecionados serão desenvolvidos. Utilizando o Excel, foi criado um modelo de Regressão Linear Simples. Primeiro, abriremos as séries temporais no Excel e definiremos as colunas (Data e Quantidade de Produtos). Em seguida, selecionou-se um gráfico de dispersão e adicionou uma linha de tendência. Depois, foi gerada uma equação linear e criada uma coluna de resultados. Por fim, foram estendidos os dados para a quantidade de meses que se quer prever.

Para o modelo de Médias Móveis, primeiro foram abertas as séries temporais no Excel e definidas as colunas (Data e Quantidade de Produtos). Em seguida, escolhidos os períodos para a média móvel (2, 3 e 6 meses). Depois, inserida a fórmula da média móvel. Após isso, replicada a fórmula e estendidos os dados para os períodos escolhidos. Depois de obter a previsão para os 3 períodos, foram aplicados os mesmos parâmetros de avaliação a fim de determinar o melhor período.

Para o modelo ARIMA, utilizou-se o Python. Primeiro, foi feita a instalação das ferramentas. Em seguida, passou-se para a preparação dos dados. A terceira etapa envolve a análise dos padrões e sazonalidade. Depois, verificou-se a estacionaridade dos dados. A quinta etapa consiste na divisão dos dados em treino e teste. Após isso, foi construído o modelo ARIMA com os parâmetros p, d e q. Na sétima etapa, feita a previsão dos dados de teste. Por fim, a avaliação do modelo.

Para o modelo SVM, foi utilizado o Python e bibliotecas adicionais. Primeiro, a instalação das ferramentas. Em seguida, a preparação dos dados. A terceira etapa envolve a análise dos padrões e sazonalidade. Depois, a divisão dos dados em treino e teste. A quinta etapa consiste na construção do modelo SVM. Após isso, a previsão dos dados de teste. Por

fim, a avaliação do modelo.

3.4 Avaliação e comparação dos modelos

A avaliação dos modelos de previsão é essencial para identificar as melhores opções. Utiliza-se a métrica de Erro Quadrático Médio da Raiz (RMSE) para a avaliação e comparação, conforme Shcherbakov et al. (2013). O RMSE penaliza mais severamente os grandes erros, proporcionando uma avaliação rigorosa da precisão do modelo. A fórmula do RMSE é:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_t^{(m)})^2} \quad (7)$$

Onde, y_i é o valor medido no tempo t e $f_t^{(m)}$ é a previsão no tempo t .

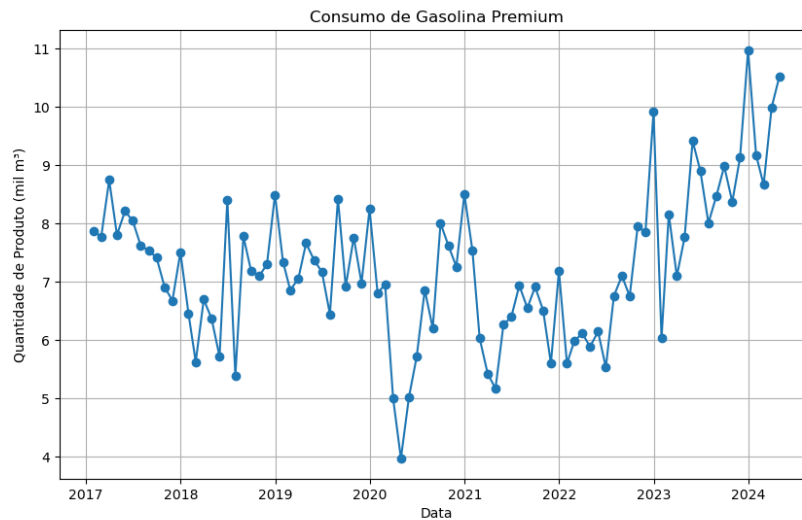
Utiliza-se o RMSE para comparar os modelos e determinar o mais eficaz para prever a demanda de combustíveis *premium* no Brasil.

4. RESULTADOS

4.1 Análise dos dados

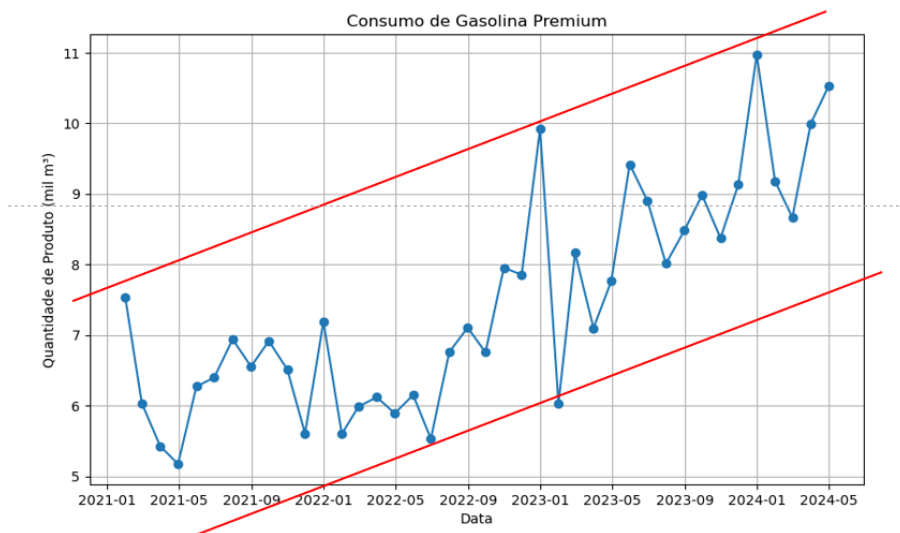
Os dados foram extraídos do site da ANP, conforme descrito na seção de obtenção e tratamento de dados. Após as transformações necessárias, iniciamos a análise de dados para identificar sazonalidades na demanda de combustíveis. A Figura 4 mostra uma sazonalidade significativa no período de 2020 a 2021, atribuída à pandemia de COVID-19, resultando em um *outlier*. Foram selecionados os dados de 2021 a 2024, onde uma sazonalidade mais consistente e uma tendência de crescimento podem ser observadas (Figura 5).

FIGURA 2 - Consumo de Gasolina Premium



Fonte: elaborada pelos autores, baseada em dados adaptados da ANP (2024).

Figura 3 - Tendência de Consumo de Gasolina Premium



Fonte: Elaborada pelos autores, baseada em dados adaptados da ANP (2024).

Como se pode observar, os dados apresentam uma sazonalidade evidente, como, por exemplo, os picos de consumo no final de cada ano, seguidos por uma queda brusca nos períodos subsequentes que marcam o início do novo ano. Isso demonstra uma relação clara entre o consumo e o período do ano.

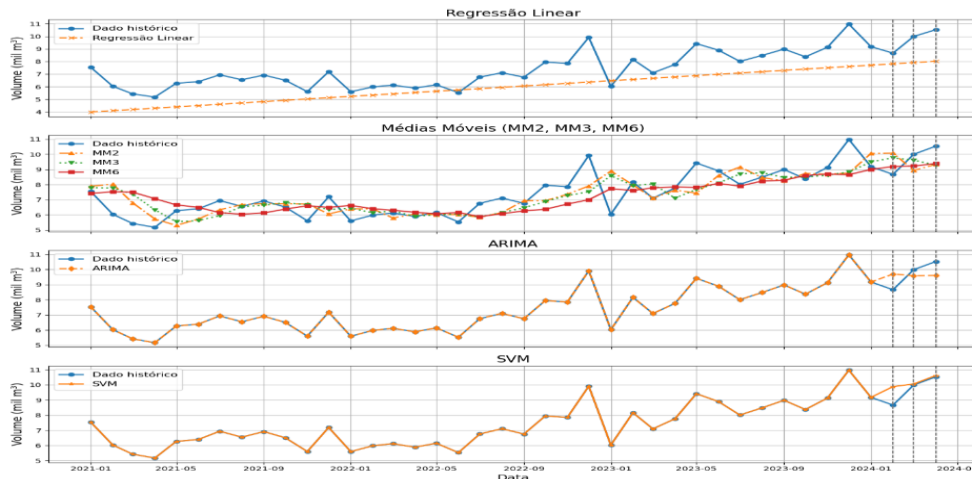
Além da perda de sazonalidade, outra importante razão para eliminar os períodos anteriores está relacionada à nossa métrica de avaliação, o RMSE, que não apresenta bom desempenho nesses casos. Para manter uma previsão mais sólida e uma métrica mais simples, o autor decidiu eliminar esses *outliers*, uma vez que a sazonalidade dos combustíveis foi identificada, assim como sua tendência, sempre se baseando no ano anterior.

O autor está ciente de que existem técnicas para correção desses dados e ajuste aos *outliers*. No entanto, esse não é o escopo deste trabalho, reservando essas técnicas para pesquisas futuras.

4.2 Aplicação dos Modelos

Os modelos selecionados foram aplicados aos dados para prever os três últimos períodos de teste e avaliar o desempenho de cada um. A Figura 6 apresenta os resultados de cada modelo.

Figura 4 - Previsões



Fonte: Elaborada pelos autores, baseada em dados adaptados da ANP (2024).

O RMSE foi utilizado para avaliar o desempenho de cada modelo (Tabela 2):

Tabela 1 - Resultados

Data	Regressão Linear	MM2	MM3	MM6	ARIMA	SVM
01/02/2024	7,82	10,08	9,76	9,19	9,70	9,89
01/03/2024	7,92	8,92	9,61	9,22	9,59	10,06
01/04/2024	8,02	9,33	9,28	9,39	9,62	10,62
RMSE	1,94	1,23	0,98	0,85	0,83	0,71

Fonte: Elaborado pelos autores (2024).

4.3 Análise dos Resultados

Os modelos apresentaram resultados distintos. A regressão linear teve a pior performance, subestimando variações sazonais e cíclicas. As médias móveis (MM2, MM3 e

MM6) suavizam flutuações de curto prazo, com MM6 destacando-se pela previsão mais estável. O modelo ARIMA capturou bem padrões de autocorrelação, fornecendo previsões precisas e um erro menor. O modelo SVM, utilizando técnicas de aprendizado de máquina, apresentou previsões robustas e equilibradas, sendo o modelo vencedor com o menor erro (Tabela 1).

4.4 Atendimento aos objetivos da Pesquisa

O primeiro objetivo foi encontrar modelos para previsão do mercado de gasolina *premium* no Brasil. Esse objetivo foi alcançado. Este trabalho comparou os modelos de regressão linear e médias móveis, além de buscar na literatura outros modelos amplamente utilizados, como ARIMA e SVM.

O segundo objetivo foi comparar os modelos encontrados e determinar o melhor para previsão. Esse objetivo foi alcançado. Após identificar os modelos, foram aplicadas técnicas e ferramentas para obter previsões. O RMSE foi utilizado para avaliar os modelos e determinar o mais preciso. O modelo SVM apresentou o menor erro, destacando-se como o mais eficaz.

5. CONSIDERAÇÕES FINAIS

O foco central deste trabalho foi propor um modelo de previsão para o mercado de gasolina *premium* no Brasil. O modelo SVM (*Support Vector Machine*) foi o mais eficaz. Usando dados de exemplo (ANP 2024), observou-se que 29,2 mil m³ foram consumidos no trimestre. A previsão com regressão linear foi de 23,76 mil m³ (18,6% abaixo do real), médias móveis de 6 períodos 27,8 mil m³ (4,8% abaixo do real), e SVM 30,57 mil m³ (4,6% acima do real), mostrando melhora significativa nas previsões. A precisão das previsões pode melhorar significativamente a performance das estratégias, resultando em maior velocidade de entrega, redução de custos e aumento na qualidade dos produtos.

Como limitação, reforça-se que apenas foram utilizados modelos univariados para fazer a previsão. No entanto, por serem produtos substitutos (gasolina *premium* e normal), de alguma forma trabalhar com uma previsão multivariada poderia trazer outras contribuições e *insights* para explicar o comportamento da demanda. Assim, a inclusão de variáveis adicionais como regiões de consumo, preço do combustível e PIB nacional e a utilização de modelos multivariados poderiam proporcionar uma análise mais abrangente e detalhada, capturando melhor as nuances do mercado e permitindo estratégias de precificação mais eficientes. Esses recursos devem ser incorporados não apenas para capturar padrões nos modelos utilizados, mas também para implementar modelos que foram abordados nesse trabalho, como, Random forest

e LSTM que possuem uma maior complexidade e necessidade de novos recursos na base de dados.

Além dessas melhorias, novas métricas de erro, técnicas de ajuste a *outliers* e a busca por mais valores históricos serão focos das futuras melhorias no modelo desenvolvido nesta pesquisa. O objetivo contínuo é otimizar as previsões e obter vantagens competitivas.

REFERÊNCIAS

GÉRON, A. (2019). **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn and TensorFlow**. Alta Books.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. 3. ed. Monash University, Australia, 2021

LORENZO-ESPEJO, A., ESCUDERO-SANTANA, A., MUÑOZ-DÍAZ, M.-L., & ROBLES-VELASCO, A. (2022). **Machine learning-based analysis of a wind turbine manufacturing operation: A case study**. *Sustainability*, 14(13), 7779.

GUIDO, S.; MUELLER, A. C. **Introduction to machine learning with python: A guide for data scientists**. Sebastopol, CA, USA: O'Reilly Media, 2016.

RAJ, R. **Supervised, Unsupervised and Semi-supervised learning**. Disponível em: <<https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning>>. Acesso em: 21 jul. 2023.

&, P.; RISHABH. **3 types of machine learning**. Disponível em: <<https://www.newtechdojo.com/3-types-of-machine-learning/>>. Acesso em: 21 jul. 2023.

ANP. **AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS**. Painel Dinâmico do Mercado Brasileiro de Combustíveis Líquidos. Disponível em: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/paineis-dinamicos-da-anp/paineis-dinamicos-do-abastecimento/painel-dinamico-do-mercado-brasileiro-de-combustiveis-liquidos>. Acesso em: 13 maio 2024.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: Forecasting and control**. 5. ed. [s.l.] John Wiley & Sons, 2016.

ASSAD, D. B. N.; CARA, J.; ORTEGA-MIER, M. Comparing short-term univariate and multivariate time-series forecasting models in infectious disease outbreak. **Bulletin of mathematical biology**, v. 85, n. 1, 2023.

BREIMAN, L. **Machine learning**, v. 45, n. 1, p. 5–32, 2001.

LI, Y. et al. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. **Knowledge-based systems**, v. 181, n. 104785, p. 104785, 2019.

SIAMI-NAMINI, S.; TAVAKOLI, N.; SIAMI NAMIN, A. A comparison of ARIMA and LSTM in forecasting time series. 2018 **17th IEEE International Conference on Machine Learning and Applications (ICMLA)**. Anais...IEEE, 2018.

DEY, B.; ROY, B.; DATTA, S.; USTUN, T. S. Forecasting ethanol demand in India to meet future blending targets: A comparison of ARIMA and various regression models. **Energy Reports**, 2023.

KUMAR, S.; SHARMA, R.; AGARWAL, P. Forecasting automobile gasoline demand in Australia using machine learning-based regression. **Energy Reports**, 2023.

LI, Y.; ZHANG, X.; ZHOU, P. Mixed random forest, cointegration, and forecasting gasoline prices. **Energy Economics**, 2023.

WANG, H.; LIU, Q.; TANG, J. An effective rolling decomposition-ensemble model for gasoline consumption forecasting. **Applied Energy**, 2023.

CHEN, Y.; WANG, R.; LIU, S. Forecasting retail fuel demand in Chinese gasoline stations: a structural double damped trend approach. **Energy Policy**, 2023.

SILVA, M. R.; OLIVEIRA, L. F.; SOUZA, A. P. Performance evaluation of forecasting models based on time series and machine learning techniques: an application to light fuel consumption in Brazil. **Renewable and Sustainable Energy Reviews**, 2023.

AN, W. et al. Predicting hourly heating load in residential buildings using a hybrid SSA–CNN–SVM approach. **Case studies in thermal engineering**, v. 59, n. 104516, p. 104516, 2024.

VAPNIK, V. N. **The nature of statistical learning theory**. Nova Iorque, NY, USA: Springer, 2014.

C8D, 6AF0 53F0 7D9D. **Download Python**. Disponível em: <<https://www.python.org/downloads/>>. Acesso em: 16 jun. 2024.

PETROBRAS. Gasolina. Disponível em: <https://petrobras.com.br/quem-somos/gasolina>. Acesso em: 20 jun. 2024.