

UNIVERSIDADE FEDERAL DE ALAGOAS – UFAL
UNIDADE ACADÊMICA CENTRO DE TECNOLOGIA CTEC
CURSO DE ENGENHARIA DE PETRÓLEO

**ANÁLISE E MODELAGEM DE SÉRIES TEMPORAIS PARA A PREVISÃO DE
PREÇOS DO BARRIL DE PETRÓLEO**

TRABALHO DE CONCLUSÃO DE CURSO

AUTOR: JÔNATHAS MAGALHÃES NUNES

Maceió, dezembro de 2024

UNIVERSIDADE FEDERAL DE ALAGOAS – UFAL
UNIDADE ACADÊMICA CENTRO DE TECNOLOGIA CTEC
CURSO DE ENGENHARIA DE PETRÓLEO

**ANÁLISE E MODELAGEM DE SÉRIES TEMPORAIS PARA A PREVISÃO DE
PREÇOS DO BARRIL DE PETRÓLEO**

Trabalho de Conclusão de Curso apresentado ao Colegiado do Curso de Engenharia de Petróleo da Universidade Federal de Alagoas, como parte integrante dos requisitos para obtenção do título de Bacharel em Engenharia de Petróleo.

Orientador: Lucas Pereira de Gouveia

Maceió, dezembro de 2024

Catlogação na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

N972a Nunes, Jônathas Magalhães.

Análise e modelagem de séries temporais para a previsão de preços do barril de petróleo / Jônathas Magalhães Nunes. – Maceió, 2024.
46 f. : il., grafs. e tabs. color.

Orientador: Lucas Pereira de Gouveia.

Monografia (Trabalho de conclusão de curso em Engenharia de Petróleo) – Universidade Federal de Alagoas. Centro de Tecnologia. Maceió, 2024.

Bibliografia: f. 44-46.

1. Análise de séries temporais. 2. Petróleo. 3. Média móvel integrada autorregressiva. 4. Auto-Regressivo integrado de médias móveis com sazonalidade.
I. Título.

CDU: 622.25:519.246

Dedico este trabalho com todo o meu coração a:

À minha mãe e meu pai, pelo amor incondicional, pela paciência e, principalmente, por nunca desistir de mim ou deste TCC, mesmo nos momentos mais difíceis. Sua força e fé em mim foram a base que me sustentou.

Aos meus professores, que foram muito mais do que educadores. Vocês acreditaram no meu potencial, mesmo quando eu duvidei. Obrigado por não desistirem de me orientar, por cada palavra de incentivo e por insistirem na minha capacidade de superar desafios.

À minha mulher, por ser minha parceira em todas as batalhas, pelo apoio, compreensão e pelo amor que ilumina meus dias. Sua presença ao meu lado tornou tudo possível.

Aos meus irmãos, que sempre estiveram comigo, seja para me apoiar ou para me lembrar de seguir em frente com coragem.

E aos meus amigos, que, com sua amizade e motivação, tornaram o caminho mais leve e significativo.

Cada um de vocês fez parte desta conquista. Este TCC é tão meu quanto é de vocês. Obrigado por acreditarem em mim.

RESUMO

Este estudo teve por objetivo realizar uma análise comparativa entre diversos métodos de previsão de séries temporais, com um foco específico nas flutuações históricas dos preços do petróleo. Nesse contexto, foram empregados modelos avançados como Média Móvel, Suavização Exponencial, ARIMA e SARIMA. Para a execução dessa análise, utilizamos a linguagem Python, juntamente com seus pacotes especializados em séries temporais. Estes pacotes ofereceram um conjunto de ferramentas sofisticadas destinadas à modelagem e à previsão de séries temporais, adaptando-se perfeitamente às complexidades associadas aos preços do petróleo.

A avaliação desses modelos foi pautada em critérios rigorosos de precisão e eficácia preditiva, incluindo a Raiz Quadrada do Erro Quadrático Médio de Previsão (RMSE) e o Critério de Informação de Akaike (AIC). Esses critérios permitiram uma comparação detalhada e criteriosa entre os diferentes métodos, garantindo a seleção do modelo mais adequado para a previsão das variações nos preços do petróleo no futuro.

Palavras-chave: Séries Temporais; Petróleo; ARIMA; SARIMA

LISTA DE ABREVIATURAS, SIGLAS E TRADUÇÕES

AIC - Critério de Informação de Akaike

Bootstrapping - Método de Reamostragem

Forecast - Previsão

Goals - Objetivos

STL - Seasonal and Trend decomposition using Loess

TBATS (Trigonometric, Box-Cox transformation, ARMA errors, Trend and Seasonal components)

Variáveis *dummy* - Variáveis criadas para representar uma variável com duas ou mais categorias

LISTA DE TABELAS

Tabela 1 - <i>Dataset</i> Preço Petróleo Brent	18
Tabela 2 - Critérios de Seleção do melhor modelo	32

LISTA DE FIGURAS

Figura 1 - Série Temporal com Padrão de Tendência	10
Figura 2 - Série Temporal com padrão Sazonal	11
Figura 3 - Série Temporal com Padrão Cíclico não regular	12
Figura 4 - Fluxograma das etapas.....	17
Figura 5 - Histograma do Preço do Petróleo Brent	20
Figura 6 - Série Temporal do Preço do Barril do Petróleo Brent.....	21
Figura 7 - Preço do Petróleo Brent - Diário, Semanal, Mensal e Anual	21
Figura 8 - Dados ausentes e outliers.....	22
Figura 9 - Produção Petróleo vs. Consumo do Petróleo ao longo do tempo.....	23
Figura 10 - Correlação Produção vs. Consumo do Petróleo	24
Figura 11 - Produção e Consumo Mundial do Petróleo e Preço do Petróleo	25
Figura 12 - Correlação Preço vs Consumo e Preço vs Produção	25
Figura 13 - Gráficos de Tendência e Padrão cíclico do preço do petróleo.....	26
Figura 14 - Previsão com Média Móvel	28
Figura 15 - Modelo de Suavização Exponencial para 0.79	29
Figura 16 - Modelo ES de Tareq	29
Figura 17 - Resultados de Tareq para o Modeo ARIMA (2, 1, 2)	30
Figura 18 - Modelo ARIMA.....	30
Figura 19 - Previsão com SARIMA	31

SUMÁRIO

1.	INTRODUÇÃO	1
1.1	COMENTÁRIOS INICIAIS.....	1
1.2	OBJETIVO.....	3
1.2.1	<i>Objetivo geral.....</i>	3
1.2.2	<i>Objetivos específicos</i>	3
1.3	JUSTIFICATIVA.....	3
1.4	DELIMITAÇÃO DO TRABALHO	4
2.	REFERENCIAL TEÓRICO	5
2.1	FORECAST	5
2.1.1	<i>O que pode ser previsto.....</i>	5
2.1.2	<i>Forecasting, planning e goals.....</i>	6
2.1.3	<i>Forecasting de dados</i>	7
2.1.4	<i>Passos básicos para uma forecasting.....</i>	8
2.2	PADRÕES (PATTERNS) DE SÉRIES TEMPORAIS	10
2.3	CORRELAÇÃO, CAUSA E PREVISÃO.....	12
2.4	DADO SEMANAL, DIÁRIO E SUB-DIÁRIO.....	13
2.5	SÉRIES MUITO LONGAS OU MUITO CURTAS.....	14
2.5.1	<i>Séries muito curtas.....</i>	14
2.5.2	<i>Séries muito longas</i>	14
2.6	VALORES AUSENTES E OUTLIERS	14
2.7	MÉTODOS	15
2.7.1	<i>Média móvel (MA).....</i>	15
2.7.2	<i>Suavização exponencial</i>	15
2.7.3	<i>ARIMA (Autoregressive Integrated Moving Average).....</i>	15
2.7.4	<i>SARIMA (Seasonal Autoregressive Integrated Moving Average).....</i>	16
3.	METODOLOGIA.....	17
3.1	PROCEDIMENTOS METODOLÓGICOS.....	17
3.1.1	<i>Coleta e preparação dos dados.....</i>	17
3.1.2	<i>Análise exploratória dos dados (AED)</i>	18
3.1.3	<i>Seleção do modelo de previsão</i>	19
3.1.4	<i>Implementação e monitoramento</i>	19
4.	RESULTADOS E DISCUSSÕES	20

4.1	ANÁLISE EXPLORATÓRIA DOS DADOS	20
4.2	ESCOLHA DO MODELO	27
4.2.1	<i>Médias Móveis</i>	27
4.2.2	<i>Suavização Exponencial</i>	28
4.2.3	<i>ARIMA</i>	29
4.2.4	<i>SARIMA</i>	31
5.	CONCLUSÕES.....	32

1. INTRODUÇÃO

1.1 COMENTÁRIOS INICIAIS

O petróleo mantém-se como a espinha dorsal energética, exercendo um papel preponderante na economia global e no Brasil, com projeções que apontam para a sua relevância contínua nas próximas décadas (YERGIN, 2011). Sua capacidade de adaptação tem sido crucial para a ampliação de suas aplicações e para sua expansão universal, atuando como um dos pilares do processo de globalização (SMIL, 2017). A análise de séries temporais surge como uma ferramenta chave na previsão dos preços do petróleo, conforme evidenciado por estudos que exploram diferentes metodologias para melhor entender e prever essas flutuações. Por exemplo, a pesquisa de Alquist, Kilian e Vigfusson (2013) emprega modelos de séries temporais para examinar a relação entre os preços do petróleo e a economia global, destacando a complexidade e a importância de previsões precisas.

O estudo do petróleo e sua relação com a economia global e brasileira é motivado pela necessidade de compreender e gerenciar a volatilidade dos preços dessa commodity, que afeta diretamente a estabilidade econômica e as estratégias energéticas. A relevância desse tema é amplificada pelo papel central do petróleo como fonte de energia primária e sua influência nos custos de transporte, produção industrial e desenvolvimento de tecnologias energéticas alternativas.

Um aspecto vital na dinâmica econômica do petróleo é a configuração de sua oferta, em que a prospecção de novos campos de petróleo representa uma empreitada que demanda significativos investimentos e enfrenta um alto nível de incerteza (KRUGMAN, 2008). A contribuição de Hamilton (2009) ao examinar os impactos dos choques de preços do petróleo na economia enfatiza a necessidade de métodos robustos de previsão, como séries temporais, para mitigar riscos econômicos.

O impacto dos preços do petróleo sobre a economia é inegável, influenciando diretamente os setores que dependem exclusivamente dessa fonte energética e, por outro lado, influenciando a viabilidade de alternativas energéticas potenciais. A importância da previsão de preços do petróleo através de séries temporais é enfatizada na literatura, com Wei et al. (2010) e Mohammadi e Su (2010) demonstrando a aplicabilidade e eficácia de modelos ARIMA e outros modelos de séries temporais na previsão de preços do petróleo, respectivamente. Esses estudos sublinham a relevância de técnicas avançadas de previsão na gestão de riscos e na formulação de políticas energéticas.

A motivação para trabalhos que tangenciam o tema reside, em grande parte, na busca por previsibilidade e controle em um mercado sujeito a múltiplas incertezas. Essas incertezas derivam de fatores geopolíticos, variações na demanda global, mudanças climáticas e o desenvolvimento de fontes de energia renováveis. A análise de séries temporais destaca-se nesse contexto como uma abordagem robusta para compreender os padrões históricos e prever flutuações futuras, proporcionando ferramentas essenciais para a tomada de decisão estratégica.

Além disso, a complexidade associada à prospecção e exploração de novos campos petrolíferos, como as reservas do pré-sal no Brasil, requer um planejamento econômico de longo prazo. Estudos que utilizam séries temporais para prever preços futuros auxiliam na avaliação de viabilidade econômica desses projetos, especialmente em cenários de instabilidade. Isso é particularmente importante para países cuja economia depende significativamente da exportação de petróleo, como o Brasil.

Especificamente para o Brasil, a estabilização dos preços do petróleo em patamares superiores a US\$100 por barril tem um impacto decisivo no desenvolvimento das reservas do pré-sal, conferindo viabilidade econômica à exploração e ao desenvolvimento dessas jazidas (BACIA ET AL., 2012). A análise de séries temporais oferece insights valiosos também neste contexto, facilitando a previsão de preços futuros e a tomada de decisões estratégicas para o aproveitamento desses recursos.

Quanto ao petróleo Brent, seu nome origina-se de um campo petrolífero operado pela Shell no Mar do Norte, denominado Brent. Atualmente, "Brent" refere-se ao petróleo proveniente dessa região, comercializado na Bolsa de Londres. Os contratos de Petróleo Brent são cotados em dólares americanos por barril, com cada contrato representando um barril de petróleo e um lote contendo 1000 barris, equivalendo a 158.988 litros por barril (OPEC, 2015). A análise de séries temporais na previsão do preço do petróleo Brent é também de grande importância, dada sua influência no mercado energético global.

O impacto de choques nos preços do petróleo sobre a economia também justifica o aprofundamento de estudos nessa área. Métodos avançados, como os modelos ARIMA e outras abordagens de séries temporais, fornecem subsídios não apenas para mitigar os riscos econômicos decorrentes de oscilações abruptas, mas também para formular políticas energéticas eficazes que promovam a diversificação e a segurança energética.

A análise do petróleo Brent, um benchmark global, reflete a importância de compreender como os preços praticados nos mercados internacionais impactam as economias locais. Prever essas variações é fundamental para antecipar impactos sobre o comércio exterior, câmbio e

inflação, justificando a contínua relevância de pesquisas focadas em previsões de séries temporais aplicadas ao setor energético.

1.2 OBJETIVO

O tema deste estudo concentra-se na análise comparativa dos métodos de previsão de séries temporais aplicadas à variação dos preços do petróleo. Com a volatilidade inerente aos mercados de energia, e especialmente ao mercado do petróleo, compreender e prever essas variações torna-se crucial para tomadores de decisão em diversos setores econômicos.

1.2.1 Objetivo geral

O objetivo geral do estudo é avaliar a eficácia de diferentes métodos de previsão de séries temporais incluindo Médias Móveis, Suavização Exponencial, ARIMA e SARIMA na precisão das previsões de preço do petróleo, utilizando o software Python e seus pacotes especializados.

1.2.2 Objetivos específicos

- Comparar a performance dos modelos de previsão de séries temporais em termos da Raiz Quadrada do Erro Quadrático Médio de Previsão (RMSE).
- Analisar a aplicabilidade dos critérios de seleção de modelos, como o Critério de Informação de Akaike (AIC) para a previsão de preços do petróleo.
- Identificar o modelo mais eficiente e preciso para a previsão de preços do petróleo.

1.3 JUSTIFICATIVA

A elaboração deste trabalho se justifica pela relevância estratégica do petróleo na economia global, marcando uma contribuição significativa tanto para a esfera econômica quanto para a literatura acadêmica. O petróleo, como um dos principais insumos energéticos mundiais, possui um impacto profundo sobre a economia global, afetando diretamente a política econômica, a segurança energética e o desenvolvimento sustentável.

A capacidade de prever as flutuações nos preços de commodities com precisão é de interesse crítico para uma ampla gama de investidores, empresas do setor energético, governos e organizações internacionais. Tais previsões podem auxiliar na formulação de políticas mais informadas e no desenvolvimento de estratégias eficazes para mitigar riscos associados à volatilidade dos preços, potencializando a tomada de decisão e o planejamento estratégico em diversos níveis.

Adicionalmente, ao explorar e comparar a eficácia de vários modelos estatísticos e computacionais, analisando o preço do CPO (óleo de Palma Bruto) e do petróleo, o estudo de KANCHYMALAY et al. 2017 empregou técnicas de aprendizado de máquina como percepção multicamada, regressão vetorial de suporte e suavização exponencial de Holt Winter para previsão de preços, onde a regressão vetorial de suporte com o algoritmo de SMO (*Sequential Minimal Optimization*) demonstrou resultados superiores em comparação com as outras técnicas. Já TULARAM, Gurudeo Anand & SAEED 2016, Tareq utilizaram os modelos de suavização exponencial, Holt-Winters e ARIMA na previsão de preços de petróleo, identificação do modelo ARIMA (2, 1, 2) como o mais eficaz para previsão de preços do petróleo, em particular, destaca a utilidade de abordagens sofisticadas em contextos caracterizados por alta incerteza e risco.

Este estudo, portanto, não apenas enriquece o conhecimento teórico e prático sobre a previsão de preços de commodities importantes, mas também oferece uma base robusta para futuras pesquisas e aplicações práticas em economia, finanças e planejamento estratégico. Assim, reitera a importância de avanços contínuos na pesquisa de métodos de previsão, especialmente em setores voláteis que têm impactos profundos na economia global e no bem-estar das sociedades.

1.4 DELIMITAÇÃO DO TRABALHO

Este trabalho delimita-se à análise de dados de preços de fechamento do petróleo Brent, considerando um período específico para a série temporal analisada. Os métodos de previsão avaliados serão restritos àqueles implementados na linguagem Python através de seus pacotes estatísticos, focando em modelos estatísticos e computacionais específicos. A discussão será centrada na comparação de modelos com base em critérios de erro e informação, sem abordar em profundidade as implicações econômicas, políticas ou ambientais das variações dos preços do petróleo.

2. REFERENCIAL TEÓRICO

2.1 FORECAST

O *forecast* estatístico, também conhecido como previsão estatística, é um método quantitativo que se baseia na utilização de dados históricos e análise estatística para fazer estimativas sobre o futuro. Este método assume que padrões observados no passado são indicativos de tendências futuras, o que permite a tomada de decisões informadas (HYNDMAN & ATHANASOPOULOS, 2021).

A previsão estatística envolve a coleta e análise de dados, aplicação de modelos matemáticos e estatísticos para compreender as relações entre variáveis, e uso desses modelos para projetar valores futuros. Dentre os métodos mais utilizados, estão a regressão linear, análise de séries temporais (incluindo ARIMA e modelos sazonais), suavização exponencial e técnicas baseadas em aprendizado de máquina.

Um dos textos fundamentais no campo da análise de séries temporais é o de Box, Jenkins e Reinsel (2015), que detalha a metodologia de Box-Jenkins, incluindo modelos ARIMA, essenciais para a previsão estatística em contextos econômicos e de negócios. Outro recurso valioso é o livro de Makridakis, Wheelwright e Hyndman (1998), que abrange uma ampla gama de métodos de previsão, desde técnicas básicas até avançadas, e oferece exemplos práticos de aplicação.

Estas referências são cruciais para entender tanto a teoria quanto a aplicação prática da previsão estatística, destacando sua utilidade em uma variedade de campos como economia, finanças e planejamento de demanda (HYNDMAN & ATHANASOPOULOS, 2021; BOX, JENKINS, & REINSEL, 2015; MAKRIDAKIS, WHEELWRIGHT, & HYNDMAN, 1998). A previsão estatística fornece ferramentas poderosas para antecipar desenvolvimentos futuros, facilitando a preparação e a tomada de decisões estratégicas.

2.1.1 O que pode ser previsto

A necessidade de previsões permeia diversas áreas, desde a planejamento de infraestrutura energética até a logística de operações cotidianas. Por exemplo, a decisão sobre a implementação de novas usinas de geração de energia baseia-se em previsões de demanda futura, refletindo a importância de métodos de previsão em grande escala (HYNDMAN & ATHANASOPOULOS, 2018). Da mesma forma, o agendamento de pessoal em call centers depende de previsões acuradas da demanda de chamadas, exemplificando a aplicabilidade da previsão em operações diárias (GARDNER, 2006). Adicionalmente, a gestão de inventário

requer a previsão de necessidades de estoque, ressaltando a relevância da previsão na logística e no suprimento (SILVER, PYKE & PETERSON, 1998).

O horizonte temporal das previsões pode variar significativamente, desde anos, no caso de investimentos de capital, até minutos, para ajustes em rotas de telecomunicações, demonstrando a flexibilidade e a necessidade da previsão como ferramenta de planejamento (BROCKWELL & DAVIS, 2002). Enquanto a previsão do nascer do sol de amanhã pode ser feita com precisão quase absoluta, a natureza imprevisível de eventos como os números da mega-sena destaca os limites da previsibilidade (MAKRIDAKIS, WHEELWRIGHT & HYNDMAN, 1998).

A capacidade de prever a demanda de eletricidade com alta precisão deve-se ao entendimento dos fatores que a influenciam, como as temperaturas e as condições econômicas, juntamente com a disponibilidade de dados históricos substanciais (BOX, JENKINS & REINSEL, 2015). Por outro lado, a previsão de taxas de câmbio é notoriamente difícil devido ao entendimento limitado dos fatores que as afetam e ao fenômeno das previsões se tornarem autorrealizáveis, ilustrando a hipótese do mercado eficiente (FAMA, 1970).

A previsão precisa encontra-se intrinsecamente ligada à capacidade de distinguir entre flutuações aleatórias e padrões genuínos nos dados históricos, uma habilidade essencial para desenvolver modelos de previsão eficazes (HYNDMAN & ATHANASOPOULOS, 2018). Assumir que os ambientes em mudança tornam a previsão impossível é um equívoco; na realidade, modelos de previsão eficazes são projetados para capturar essas mudanças (TALEB, 2007).

As situações de previsão variam amplamente, desde métodos simples baseados em dados mais recentes até modelos complexos como redes neurais e sistemas econométricos de equações simultâneas, demonstrando a diversidade das abordagens de previsão disponíveis (ZHANG, PATUWO & HU, 1998). A escolha do método de previsão apropriado depende dos dados disponíveis e da previsibilidade do fenômeno em questão (ARMSTRONG, 2001).

2.1.2 *Forecasting, planning e goals*

Forecasting é uma função estatística essencial nos negócios, fundamentando decisões relativas à produção, logística e gestão de pessoal, além de orientar o planejamento estratégico de longo prazo (MENTZER & MOON, 2005). Contudo, a prática de *forecasting* nos negócios frequentemente se confunde com planejamento e definição de objetivos, embora sejam conceitos distintos. *Forecasting* refere-se à previsão do futuro com a maior precisão possível, utilizando dados históricos e o conhecimento de eventos futuros potenciais (ARMSTRONG, 2001).

Goals, ou objetivos, representam as aspirações da organização. Idealmente, esses objetivos devem ser informados por previsões e planos concretos, embora, na prática, muitas vezes sejam estabelecidos sem um caminho claro para sua realização ou uma avaliação realista de sua viabilidade (KAPLAN & NORTON, 1996).

Planning, ou planejamento, atua como a ponte entre as previsões e os objetivos, determinando as ações necessárias para alinhar as projeções futuras com as metas desejadas. Este processo envolve uma série de decisões estratégicas sobre como alocar recursos de maneira eficaz para atingir os objetivos estabelecidos (PORTER, 1985).

O *Forecasting* deve ser integrado ao modelo de tomada de decisão da organização, apoiando a gestão em diversos níveis. As empresas modernas necessitam de previsões de curto, médio e longo prazo, adaptadas às especificidades de suas operações (WRIGHT & GOODWIN, 2009).

- Previsões de **curto** prazo são cruciais para operações diárias, como agendamento de pessoal e logística de produção (SNIJDERS, MATZAT, & REIPS, 2012).
- Previsões de **médio** prazo ajudam a determinar necessidades futuras de recursos, fundamentais para a aquisição de matérias-primas, contratação de pessoal e investimentos em equipamentos (CHASE, 1998).
- Previsões de **longo** prazo são utilizadas no planejamento estratégico, exigindo uma análise das tendências de mercado, fatores ambientais e capacidades internas (GRANT, 2003).

Desenvolver um sistema eficaz de *forecasting* requer um esforço especializado para identificar problemas de previsão, aplicar uma gama de métodos preditivos, selecionar as abordagens mais adequadas para cada situação e continuamente avaliar e refinar esses métodos (HYNDMAN & ATHANASOPOULOS, 2018). O suporte organizacional robusto é crucial para a aplicação bem-sucedida de métodos formais de previsão, assegurando que as previsões informem efetivamente o planejamento e a tomada de decisão (FILDES & GOODWIN, 2007).

2.1.3 *Forecasting* de dados

A escolha do modelo de previsão adequado depende fundamentalmente dos dados disponíveis, que podem ser categorizados em dois grupos principais:

- **Previsão Qualitativa:** Esta abordagem é adotada quando não existem dados numéricos disponíveis ou os dados existentes não são diretamente relevantes para as previsões em questão. Métodos qualitativos frequentemente se baseiam em julgamento de

especialistas ou opinião para gerar previsões (HYNDMAN & ATHANASOPOULOS, 2021).

- Previsão Quantitativa: Aplicável quando se dispõe de dados numéricos sobre o passado e é razoável supor que certos padrões ou tendências persistirão no futuro. Este grupo abrange uma ampla gama de técnicas estatísticas e de aprendizado de máquina que podem ser utilizadas para modelar e prever séries temporais (TAYLOR & LETHAM, 2021).

Cada método de previsão quantitativa possui suas características distintas, precisão e custos associados, que devem ser cuidadosamente considerados ao selecionar a técnica mais apropriada para um problema específico (SEEGER, SALINAS & FLUNKERT, 2021).

Para a previsão de dados futuros, especialmente no domínio de séries temporais, uma diversidade de técnicas pode ser aplicada. Abaixo temos alguns exemplos de séries temporais:

- O preço diário das ações da IBM
- A precipitação de chuvas mensal
- Vendas trimestrais da Amazon
- O lucro anual da Google
- O preço do Barril do Petróleo Brent

Cada uma dessas séries temporais é observada em sequência ao longo do tempo e pode ser analisada para identificar padrões, tendências e ciclos que informam as previsões futuras.

Recentemente, avanços significativos foram feitos na previsão de séries temporais, especialmente com o uso de métodos de aprendizado profundo, que demonstraram capacidade para capturar complexidades em dados de séries temporais de maneira mais eficaz do que os métodos tradicionais (ORESHKIN, CARPOV, CHAPADOS & BENGIO, 2020). Além disso, a aplicação de técnicas de aprendizado de máquina em larga escala tem permitido a modelagem mais precisa de séries temporais complexas, como as relacionadas ao clima, mercado de ações e padrões de consumo (LI ET AL., 2021).

2.1.4 Passos básicos para uma *forecasting*

2.1.4.1 Definição do problema

A definição clara do problema de previsão é crítica, pois orienta todo o processo subsequente. Compreender como as previsões serão utilizadas é fundamental para definir o escopo e os objetivos da modelagem (BOWERMAN ET AL., 2019). A comunicação com as

partes interessadas envolvidas na coleta de dados e na aplicação das previsões é essencial para garantir que o modelo atenda às necessidades da organização (HYNDMAN & ATHANASOPOULOS, 2021).

2.1.4.2 Coletando informação

A coleta de informações tanto quantitativas quanto qualitativas é crucial para o desenvolvimento de um modelo preditivo eficaz. A seleção de dados relevantes e a avaliação da sua suficiência e qualidade são passos críticos (HYNDMAN & ATHANASOPOULOS, 2021). Métodos de previsão baseados em julgamento podem ser úteis quando dados históricos são escassos ou não refletem mudanças estruturais recentes (MAKRIDAKIS ET AL., 2020).

2.1.4.3 Análise Preliminar (Exploratória)

A análise exploratória de dados é uma etapa inicial crucial para identificar padrões, tendências, ciclos sazonais e potenciais outliers nos dados. Esta fase envolve técnicas visuais e estatísticas para compreender as dinâmicas subjacentes aos dados (TUKEY, 1977; CLEVELAND, 1993).

2.1.4.4 Escolhendo e Ajustando Modelos

A seleção do modelo adequado depende de uma compreensão profunda dos dados e do problema em questão. Comparar vários modelos potenciais permite identificar a abordagem mais eficaz, considerando a precisão das previsões e a complexidade do modelo (BURNHAM & ANDERSON, 2002; JAMES ET AL., 2021). A ajuste do modelo envolve a estimativa de parâmetros que melhor se alinham aos dados históricos.

2.1.4.5 Utilizando e Avaliando o Modelo de Previsão

Após a seleção e ajuste do modelo, este é utilizado para realizar previsões. A validação do modelo é fundamental e geralmente é realizada através da comparação das previsões geradas com dados reais subsequentes. Métodos de avaliação da precisão de previsão, como o erro quadrático médio (MSE) e o erro absoluto médio (MAE), são comumente utilizados (HYNDMAN & KOEHLER, 2006). Além disso, enfrentar desafios organizacionais e práticos, como lidar com dados ausentes e outliers, é parte integrante da implementação eficaz de previsões na prática (FILDES ET AL., 2019).

2.2 PADRÕES (*PATTERNS*) DE SÉRIES TEMPORAIS

A análise de séries temporais frequentemente envolve a identificação de padrões específicos que podem ajudar na previsão de futuras observações. Estes padrões, como tendência, sazonalidade e ciclos, são fundamentais para compreender a dinâmica subjacente aos dados.

- **Tendência:** Uma tendência reflete um movimento de longo prazo nos dados, que pode ser de crescimento ou de declínio conforme Figura 1. Essas mudanças não necessariamente seguem um padrão linear e podem indicar uma "mudança de direção" significativa no comportamento da série temporal (CHATFIELD, 2016). A tendência pode ser impactada por uma variedade de fatores, incluindo mudanças econômicas, sociais ou tecnológicas.

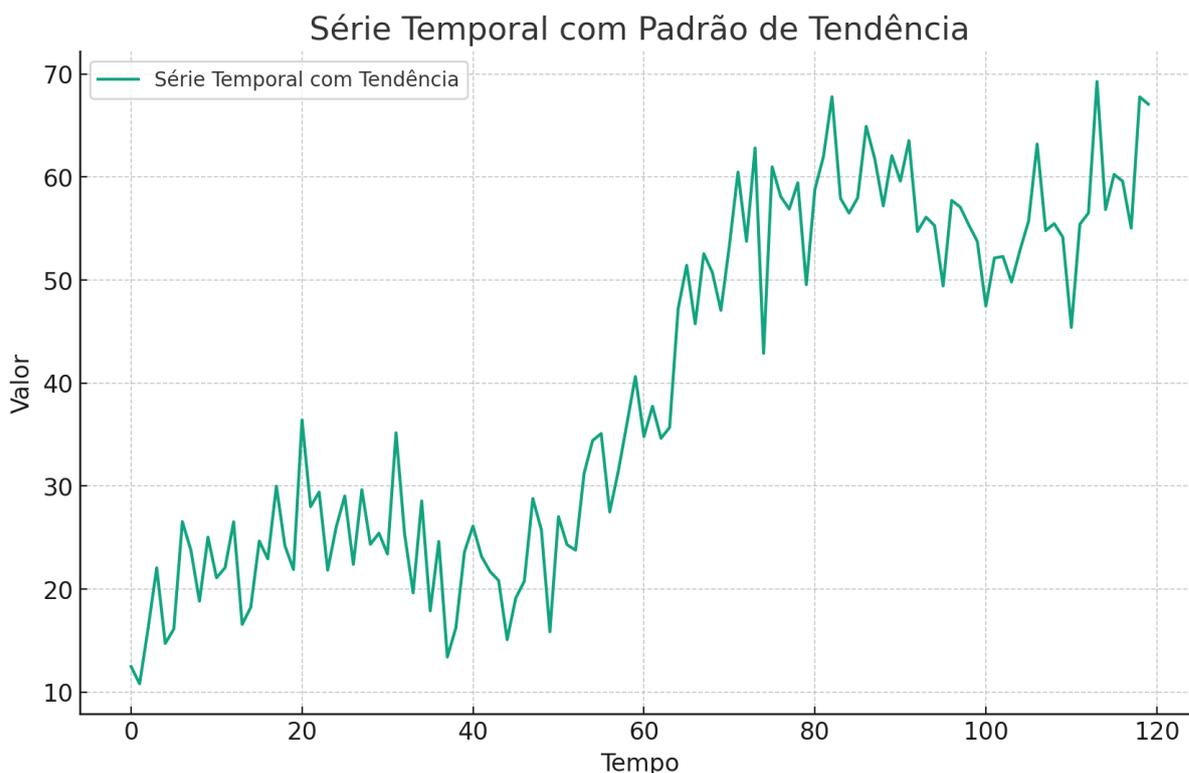


Figura 1 - Série Temporal com Padrão de Tendência

- **Sazonal:** Padrões sazonais ocorrem quando uma série temporal é influenciada por fatores sazonais repetitivos, como épocas do ano, dias da semana ou até mesmo horas do dia conforme Figura 2. Essa regularidade sazonal permite que modelos de previsão incorporem esses padrões para melhorar a acurácia das previsões (HYNDMAN & ATHANASOPOULOS, 2021).

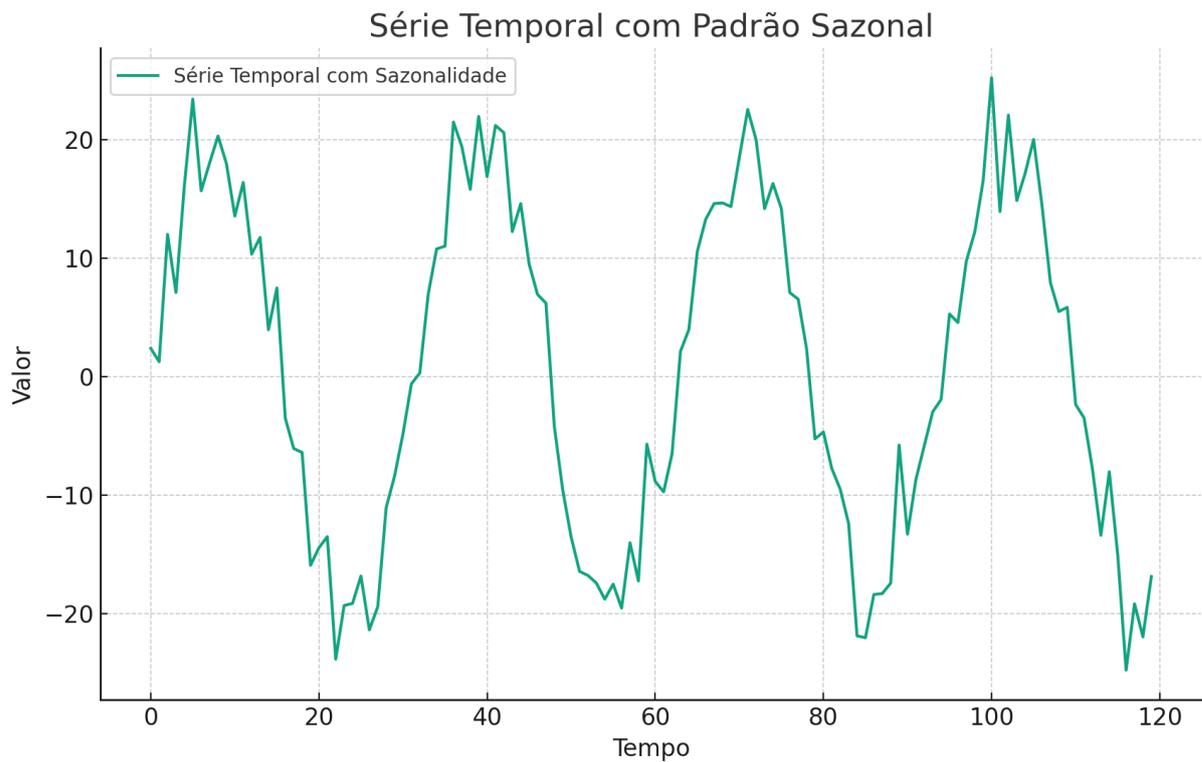


Figura 2 - Série Temporal com padrão Sazonal

- **Cíclico:** Diferente dos padrões sazonais, que têm uma frequência fixa, os ciclos são caracterizados por flutuações que não se encaixam em um período sazonal regular conforme Figura 3. Estes são frequentemente atribuídos a ciclos econômicos ou a fatores externos que afetam a série temporal em intervalos irregulares, podendo durar mais de dois anos (MILLS, 2019).

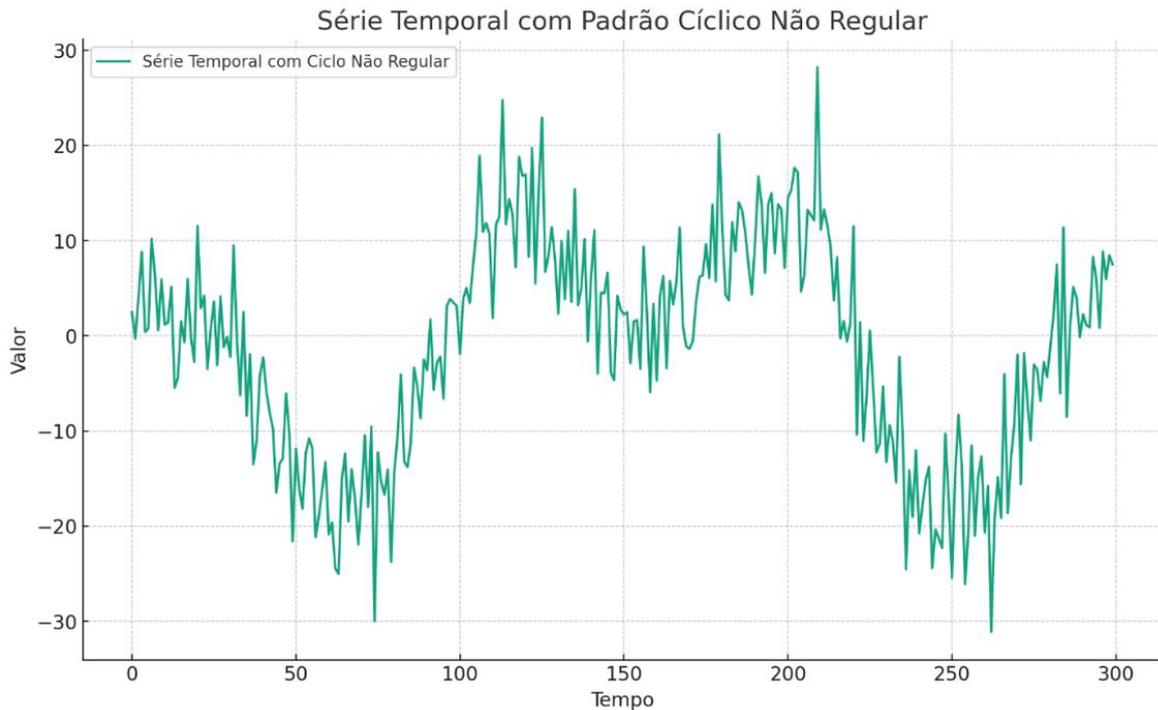


Figura 3 - Série Temporal com Padrão Cíclico não regular

A identificação e modelagem corretas desses padrões são essenciais para a construção de modelos de previsão eficazes. Utilizando técnicas avançadas de análise de séries temporais, como decomposição STL ou modelos ARIMA, é possível capturar essas dinâmicas e gerar previsões mais precisas (BROCKWELL & DAVIS, 2016).

2.3 CORRELAÇÃO, CAUSA E PREVISÃO

A compreensão da relação entre correlação e causalidade é fundamental na análise de dados e na formulação de previsões. A presença de uma correlação entre duas variáveis não implica necessariamente uma relação causal (PEARL, 2009). Por exemplo, enquanto a variável x pode estar correlacionada com a variável y , isso não indica automaticamente que x cause y . Inversamente, y pode ser a causa de x , ou a correlação pode ser resultante de ambas as variáveis sendo influenciadas por um terceiro fator não considerado (PEARL & MACKENZIE, 2018).

A questão dos preditores confusos surge quando não é possível distinguir os efeitos independentes de duas ou mais variáveis correlacionadas sobre a variável de interesse. Isso ocorre frequentemente em dados complexos, onde múltiplas variáveis interagem de maneiras não óbvias (MORGAN & WINSHIP, 2015). Embora a confusão entre preditores não impeça necessariamente a formulação de previsões, ela complica a interpretação das relações de causa

e efeito e pode levar a conclusões enganosas se não for adequadamente tratada (IMBENS & RUBIN, 2015).

Em cenários de previsão, é crucial considerar as interações entre variáveis para desenvolver modelos preditivos robustos e interpretáveis. A utilização de técnicas estatísticas avançadas, como modelos de regressão multivariada, métodos de variáveis instrumentais e análise de caminho, pode ajudar a desvendar as relações complexas entre variáveis e a identificar relações causais genuínas (ANGRIST & PISCHKE, 2009).

2.4 DADO SEMANAL, DIÁRIO E SUB-DIÁRIO

Trabalhar com dados semanais apresenta desafios únicos devido ao número não inteiro de semanas em um ano, tipicamente 52.18, o que complica a aplicação de métodos sazonais padrão que esperam períodos sazonais inteiros. Uma solução comum para esse problema é a aplicação da decomposição STL (*Seasonal and Trend decomposition using Loess*), que pode se ajustar flexivelmente a padrões sazonais de diferentes comprimentos, ou o uso de regressão harmônica dinâmica, que modela a sazonalidade através de senos e cossenos, permitindo maior flexibilidade na modelagem de períodos sazonais (CLEVELAND ET AL., 1990; TAYLOR ET AL., 2021).

Quando se trata de dados diários ou sub-diários, os desafios são amplificados pela presença de múltiplas sazonalidades, como padrões diários, semanais e anuais. Métodos como STL, regressão dinâmica harmônica e TBATS (*Trigonometric, Box-Cox transformation, ARMA errors, Trend and Seasonal components*) foram desenvolvidos especificamente para abordar essas complexidades, permitindo a modelagem de séries temporais com múltiplas sazonalidades (DE LIVERA ET AL., 2011; HYNDMAN & ATHANASOPOULOS, 2021).

Modelar a sazonalidade relacionada a eventos móveis, como feriados e celebrações que não ocorrem em datas fixas todos os anos, requer abordagens mais sofisticadas. Variáveis *dummy*, ou indicadoras, representam uma técnica comum para incorporar esses efeitos nos modelos, ajustando-se assim para variações atípicas nos dados que são atribuíveis a esses eventos (HYNDMAN & ATHANASOPOULOS, 2021). Esse método é particularmente útil em modelos de regressão dinâmica harmônica, onde os efeitos dos feriados podem ser explicitamente modelados junto com a sazonalidade regular.

2.5 SÉRIES MUITO LONGAS OU MUITO CURTAS

2.5.1 Séries muito curtas

Modelar séries temporais com poucos pontos de dados é um desafio notável devido à limitação na capacidade de identificar padrões estatisticamente significativos. A escolha do modelo deve ser guiada pela simplicidade para evitar o ajuste excessivo, com uma preferência por modelos com menos parâmetros a serem estimados (SHMUELI & LICHTENDAHL, 2017). O uso de técnicas de *bootstrapping* também pode oferecer uma maneira de avaliar a incerteza das previsões nessas situações (EFRON & TIBSHIRANI, 1994).

2.5.2 Séries muito longas

Para séries temporais extensas, o desafio muitas vezes reside na variabilidade dos dados ao longo do tempo e na possibilidade de mudanças estruturais. Métodos de janela móvel ou algoritmos de aprendizado de máquina que podem se adaptar ao longo do tempo, como o aprendizado profundo, podem ser particularmente úteis (GOODFELLOW, BENGIO, & COURVILLE, 2016). No entanto, a otimização de modelos complexos se torna computacionalmente intensiva com grandes conjuntos de dados.

2.6 VALORES AUSENTES E OUTLIERS

Valores ausentes são uma realidade comum e podem distorcer análises se não forem adequadamente tratados. Técnicas como imputação múltipla oferecem métodos para estimar valores faltantes com base na informação disponível, permitindo uma análise mais robusta dos dados (RUBIN, 2004). Modelos que incorporam variáveis *dummy* para indicar a presença de dados ausentes ou eventos específicos, como feriados, podem ajustar suas previsões para refletir essas peculiaridades (LITTLE & RUBIN, 2002).

A presença de *outliers* pode ter um impacto significativo na análise de séries temporais, podendo indicar tanto erros de medição quanto eventos atípicos genuínos. Métodos robustos de análise de séries temporais, que minimizam a influência de outliers, ou a aplicação de técnicas de detecção e tratamento de outliers, são essenciais para manter a integridade do modelo (ROUSSEEUW & LEROY, 2005).

2.7 MÉTODOS

Os métodos de Média Móvel, Suavização Exponencial, ARIMA e SARIMA são técnicas fundamentais na análise de séries temporais, cada um com suas particularidades e aplicações específicas.

2.7.1 Média móvel (MA)

O método de Média Móvel Equação 1 é utilizado para suavizar séries temporais e identificar tendências de longo prazo, calculando a média de um número específico de termos consecutivos ao longo da série. Este método é especialmente útil para eliminar flutuações de curto prazo e destacar tendências de longo prazo (CHATFIELD, 2016).

Modelagem:

$$MA_t = \frac{1}{n} \sum_{i=0}^{n-1} X_{t-i}$$

Equação 1 - Média Móvel

onde (MA_t) é o valor da média móvel no tempo (t), (X_{t-i}) são os valores da série temporal, e (n) é o número de períodos na média.

2.7.2 Suavização exponencial

A Suavização Exponencial Equação 2 é um método que aplica pesos decrescentes exponencialmente aos dados mais antigos. É particularmente eficaz para séries temporais sem tendência ou sazonalidade clara. A Suavização Exponencial Simples dá maior peso aos dados mais recentes, tornando-a sensível a mudanças recentes na tendência dos dados (HYNDMAN & ATHANASOPOULOS, 2021).

Modelagem:

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$$

Equação 2 - Suavização Exponencial

onde (S_t) é o valor suavizado no tempo (t), (X_t) é o valor observado no tempo (t), (S_{t-1}) é o valor suavizado anterior, e (α) é o parâmetro de suavização.

2.7.3 ARIMA (Autoregressive Integrated Moving Average)

O modelo ARIMA Equação 3 combina autorregressão (AR), diferenciação (I) para alcançar estacionariedade e média móvel (MA). É capaz de modelar uma ampla variedade de

séries temporais. A autorregressão modela a correlação entre uma observação e um número de observações defasadas. A diferenciação remove as tendências e sazonalidades, e a média móvel modela o erro como uma combinação linear de erros passados (BOX ET AL., 2015).

Modelagem:

$$\left(1 + \sum_{i=1}^p \phi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

Equação 3 – ARIMA

onde (L) é o operador de defasagem, (d) é o grau de diferenciação, (p) e (q) são as ordens dos componentes autoregressivos e de média móvel, respectivamente, (ϕ) e (θ) são os parâmetros desses componentes, e (ε_t) é o termo de erro.

2.7.4 SARIMA (Seasonal Autoregressive Integrated Moving Average)

O modelo SARIMA Equação 4 estende o ARIMA ao incluir componentes sazonais. É adequado para séries temporais com padrões sazonais claros, permitindo a modelagem simultânea da sazonalidade e da não sazonalidade na série (HYNDMAN & ATHANASOPOULOS, 2021).

Modelagem:

$$\left(1 + \sum_{i=1}^P \Phi_i L^{iS}\right) \left(1 + \sum_{i=1}^p \phi_i L^i\right) (1-L)^d (1-L^S)^D X_t = \left(1 + \sum_{i=1}^Q \Theta_i L^{iS}\right) \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

Equação 4 – SARIMA

onde (S) é o período sazonal, (P) , (D) , e (Q) são as ordens dos componentes sazonais autoregressivos, de diferenciação e de média móvel, respectivamente, e (Φ) e (Θ) são os parâmetros desses componentes sazonais.

3. METODOLOGIA

3.1 PROCEDIMENTOS METODOLÓGICOS

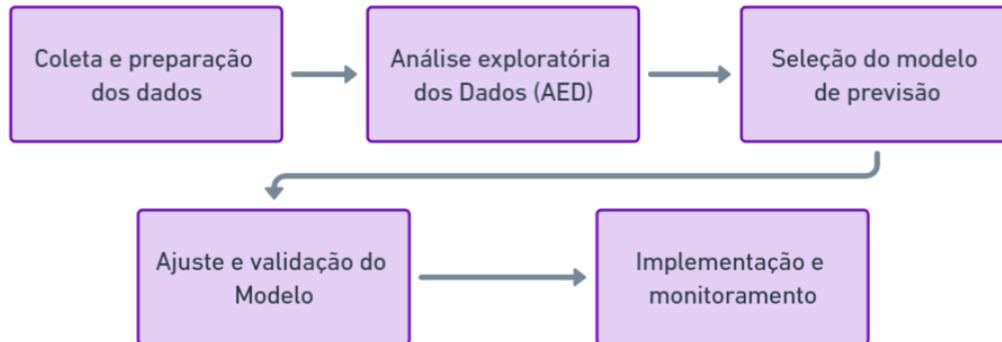


Figura 4 - Fluxograma das etapas

3.1.1 Coleta e preparação dos dados

Conforme o fluxograma de etapas na Figura 4, a coleta de dados envolve a reunião de informações relevantes que serão utilizadas para a análise. É importante garantir que os dados coletados sejam de alta qualidade, completos e pertinentes ao objetivo da previsão. A preparação dos dados pode incluir a limpeza (remoção de outliers e tratamento de valores faltantes) e a transformação dos dados para garantir sua adequação para análise.

Os dados escolhidos para análise correspondem aos valores de fechamento diário do barril de petróleo Brent, obtidos através da Bolsa de Valores, além de valores oficiais de consumo de produção mundiais do petróleo. Estes dados foram extraídos diretamente do site da Administração de Informação Energética dos Estados Unidos (*EIA - U.S. Energy Information Administration*), uma fonte oficial do governo dos Estados Unidos, no dia 19 de março de 2024. A base de dados disponibilizada pela EIA abrange um período extenso, iniciando em 20 de maio de 1987, e é atualizada regularmente todos os dias úteis após o encerramento das atividades da Bolsa de Valores.

Para o processamento e análise desses dados, foi utilizada a biblioteca *pandas* do Python, uma ferramenta poderosa para manipulação e análise de dados em Python. Foram realizadas transformações nas datas e valores para alcançar uma padronização consistente dos dados. Também foi observado que os dados estavam livres de ruídos, eliminando, portanto, a necessidade de procedimentos adicionais de limpeza. A Tabela 1 apresenta uma amostra dos dados tratados para oferecer uma visão preliminar do conjunto de dados trabalhado.

Tabela 1 - Dataset Preço Petróleo Brent

Index	Data	Preço
0	1987-05-20	18,63
1	1987-05-21	18,45
2	1987-05-22	18,55
3	1987-05-25	18,6
4	1987-05-26	18,63
...
9342	2024-03-12	83,66
9343	2024-03-13	83,95
9344	2024-03-14	86,16
9345	2024-03-15	85,39
9346	2024-03-18	86,47

3.1.2 Análise exploratória dos dados (AED)

Nesta etapa, foram adotadas técnicas sofisticadas de visualização de dados para realizar uma Análise Exploratória de Dados (AED) com o objetivo de identificar o modelo preditivo mais eficaz para a série temporal do petróleo Brent. Para isso, foram utilizadas as bibliotecas Python *sklearn*, *statsmodels*, *matplotlib* e *scipy*, que são amplamente reconhecidas por sua eficácia em análises estatísticas e gráficas.

Posteriormente foi iniciada a análise decompondo a série temporal em duas componentes principais: tendência, sazonalidade e ciclos. Esse processo é crucial para entender o comportamento intrínseco dos dados ao longo do tempo e identificar padrões que podem não ser imediatamente aparentes. Utilizamos o método de decomposição clássica proporcionado pela biblioteca *statsmodels*, que separa a série em seus componentes de forma visualmente interpretável.

Além da decomposição, aplicamos técnicas de plotagem para gerar histogramas e diversos outros gráficos. Os histogramas nos permitiram visualizar a distribuição de frequências dos dados, auxiliando na identificação de outliers e na compreensão da distribuição normal dos componentes da série temporal. Os gráficos adicionais, como gráficos de dispersão e de linha, ajudaram a correlacionar variáveis e a observar a evolução dos dados ao longo do tempo.

Esta abordagem metodológica integrada e baseada em visualizações detalhadas permitiu selecionar de forma informada o modelo que oferece o melhor equilíbrio entre precisão e praticidade para a previsão de preços do petróleo Brent.

3.1.3 Seleção do modelo de previsão

3.1.3.1 Ajuste e validação do modelo

Após a seleção do modelo, este foi ajustado aos dados de teste, incluindo a estimação dos parâmetros do modelo e a sua calibração para maximizar a precisão das previsões.

3.1.4 Implementação e monitoramento

Com o modelo ajustado e validado, o modelo foi implementado para fazer previsões futuras, monitorando continuamente o desempenho do modelo ao longo do tempo, ajustando-o conforme necessário para garantir que as previsões permaneçam precisas e relevantes.

4. RESULTADOS E DISCUSSÕES

Nesta seção, serão discutidos detalhadamente os resultados alcançados por meio da metodologia aplicada, focando na análise de séries temporais do petróleo Brent. Ao longo deste estudo, empregamos técnicas estatísticas e de aprendizado de máquina para modelar e prever comportamentos futuros dessa série temporal, crucial para a compreensão das dinâmicas do mercado de petróleo.

4.1 ANÁLISE EXPLORATÓRIA DOS DADOS

Para melhor compreensão dos dados foi realizada a análise exploratória nas amostras utilizadas nesse trabalho, o resultado pode ser observado a seguir.

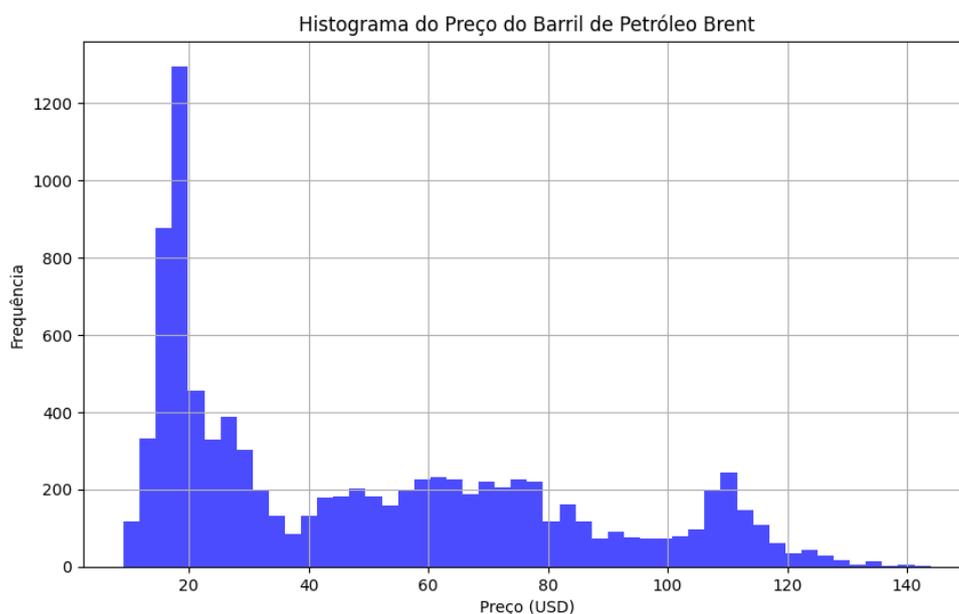


Figura 5 - Histograma do Preço do Petróleo Brent

O histograma na Figura 5 - Histograma do Preço do Petróleo Brent representa o preço do barril de petróleo Brent e mostra uma concentração de valores na faixa mais baixa, indicando longos períodos de estabilidade nos preços em torno de 20 a 30 dólares, com modas claras que sugerem frequências de estabilização de preços. Observam-se também picos de preços mais altos, provavelmente outliers relacionados a momentos de alta volatilidade devido a crises geopolíticas ou desastres naturais.

A análise estatística dos preços do petróleo Brent revela uma média de USD 49.65 e uma mediana de USD 42.19, indicando uma distribuição com cauda à direita, onde a média é puxada para cima por valores mais altos. Os quartis são distribuídos como segue: o primeiro quartil (25%) está em USD 19.25, o segundo quartil (50% - mediana) em USD 42.19, e o terceiro

quartil (75%) em USD 73.46. O desvio padrão, que é de USD 32.91, aponta para uma variação considerável nos preços, refletindo a volatilidade característica do mercado de petróleo ao longo do tempo.

Ademais, também foram utilizados os dados históricos de consumo e produção mundiais de petróleo para analisar a influência dessas duas variáveis no preço do petróleo Brent.

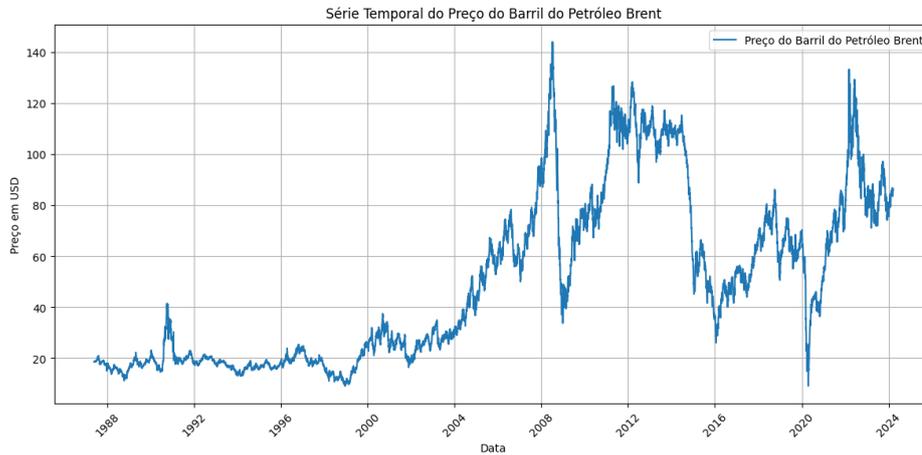


Figura 6 - Série Temporal do Preço do Barril do Petróleo Brent

A Figura 6 - Série Temporal do Preço do Barril do Petróleo Brent apresenta a variação dos preços diários de fechamento do barril do petróleo ao longo do tempo, desde sua inclusão na bolsa de valores até o período da coleta.

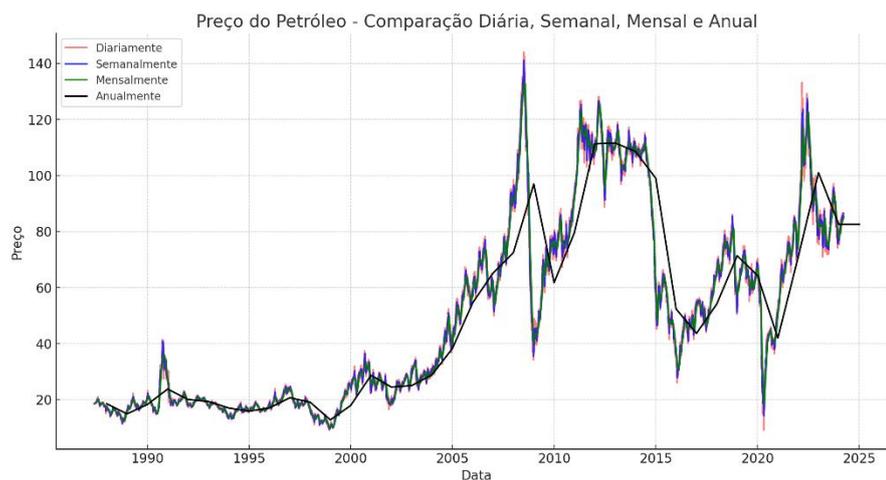


Figura 7 - Preço do Petróleo Brent - Diário, Semanal, Mensal e Anual

Adicionalmente a Figura 7 - Preço do Petróleo Brent - Diário, Semanal, Mensal e Anual ilustra uma análise temporal detalhada, apresentando uma sobreposição dos preços de fechamento diários do petróleo Brent com suas respectivas médias calculadas em diferentes

períodos: semanais, mensais e anuais. Este comparativo permite observar a volatilidade inerente às transações diárias em contraste com as tendências mais suavizadas que emergem quando os preços são agregados em intervalos de tempo mais longos. Tal visualização destaca as flutuações de curto prazo frente à estabilidade relativa e às mudanças de tendência observadas em uma escala temporal ampliada, fornecendo um panorama abrangente da dinâmica do mercado de petróleo ao longo do período analisado.

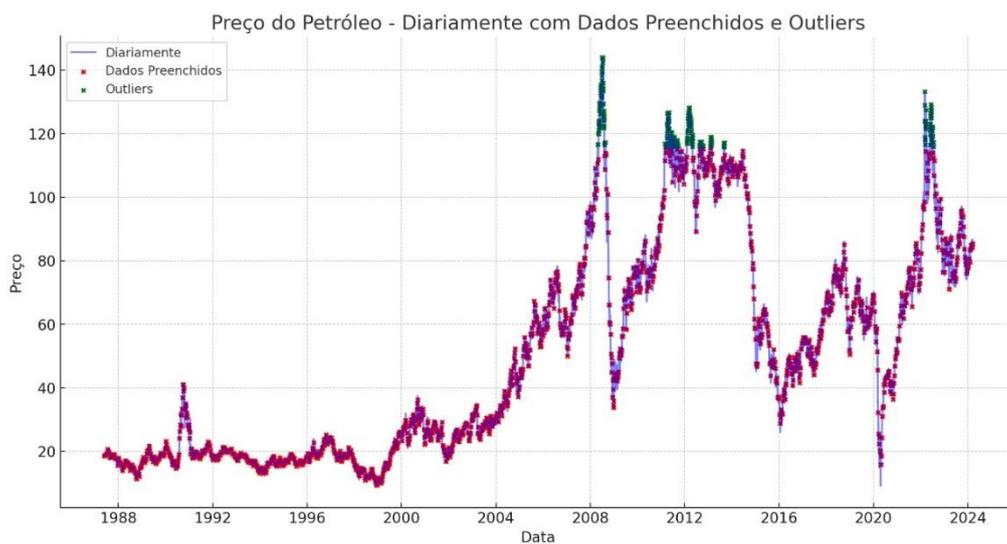


Figura 8 - Dados ausentes e outliers

A Figura 8 - Dados ausentes e outliers oferece um gráfico detalhado que destaca as anomalias e lacunas nos dados históricos dos preços do petróleo Brent. Os *outliers*, marcados em verde, representam valores que se desviam significativamente do padrão geral, potencialmente indicando eventos atípicos ou distorções no mercado. Complementarmente, os dados ausentes, identificados em vermelho, correspondem aos dias em que não houve atividade comercial devido a finais de semana ou feriados, quando as bolsas de valores estão fechadas. Esta representação gráfica não apenas aponta para as irregularidades e a não uniformidade do registro de dados, mas também sublinha a importância de técnicas de preenchimento e correção para uma análise mais consistente e robusta das tendências de mercado.

Na análise da série temporal do preço do petróleo Brent na bolsa de valores, foram observados desafios específicos associados à natureza intermitente dos dados. As lacunas nos dados são decorrentes de dias não operacionais, como fins de semana e feriados, que são comuns em dados financeiros.

Em relação aos dados ausentes, o preenchimento dessas lacunas não foi realizado. Esta decisão foi baseada na observação de que a ausência de transações em dias não operacionais não altera significativamente as previsões futuras, ou seja, o modelo de previsão escolhido mostrou-se robusto o suficiente para lidar com essas interrupções sem prejuízo significativo à precisão do *forecast*.

Quanto aos *outliers*, foi detectado que eles correspondem a picos significativos nos preços das ações, que estão diretamente relacionados a eventos específicos que impactaram o mercado de petróleo durante o período analisado. Tais eventos podem incluir crises geopolíticas, alterações regulatórias, ou mudanças significativas na oferta e demanda. Notou-se que tais picos não são apenas anomalias; eles são indicativos de possíveis eventos futuros semelhantes que podem afetar novamente o mercado. Portanto, esses outliers nos dados foram mantidos, já que incorporam informações valiosas sobre a volatilidade do mercado e tendências que são cruciais para modelar a dinâmica de preços das ações de forma realista.

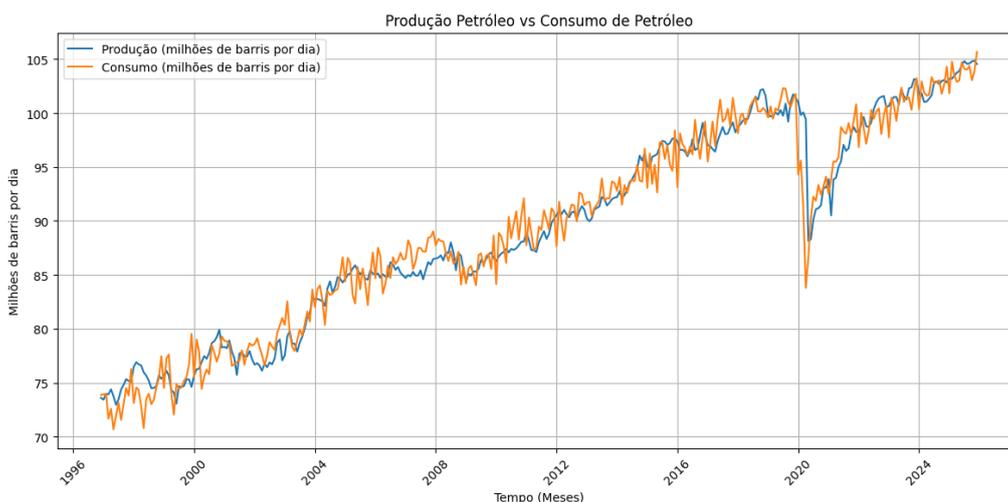


Figura 9 - Produção Petróleo vs. Consumo do Petróleo ao longo do tempo

A Figura 9 - Produção Petróleo vs. Consumo do Petróleo ao longo do tempo apresenta os dados de Consumo do Petróleo e Produção do Petróleo ao longo do tempo, é possível observar que aparentemente existe uma correlação visual entre as duas variáveis. Para determinar com exatidão tal correlação, foi aplicada a correlação de Spearman onde o resultado pode ser observado na Figura 10 - Correlação Produção vs. Consumo do Petróleo.

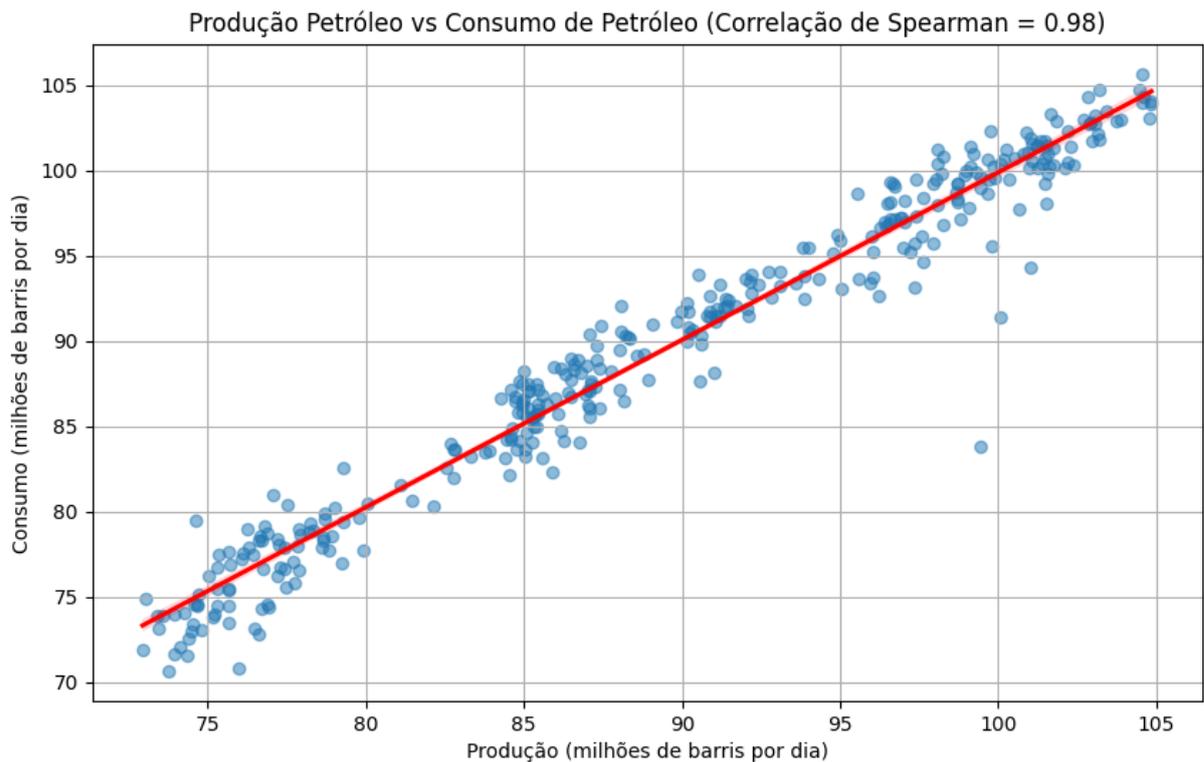


Figura 10 - Correlação Produção vs. Consumo do Petróleo

O gráfico de dispersão com a linha de regressão ilustra a relação entre a produção e o consumo mundial de petróleo. A correlação de Spearman entre essas duas variáveis é muito alta, com um valor de aproximadamente 0.98. Isso indica uma forte associação positiva, onde aumentos na produção de petróleo geralmente correspondem a aumentos no consumo.

Após essa análise inicial, investigou-se a existência de correlações entre o preço do petróleo Brent e as variáveis de Consumo e Produção Mundial de Petróleo. A Figura 11 - Produção e Consumo Mundial do Petróleo e Preço do Petróleo ilustra o gráfico que compara o Preço do Petróleo Brent com o Consumo e a Produção Mundial. Em certos intervalos, é possível notar um comportamento semelhante entre as variáveis. Para explorar mais detalhadamente essas semelhanças, novamente aplicou-se a correlação de Spearman entre as variáveis mencionadas.

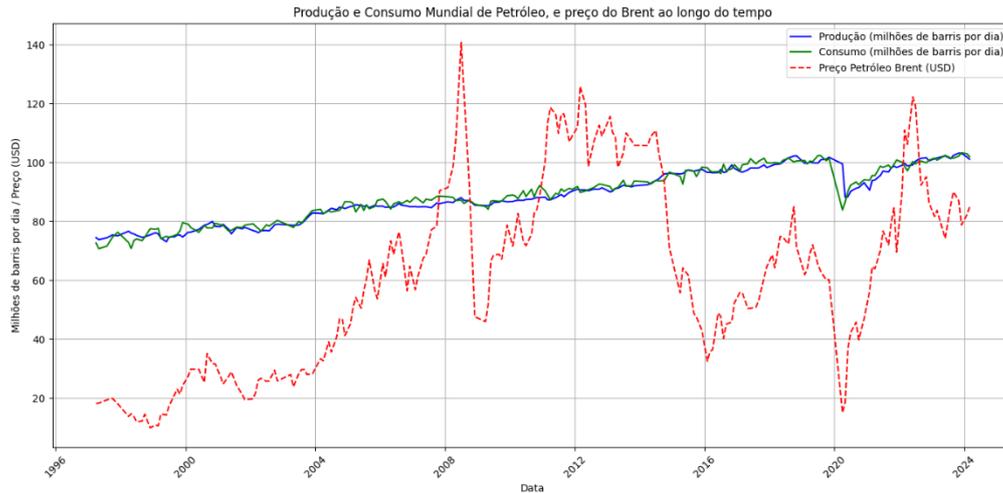


Figura 11 - Produção e Consumo Mundial do Petróleo e Preço do Petróleo

A Figura 12 - Correlação Preço vs Consumo e Preço vs Produção representa os gráficos de dispersão e estes mostram as relações entre o preço do petróleo Brent e as variáveis de produção e consumo de petróleo. A correlação de Spearman entre o preço do Brent e a produção de petróleo é de aproximadamente 0.62, indicando uma associação positiva moderada. Da mesma forma, como esperado devido à forte associação positiva entre o consumo e produção, a correlação entre o preço do Brent e o consumo de petróleo também é de cerca de 0.62.

Esses resultados sugerem que tanto a produção quanto o consumo tendem a aumentar quando o preço do Brent é mais alto, embora a relação não seja extremamente forte. Isso pode refletir fatores de mercado, onde preços mais altos incentivam uma maior produção e, ao mesmo tempo, podem estar associados a períodos de maior demanda

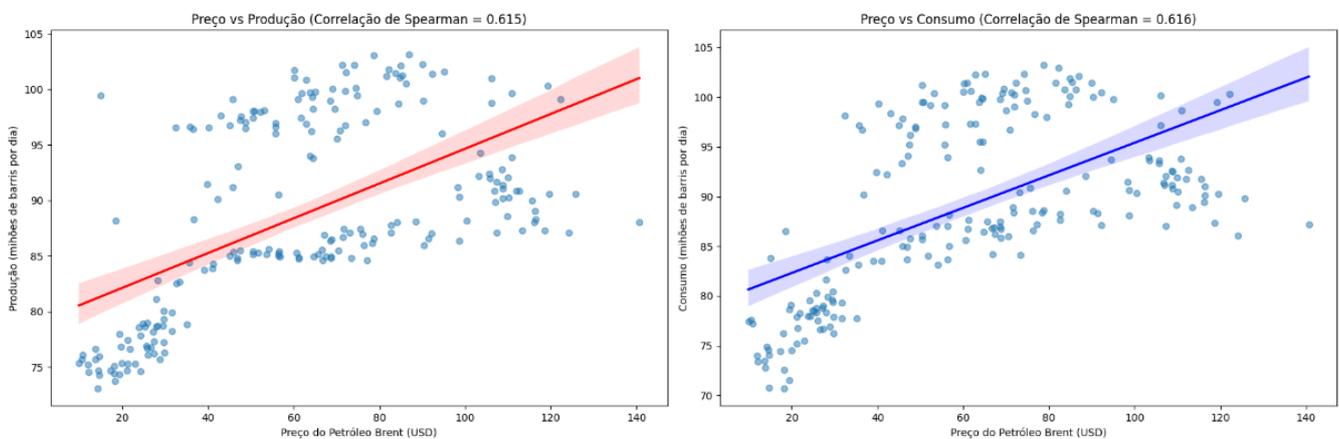


Figura 12 - Correlação Preço vs Consumo e Preço vs Produção

Na Figura 13 - Gráficos de Tendência e Padrão cíclico do preço do petróleo o gráfico de tendência do preço do barril de petróleo Brent revela um aumento consistente até 2008, seguido por um declínio acentuado durante a crise financeira global. Após esse período, há uma recuperação significativa que culmina em um pico por volta de 2012. No entanto, a partir de 2014, os preços entram em uma nova fase de declínio, com recuperações moderadas subsequentes, sem retornar aos níveis anteriores. Após 2020, observa-se uma leve recuperação, provavelmente associada aos efeitos da pandemia e à retomada da economia global.

O padrão cíclico, por sua vez, evidencia as oscilações em torno da tendência, refletindo eventos de curto prazo que impactaram o mercado. As maiores flutuações estão associadas a eventos como a crise financeira de 2008 e o choque de demanda causado pela pandemia em 2020. Em períodos de maior estabilidade, como no final da década de 1990 e início dos anos 2000, as variações são menos intensas. Esses ciclos destacam a sensibilidade dos preços do petróleo a fatores econômicos, geopolíticos e ambientais.

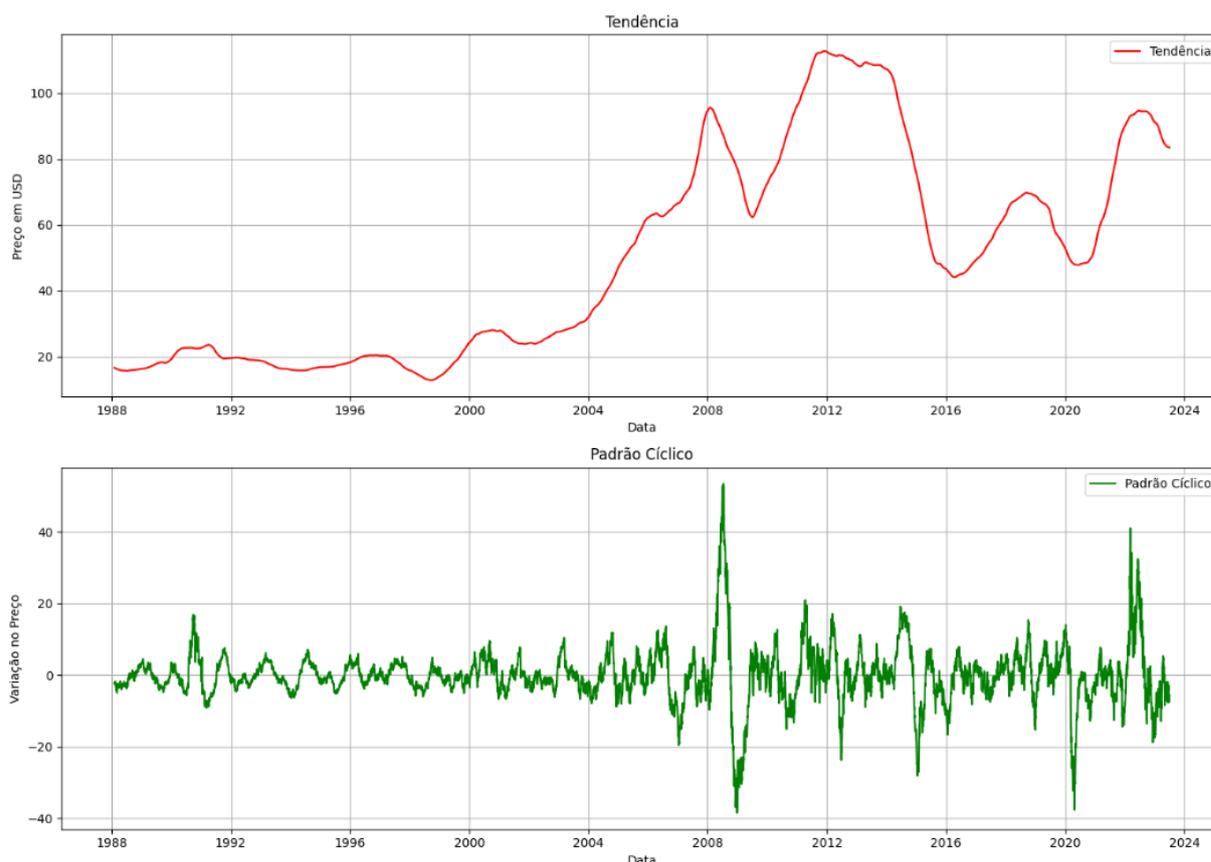


Figura 13 - Gráficos de Tendência e Padrão cíclico do preço do petróleo

Para fins de análise comparativa, este estudo examina os resultados obtidos por

TULARAM, Gurudeo Anand & SAEED 2016, Tareq, que investigaram os preços do barril de petróleo WTI (West Texas Intermediate) em uma janela temporal que compreende o período de outubro de 2015 a março de 2016. Os modelos em comum com os autores são: **Suavização Exponencial e ARIMA**.

4.2 ESCOLHA DO MODELO

A escolha de um modelo para prever o preço do petróleo Brent exige uma análise criteriosa dos padrões históricos da série temporal. O petróleo é uma commodity cujo preço é altamente influenciado por fatores externos, incluindo sazonalidade (por exemplo, padrões anuais de oferta e demanda), ciclos econômicos e choques externos, como conflitos geopolíticos. Modelos como Médias Móveis, Suavização Exponencial e ARIMA são amplamente utilizados em séries temporais, mas a SARIMA (*Seasonal ARIMA*) é frequentemente mais adequada para séries como esta devido à sua capacidade de modelar sazonalidades explícitas.

4.2.1 Médias Móveis

O modelo de Médias Móveis é uma abordagem simples que utiliza a média de observações passadas para suavizar flutuações e capturar tendências recentes. Apesar de ser útil para dados estacionários e para capturar oscilações curtas, este método apresenta limitações significativas:

- Não modela sazonalidades ou ciclos de longo prazo.
- Não é robusto para capturar padrões complexos como choques econômicos ou sazonalidade.
- Podemos observar na Figura 14 - Previsão com Média Móvel que a média móvel replica as médias dos últimos registros, apresentando uma tendência a se tornar uma linha reta ao longo de períodos mais extensos. Embora a técnica ilustre uma acurácia razoável no cálculo das médias móveis, observa-se que os valores reais acompanham predominantemente o limite inferior do intervalo de confiança na primeira metade do período analisado, enquanto na segunda metade tendem a se aproximar do limite superior. Apesar dessa proximidade com os limites do intervalo, a previsão com média móvel, ao ser comparada com outras abordagens, não oferece uma predição suficientemente satisfatória para capturar as variações reais dos dados.

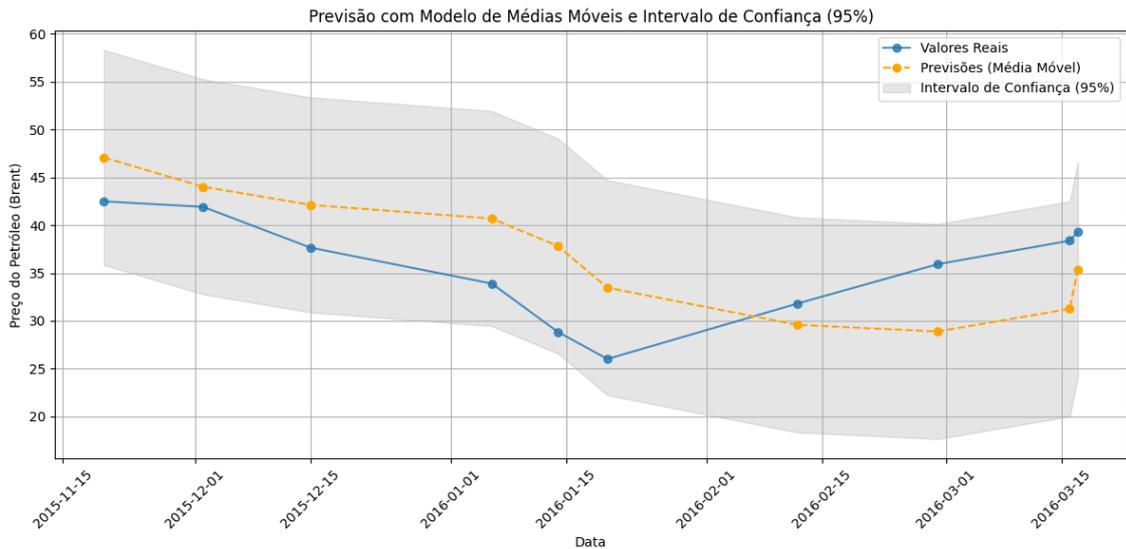


Figura 14 - Previsão com Média Móvel

4.2.2 Suavização Exponencial

A Suavização Exponencial melhora em relação às Médias Móveis ao ponderar mais fortemente as observações recentes, capturando tendências e, em alguns casos, sazonalidade.

No entanto:

- Este modelo assume que as tendências e padrões são relativamente estáveis ao longo do tempo.
- Ele não modela explicitamente ciclos ou sazonalidade complexa, que são comuns em preços de commodities como o petróleo.
- A projeção do preço do petróleo para os próximos três meses na Figura 15, feita a partir da série histórica. Com um parâmetro de suavização de 0.79, a previsão dá maior peso às observações mais recentes, refletindo possíveis mudanças no mercado. Quando comparado ao modelo de Médias Móveis, a figura ilustra um melhor encaixe. Observando os intervalos de confiança, pode-se observar que o comportamento do preço real acompanha as tendências tanto dos intervalos, quanto do valor previsto. Além disso, apresenta um resultado semelhante ao de Tareq Figura 16, mesmo considerando *datasets* distintos, ao utilizar os mesmos parâmetros no modelo.
- A suavização não considera explicitamente a sazonalidade ou eventos externos que podem impactar os preços do petróleo, como mudanças geopolíticas, desastres naturais ou decisões da OPEP.

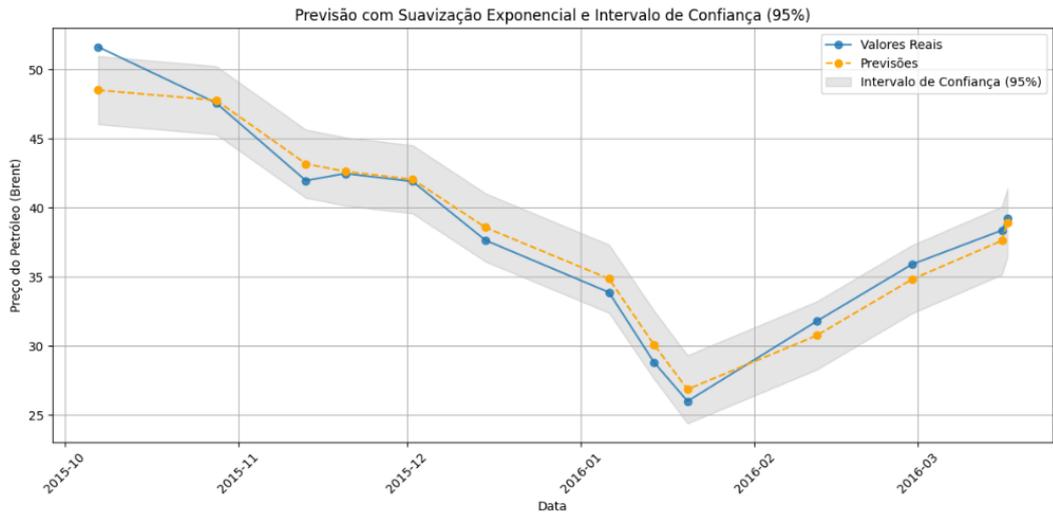


Figura 15 - Modelo de Suavização Exponencial para 0.79

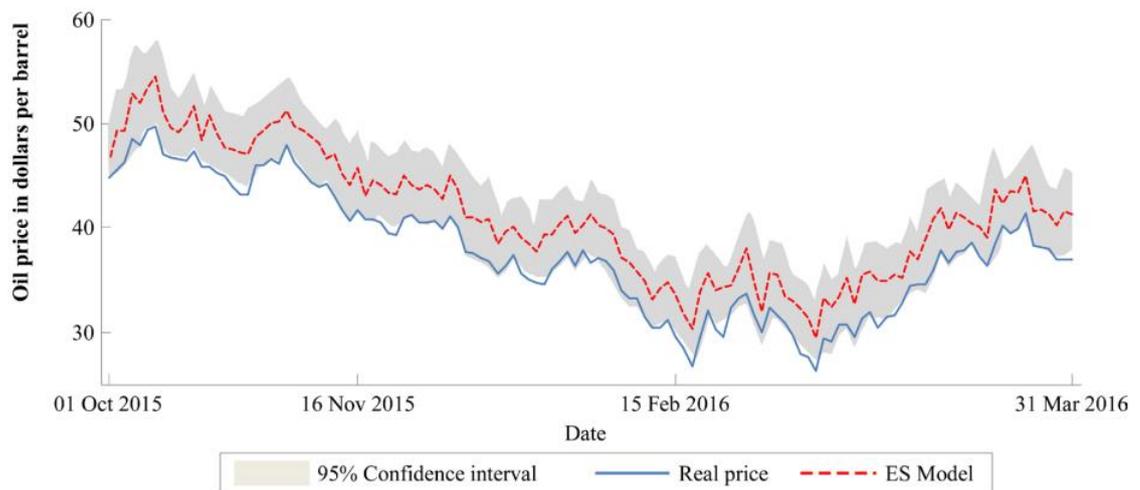


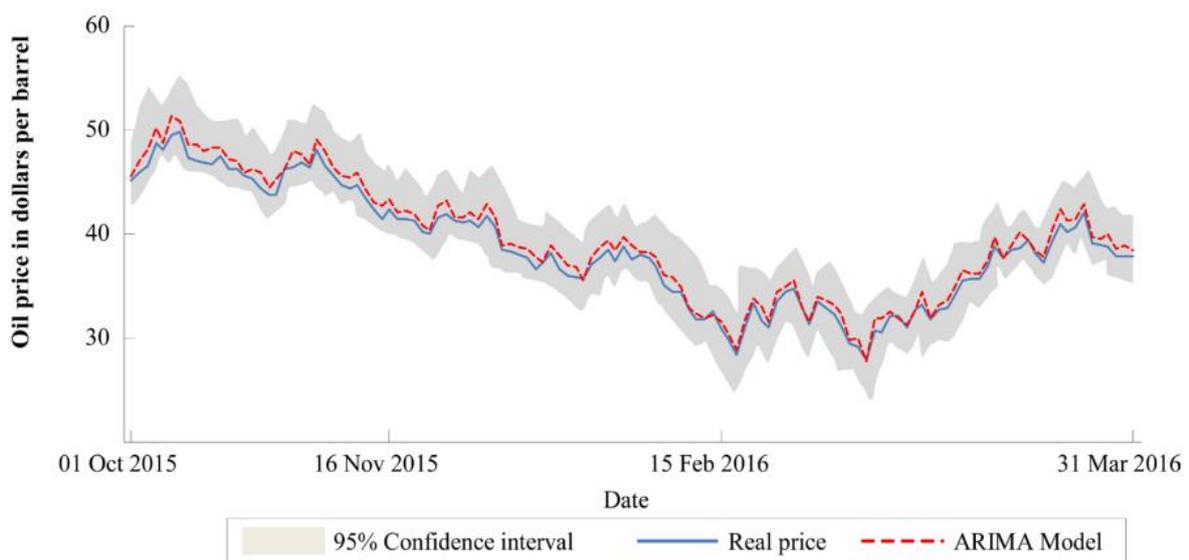
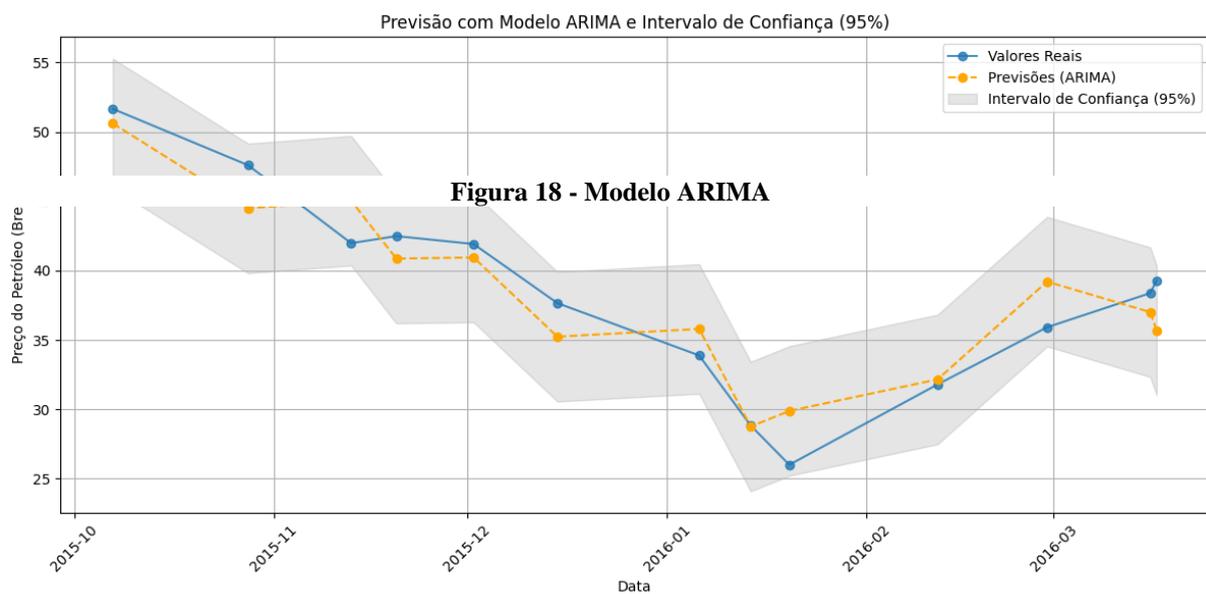
Figura 16 - Modelo ES de Tareq

4.2.3 ARIMA

O modelo ARIMA é uma abordagem robusta que considera dependências temporais entre os valores observados, capturando tanto padrões autorregressivos quanto médias móveis. Apesar disso, o ARIMA possui limitações importantes:

- Ele é projetado para dados não sazonais e não incorpora explicitamente a sazonalidade, o que é essencial para séries como o preço do petróleo.
- Ao forçar a série a se tornar estacionária, o modelo pode ignorar padrões cíclicos de longo prazo que não se ajustam bem a uma única tendência.

- Na Figura 18 - Modelo ARIMA, é possível observar que as previsões geradas pelo modelo estão relativamente próximas dos valores reais, mantendo-se dentro dos intervalos de confiança e acompanhando as tendências dos dados. O modelo adotado utilizou os parâmetros ARIMA (2, 1, 2), resultando em um desempenho considerado satisfatório. Além disso, ao comparar os resultados obtidos com os apresentados por Tareq Figura 17 - Resultados de Tareq para o Modeo ARIMA (2, 1, 2), nota-se um comportamento semelhante, reforçando a consistência do modelo em reproduzir as características identificadas na literatura.



4.2.4 SARIMA

O modelo SARIMA (*Seasonal ARIMA*) estende o ARIMA para capturar sazonalidades de forma explícita. Ele considera dependências sazonais além das relações temporais e tendências. Isso o torna ideal para o preço do petróleo Brent, que frequentemente exhibe padrões sazonais anuais devido a flutuações de demanda (como no inverno e verão) e à produção de grandes exportadores:

- **Modelagem sazonal explícita:** SARIMA incorpora parâmetros sazonais (P,D,QP, D, QP,D,Q) que permitem capturar ciclos sazonais e padrões repetitivos.
- **Flexibilidade:** Ele pode modelar tendências, ciclos e sazonalidades, tornando-o mais robusto para séries temporais complexas.
- **Melhor ajuste aos dados:** Para séries como a do petróleo Brent, o SARIMA frequentemente apresenta melhores métricas de desempenho, como AIC e RMSE, quando comparado a modelos menos anteriores em longas janelas de tempo Figura 19.

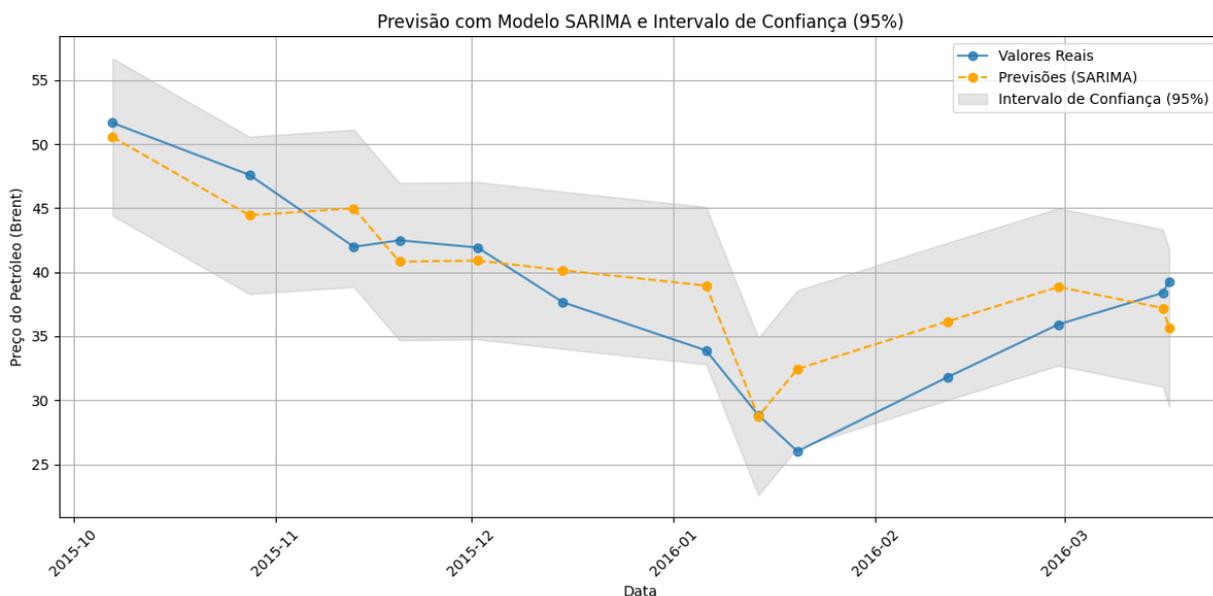


Figura 19 - Previsão com SARIMA

Tabela 2 - Critérios de Seleção do melhor modelo

<i>Critérios</i>	<i>Médias Móveis</i>	<i>Suavização Exponencial</i>	<i>ARIMA</i>	<i>SARIMA</i>
<i>RMSE</i>	5.9	1.19	1.05	1.1
<i>AIC</i>	37.52	6.57	3.27	4.48
<i>Ranking do Modelo</i>	4°	3°	1°	2°

A Tabela 2 representa o comparativo dos 4 modelos e a performance das previsões destes. Os resultados demonstram que a performance do modelo ARIMA é melhor do que os outros modelos para a série temporal do preço petróleo Brent. A abordagem de previsão no modelo de Média Móvel baseia-se na média simples ou ponderada de um número fixo de observações passadas, sem dar prioridade maior a valores mais recentes, o que pode limitar sua capacidade de captar tendências em dados dinâmicos. Já no modelo ES, a previsão é feita com base em uma média ponderada dos valores observados no passado, em que os pesos diminuem exponencialmente. Isso significa que as observações mais recentes contribuem de forma significativamente maior para a previsão em comparação com os dados mais antigos.

O modelo SARIMA (ARIMA Sazonal) expande o ARIMA tradicional ao incluir componentes sazonais, permitindo captar padrões cíclicos nos dados, além das componentes de autorregressão, integração e média móvel. O modelo ARIMA, por sua vez, baseia-se em três partes principais: autorregressão, integração e média móvel, em que o valor futuro de uma variável é calculado como uma combinação linear dos valores passados e dos erros associados. Essas diferenças metodológicas explicam os distintos comportamentos de previsão entre os modelos.

5. CONCLUSÕES

Com base na análise comparativa dos diferentes modelos aplicados para a predição do preço do barril de petróleo Brent, o modelo ARIMA demonstrou ser a abordagem mais eficaz. A escolha do ARIMA em vez do SARIMA pode ser justificada em situações em que o padrão sazonal não é evidente ou significativo nos dados analisados. O ARIMA é uma abordagem mais simples e direta, projetada para capturar as relações lineares entre os valores passados

(autorregressão), as diferenças entre observações consecutivas (integração) e os erros residuais (média móvel). Isso o torna ideal para séries temporais que não apresentam variações sazonais ou onde a sazonalidade é mínima e pode ser tratada adequadamente com o modelo básico. Além disso, métricas como AIC, e RSME apresentaram valores consistentemente mais baixos para o modelo ARIMA, indicando um melhor equilíbrio entre ajuste e simplicidade.

Os outros modelos avaliados, incluindo Médias Móveis, Suavização Exponencial e SARIMA, apresentaram limitações significativas. Enquanto as Médias Móveis e a Suavização Exponencial são úteis para capturar tendências curtas e suavizar flutuações, ambos carecem da capacidade de modelar sazonalidades explícitas, limitando sua aplicabilidade para séries temporais com padrões cíclicos complexos como os preços do petróleo. O modelo SARIMA, embora o mais robusto para dados sazonais, na janela adotada, não há sazonalidade, conseqüentemente, resultou visões menos alinhadas com as características da série histórica do petróleo Brent quando comparado ao ARIMA.

O ARIMA possui menor complexidade estrutural em comparação ao SARIMA, o que facilita tanto a parametrização quanto a implementação. Essa simplicidade é vantajosa, especialmente quando os dados disponíveis são limitados ou quando não há múltiplos ciclos sazonais para análise. Em situações em que a série temporal é curta ou incompleta, o SARIMA pode não capturar adequadamente as características sazonais, enquanto o ARIMA fornece uma alternativa confiável e eficiente.

Por fim, em aplicações práticas voltadas para projeções de curto prazo ou onde a sazonalidade não exerce influência determinante, o ARIMA frequentemente atende aos objetivos de maneira eficaz. Ele entrega resultados satisfatórios com menor esforço computacional, evitando a complexidade adicional do SARIMA. Assim, a escolha do ARIMA é estratégica, alinhada às características dos dados, ao objetivo da previsão e à relação custo-benefício de adotar um modelo mais simples e direto.

REFERENCIAS

- ALQUIST, R.; KILIAN, L.; VIGFUSSON, R.J. **Forecasting the price of oil**. Handbook of Economic Forecasting, v. 2, p. 427-507, 2013.
- HAMILTON, J.D. **Understanding crude oil prices**. The Energy Journal, v. 30, n. 2, p. 179-206, 2009.
- KRUGMAN, P. **The oil nonbubble**. The New York Times, 12 maio 2008.
- MOHAMMADI, H.; SU, L. **International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models**. Energy Economics, v. 32, n. 5, p. 1001-1008, Set. 2010.
- SMIL, V. **Energy and civilization: A history**. Cambridge, Massachusetts: The MIT
- ANGRIST, J.D.; PISCHKE, J.-S. **Mostly Harmless Econometrics: An Empiricist's Companion**. Princeton: Princeton University Press, 2009.
- ARMSTRONG, J.S. **Principles of forecasting: A handbook for researchers and practitioners**. Kluwer Academic Publishers, 2001.
- BACIA et al. **The Economics of Oil Exploration in Deepwater**. Journal of Petroleum Science and Engineering, 2012.
- BOWERMAN, B.; O'CONNELL, R.; KOEHLER, A.B.; HYNDMAN, R.J. **Forecasting: Principles and Practice**. OTexts, 2019.
- BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. **Time Series Analysis: Forecasting and Control**. 5. ed. Wiley, 2015.
- BROCKWELL, P.J.; DAVIS, R.A. **Introduction to Time Series and Forecasting**. Springer, 2002.
- BROCKWELL, P.J.; DAVIS, R.A. **Introduction to Time Series and Forecasting**. 3. ed. Springer, 2016.
- BURNHAM, K.P.; ANDERSON, D.R. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. Springer, 2002.
- CHASE, C.W. **Demand-driven forecasting: A structured approach to forecasting**. Wiley, 1998.
- CHATFIELD, C. **The Analysis of Time Series: An Introduction**. 7. ed. CRC Press, 2016.
- CLEVELAND, R.B.; CLEVELAND, W.S.; McRAE, J.E.; TERPENNING, I. **STL: A Seasonal-Trend Decomposition Procedure Based on Loess**. Statistics Research Report. AT&T Bell Laboratories, 1990.
- CLEVELAND, W.S. **Visualizing Data**. Hobart Press, 1993.
- DE LIVERA, A.M.; HYNDMAN, R.J.; SNYDER, R.D. **Forecasting time series with complex seasonal patterns using exponential smoothing**. Journal of the American Statistical Association, v. 106, n. 496, p. 1513-1527, 2011.
- EFRON, B.; TIBSHIRANI, R.J. **An Introduction to the Bootstrap**. CRC Press, 1994.

- GARDNER, E.S. Jr. **Exponential smoothing: The state of the art – Part II.** International Journal of Forecasting, v. 22, n. 4, p. 637-666, 2006.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning.** MIT Press, 2016.
- GRANT, R.M. **Contemporary Strategy Analysis.** 5. ed. Blackwell Publishing, 2003.
- HAMILTON, J.D. **Causes and Consequences of the Oil Shock of 2007–08.** Brookings Institution, 2009.
- HYNDMAN, R.J.; ATHANASOPOULOS, G. **Forecasting: principles and practice.** OTexts, 2018.
- HYNDMAN, R.J.; ATHANASOPOULOS, G. **Forecasting: principles and practice.** 3. ed. OTexts, 2021.
- HYNDMAN, R.J.; KOEHLER, A.B. **Another look at measures of forecast accuracy.** International Journal of Forecasting, v. 22, n. 4, 2006.
- IMBENS, G.W.; RUBIN, D.B. **Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.** Cambridge University Press, 2015.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications** in R. Springer, 2021.
- KANCHYMALAY, Kasturi et al. **Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques.** In: International Research and Innovation Summit (IRIS2017). IOP Conference Series: Materials Science and Engineering, 2017.
- KAPLAN, R.S.; NORTON, D.P. **The balanced scorecard: Translating strategy into action.** Harvard Business School Press, 1996.
- KRUGMAN, P. **The Return of Depression Economics and the Crisis of 2008.** W.W. Norton & Company, 2008.
- LI, X.; ZHAO, L.; WEI, Y.; YANG, M.H.; WU, F.; ZHUANG, Y.; LING, H.; WANG, J. **On the versatility and ease of use of deep learning models in earth system science.** Nature Communications, 2021.
- LITTLE, R.J.; RUBIN, D.B. **Statistical Analysis with Missing Data.** 2. ed. John Wiley & Sons, 2002.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.C.; HYNDMAN, R.J. **Forecasting: methods and applications.** John Wiley & Sons, 1998.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.C.; HYNDMAN, R.J. **Forecasting Methods and Applications.** 2. Ed. John Wiley & Sons, 2020.
- MENTZER, J.T.; MOON, M.A. **Sales forecasting management: A demand management approach.** Sage Publications, 2005.
- MILLS, T.C. **Time Series Techniques for Economists.** Cambridge University Press, 2019.
- MORGAN, S.L.; WINSHIP, C. **Counterfactuals and Causal Inference: Methods and Principles for Social.** Research. Cambridge University Press, 2015.
- OPEC. **Annual Statistical Bulletin.** Organization of the Petroleum Exporting Countries, 2015.
- ORESHKIN, B.N.; CARPOV, D.; CHAPADOS, N.; BENGIO, Y. **N-BEATS: Neural basis expansion analysis for interpretable time series forecasting.** International Conference on Learning Representations, 2020.

- PEARL, J. **Causality: Models, Reasoning, and Inference**. Cambridge University Press, 2009.
- PEARL, J.; MACKENZIE, D. **The Book of Why: The New Science of Cause and Effect**. Basic Books, 2018.
- PORTER, M.E. **Competitive advantage: Creating and sustaining superior performance**. Free Press, 1985.
- ROUSSEEUW, P.J.; LEROY, A.M. **Robust Regression and Outlier Detection**. John Wiley & Sons, 2005.
- RUBIN, D.B. **Multiple Imputation for Nonresponse in Surveys**. John Wiley & Sons, 2004.
- SEEGER, M.; SALINAS, D.; FLUNKERT, V. **DeepAR: Probabilistic forecasting with autoregressive recurrent networks**. *International Journal of Forecasting*, 2021.
- SHMUELI, G.; LICHTENDAHL JR., K.C. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. O'Reilly Media, 2017.
- SILVER, E.A.; PYKE, D.F.; PETERSON, R. **Inventory Management and Production Planning and Scheduling**. Wiley, 1998.
- SMIL, V. **Energy and Civilization: A History**. MIT Press, 2017.
- SNIJDERS, C.; MATZAT, U.; REIPS, U.D. **'Big Data': Big gaps of knowledge in the field of Internet**. *International Journal of Internet Science*, v. 7, n. 1, p. 1-5, 2012.
- STERN, N. **The Economics of Climate Change: The Stern Review**. Cambridge University Press, 2007.
- TAYLOR, S.J.; LETHAM, B.; **Prophet: Forecasting at scale**. *The American Statistician*, v. 72, n. 1, p. 37-43, 2021.
- TUKEY, J.W. **Exploratory Data Analysis**. Addison-Wesley, 1977.
- TULARAM, Gurudeo Anand; SAEED, Tareq. **Oil-Price Forecasting Based on Various Univariate Time-Series Models**. *American Journal of Operations Research*, [s.l.], v. 6, p. 226-235, 2016.
- WRIGHT, G.; GOODWIN, P. **Decision making and planning under low levels of predictability: Enhancing the scenario method**. *International Journal of Forecasting*, v. 25, n. 4, p. 813-825, 2009.
- YERGIN, D. **The Quest: Energy, Security, and the Remaking of the Modern World**. Penguin Books, 2011.