

Universidade Federal de Uberlândia

Instituto de Ciências Exatas e Naturais do Pontal

Curso de Matemática

Trabalho de Conclusão de Curso

**Modelagem e Previsão de Preços do
Minério de Ferro: Um Estudo
Utilizando ARIMA e *Random Forest*
*Regressor***

por

Giulio Cesar Mirabella Maramarque

Bacharelado em Matemática - Ituiutaba - MG

Orientadora: Profa. Dra. Milena Almeida L. Brandão

Co-orientadora: Arianne Alves S. Mundim

Giulio Cesar Mirabella Maramarque

Modelagem e Previsão de Preços do Minério de Ferro: Um
Estudo Utilizando ARIMA e *Random Forest Regressor*

Trabalho de Conclusão de Curso apresentado
ao Instituto de Ciências Exatas e Naturais
do Pontal da Universidade Federal de
Uberlândia como requisito parcial para ob-
tenção do título de Bacharel em Matemática.

Orientadora: Profa. Dra. Milena Almeida
Leite Brandão

Co-orientadora: Arianne Alves S. Mundim

Ituiutaba, Abril de 2024

Giulio Cesar Mirabella Maramarque

Modelagem e Previsão de Preços do Minério de Ferro: Um
Estudo Utilizando ARIMA e *Random Forest Regressor*

Trabalho de Conclusão de Curso apresentado ao Instituto de Ciências Exatas e Naturais do Pontal da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de Bacharel em Matemática.

Orientadora: Profa. Dra. Milena Almeida Leite Brandão

Co-orientadora: Arianne Alves S. Mundim

BANCA EXAMINADORA:

Profa. Dra. Milena Almeida Leite Brandão

Profa. Dra. Gabriella de Freitas Alves

Prof. Dr. Homero Ghioti da Silva

Ituiutaba, Abril de 2024

*Dedico este trabalho aos meus
amados avós, Girolamo
Mirabella e Maria Pia Tiezzi
Mirabella, cujo amor, apoio e
inspiração são fontes
inesgotáveis de força e
sabedoria em minha vida.
Seus valores moldaram meu
ser, e é com imensa gratidão
que compartilho este trabalho
com vocês. Obrigado por
serem minha luz e minha
âncora.*

AGRADECIMENTOS

Gostaria de expressar minha sincera gratidão a todos os meus professores da graduação por seu apoio e orientação ao longo desta jornada acadêmica. Não me restam dúvidas de que, sem seu apoio, eu não estaria apto a concluir esta etapa da minha vida. Serei eternamente grato a todos vocês e jamais poderei retribuir tudo o que me fizeram. Suas contribuições foram inestimáveis e fundamentais para meu crescimento profissional e pessoal.

Em especial, gostaria de agradecer aos Professores Drs.: Alisson, Milena e Homero. Também um agradecimento especial para Arianne Mundim por aceitar o convite para coorientar este trabalho. Sua dedicação e insights foram essenciais para o sucesso deste projeto.

Também gostaria de estender meu agradecimento à Exed Consulting, empresa onde trabalho, por fornecer o ambiente e os recursos necessários para minha atuação na área e meu desenvolvimento profissional. Sua contínua colaboração e suporte foram fundamentais para minha formação.

Finalmente, gostaria de agradecer a todos os amigos, familiares e colegas que me apoiaram ao longo desta jornada. Suas palavras de incentivo e apoio foram valiosas e muito apreciadas.

RESUMO

Este estudo analisa o desempenho de dois modelos de previsão estatística aplicados ao mercado de minério de ferro. O primeiro modelo adotado é o ARIMA, uma abordagem tradicional amplamente utilizada em previsão temporal, enquanto o segundo é o *Random Forest Regressor*, uma técnica de aprendizado de máquina mais recente e flexível. O objetivo é avaliar a capacidade de cada modelo em prever os preços futuros do minério de ferro com base em dados históricos. Para isso, foram empregadas duas métricas de erro, são elas: MAPE (Erro Percentual Absoluto Médio) e MAE (Erro Absoluto Médio). A análise foi conduzida utilizando o histórico de preços do minério de ferro desde janeiro de 1980 até outubro de 2023 e o incremento de variáveis externas como: Produção de Aço e Consumo de Aço, ambas no mercado brasileiro com dados de janeiro de 2013 a outubro 2023. Os resultados obtidos foram submetidos a uma análise comparativa detalhada, destacando as vantagens e desvantagens de cada modelo em termos de precisão de previsão. Esta pesquisa proporciona uma análise para o mercado de minério de ferro e oferece uma base para futuros estudos e desenvolvimentos na área de previsão de preços de commodities.

Palavras-chave: previsão estatística, séries temporais, aprendizado de máquina.

CONTEÚDO

1	Introdução	1
1.1	Contexto	1
1.2	GAP	2
1.3	Objetivos	3
1.3.1	Objetivo Geral	3
1.3.2	Objetivos Específicos	3
2	Materiais e Métodos	4
2.1	Métodos	4
2.1.1	Previsão de Demanda	4
2.1.2	Dados	5
2.1.3	<i>Machine Learning</i>	6
2.1.4	Análise de Séries Temporais	7
2.1.5	ARIMA	8
2.1.6	<i>Random Forest Regressor</i>	9
2.2	Métricas	10
2.2.1	MAPE	10
2.2.2	MAE	11
3	Experimentos Computacionais	12
3.1	Códigos	12

3.1.1	<i>Python</i> - Previsão estatística com ARIMA	12
3.1.2	<i>Python</i> - Previsão estatística com <i>Random Forest Regressor</i>	14
3.2	Análise do Histórico de Preços	19
3.2.1	Histórico	19
3.2.2	Identificação de <i>Outliers</i>	20
3.3	Análise de Correlação	22
3.3.1	Correlação do Preço do Minério de Ferro com Produção de Aço no Brasil	22
3.3.2	Correlação do Preço do Minério de Ferro com Consumo de Aço no Brasil	23
3.3.3	Comparação de Correlações	24
3.4	Resultados dos Modelos	26
3.4.1	Métricas Modelo ARIMA	26
3.4.2	Métricas Modelo <i>Random Forest Regressor</i>	27
3.4.3	Comparação das Métricas em Média	31
3.5	Assertividade dos Modelos	32
3.5.1	Gráfico de Erros	32
4	Considerações Finais	34
4.1	Conclusão	34
4.2	Trabalhos Futuros	35
	Referências	36

LISTA DE FIGURAS

3.1	Código ARIMA PT.1	13
3.2	Código ARIMA PT.2	14
3.3	Código RFR PT.1	15
3.4	Código RFR PT.2	16
3.5	Código RFR PT.3	17
3.6	Código RFR PT.4	18
3.7	Preço Médio Anual.	19
3.8	Variação Últimos 15 anos.	20
3.9	<i>Outliers</i>	21
3.10	Correlação de Produção e Preço	23
3.11	Correlação de Consumo e Preço	24
3.12	Comparação de Correlações	25
3.13	Métricas Modelo ARIMA.	26
3.14	Métricas Modelo <i>Random Forest Regressor</i> - Cenário 1.	28
3.15	Métricas Modelo <i>Random Forest Regressor</i> - Cenário 2.	30
3.16	Métricas Médias por Modelo.	32
3.17	Erro Médio dos Modelos.	33

CAPÍTULO 1

INTRODUÇÃO

1.1 Contexto

A previsão de demanda é uma tarefa desafiadora devido à complexidade e volatilidade inerentes aos mercados contemporâneos. Tradicionalmente, métodos como Autorregressivo Integrado de Médias Móveis - ARIMA (Auto Regressive Integrated Moving Average) e Autorregressivo Integrado de Médias Móveis com Sazonalidade - SARIMA (Seasonal Autorregressive Integrated Moving Average) são empregados para realizar estimativas. No entanto, com o advento de grandes conjuntos de dados e avanços na capacidade computacional, novas técnicas de aprendizado de máquina têm se destacado, proporcionando uma abordagem mais dinâmica e adaptativa para lidar com a variabilidade dos padrões de demanda.

A utilização de algoritmos de aprendizado profundo, como Redes Neurais, Inferência Causal e Árvores Aleatórias de Regressão vem ganhando espaço na modelagem de séries temporais, como foi evidenciado por [1], possibilitando a identificação de padrões temporais de curto, médio e longo prazo. Essa capacidade de representação hierárquica e aprendizado de dependências temporais complexas contribui para a melhoria da precisão das previsões, especialmente em ambientes empresariais onde a demanda é afetada por múltiplos fatores interconectados.

Fazer previsão estatística de demanda, significa estimar quantidades futuras de pro-

dados necessárias, o que influencia diretamente outras etapas da gestão da cadeia de suprimentos de empresas e indústrias.

Ao longo do capítulo de Materiais e Métodos, serão apresentados os conceitos atribuídos a este processo de previsão, são eles: Previsão de Demanda, *Machine Learning*, Análise de Séries Temporais, além dos algoritmos ARIMA e *Random Forest Regressor* e da motivação e importância da utilização do Minério de Ferro como base de dados.

No capítulo Experimentos Computacionais, serão apresentados os códigos em *Python* utilizados para ler, processar e prever os dados. Também serão apresentados os resultados gerados por cada modelo, as métricas de erro em tabela e um gráfico comparando as barras das médias dos erros.

Por fim, na Conclusão será realizado o fechamento do trabalho, explicitando os principais ganhos com a possibilidade de realizar a previsão estatística com técnicas de aprendizado de máquina. Ademais, será feita uma provocação final, deixando o caminho aberto para novos trabalhos de mesmo viés, porém utilizando algoritmos novos, visto que esses algoritmos são atualizados e lançados com frequência.

1.2 GAP

O presente estudo concentra-se na análise de desempenho de um modelo tradicional de previsão estatística (ARIMA), e de uma técnica mais recente de aprendizado de máquina, *Random Forest Regressor*, no contexto da previsão estatística.

No entanto, para abordar as lacunas (GAPs) identificadas no universo da previsão estatística, este trabalho se propõe a realizar uma análise sobre novas tecnologias disponíveis para aprimorar a precisão das previsões estatísticas utilizando a inteligência artificial por meio de técnicas de aprendizado de máquina.

A abordagem das lacunas inclui a exploração do método *Random Forest Regressor*, destacado por [2], com a possibilidade de utilizar variáveis independentes que façam sentido com o contexto de negócio analisado. Por isso, em seu modelo preditivo criamos dois cenários. No primeiro consideramos a variável de Produção do Aço no Brasil, já no segundo, a variável de Consumo de Aço no Brasil, ambas disponíveis em: [3]. A escolha dessas variáveis pode ser embasada na opinião de profissionais experientes do setor ou por meio do cálculo de correlação, apresentado no capítulo 3. Essas variáveis compõem um modelo de previsão multivariado, que podemos acompanhar por [4].

Ao incorporar essa análise, o estudo visa contribuir significativamente para a compreensão do estado atual das tecnologias aplicadas à previsão estatística.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo deste trabalho de conclusão de curso é avaliar a aplicabilidade e desempenho dos modelos: ARIMA e *Random Forest Regressor*, bem como apresentar as novas possibilidades que uma técnica de aprendizado de máquina pode nos proporcionar.

1.3.2 Objetivos Específicos

Avaliar a assertividade dos modelos de previsão com *Random Forest Regressor* e com o tradicional modelo ARIMA, por meio das métricas: Erro Percentual Absoluto Médio - MAPE (*Mean Absolute Percentual Error*), Erro Médio Absoluto - MAE (*Mean Absolute Error*).

Ao decorrer do trabalho explicamos como as fórmulas matemáticas se aplicam nos cálculos de todos esses algoritmos e também nas medidas de erro.

CAPÍTULO 2

MATERIAIS E MÉTODOS

2.1 Métodos

2.1.1 Previsão de Demanda

O sucesso de qualquer empreendimento está intrinsecamente ligado à sua capacidade de antecipar e responder às demandas do mercado. No âmbito empresarial, essa capacidade é amplamente respaldada pela previsão estatística, um processo vital que permeia diversas áreas funcionais, incluindo Logística, Marketing, Produção e Finanças.

A previsão de demanda, como visto em [5], é mais do que uma mera estimativa de volumes futuros; é uma entrada fundamental para o planejamento estratégico e o controle operacional. Em um ambiente competitivo e dinâmico, a precisão das estimativas de demanda é essencial para a redução de custos e o aprimoramento dos níveis de serviço oferecidos aos clientes, entre outros benefícios.

Uma das principais implicações da previsão de demanda é a gestão eficaz do *LeadTime*. Essa medida, que representa o tempo decorrido entre o pedido e a entrega de um produto, é crucial para a eficiência operacional e a satisfação do cliente.

A previsão de demanda, essencialmente, busca antecipar as necessidades futuras, mitigando assim os riscos associados à tomada de decisão. Segundo [6], é um processo metodológico que se baseia em modelos matemáticos e especulações sobre o comportamento do mercado consumidor.

Como apresentado em [6], é importante distinguir entre eventos externos incontroláveis e internos controláveis ao realizar previsões de demanda. Enquanto eventos externos, como mudanças na economia ou no clima, são incertos e exigem estratégias adaptativas, eventos internos, como iniciativas de marketing ou ajustes na cadeia de suprimentos, podem ser influenciados e otimizados para atender às demandas previstas.

De acordo com [6], existem várias abordagens para a previsão de demanda, podemos dividi-las em métodos qualitativos, de séries temporais e de previsão causal. Cada método possui suas próprias vantagens e limitações, e a escolha adequada depende da natureza dos dados disponíveis, modelo de negócio e das características específicas do mercado de cada cliente.

Além disso, a previsão de demanda pode ser realizada em diferentes horizontes de tempo, abrangendo curto, médio e longo prazo. Isso permite que as organizações adaptem suas estratégias de acordo com as tendências de mercado e os objetivos de longo prazo.

Em resumo, a previsão de demanda é uma ferramenta essencial para a gestão eficaz de negócios, apoiando a melhoria contínua no planejamento estratégico. Ao integrar abordagens analíticas e conhecimento contextual, as empresas podem posicionar-se de forma competitiva em um cenário empresarial em constante evolução.

2.1.2 Dados

Neste estudo, utilizamos a variação do preço do minério de ferro, disponível em [7], como a variável alvo a ser prevista, utilizando duas abordagens estatísticas distintas: ARIMA e *Random Forest Regressor*. A escolha dessa variável como base de dados se deve à sua relevância tanto em escala nacional quanto global, dadas suas amplas implicações econômicas e industriais. Outro fator determinante para escolhermos a variável, é o fato dela ser atualizada mensalmente e com isso podemos acompanhar o desempenho dos modelos ao longo do tempo e inclusive ao decorrer do trabalho.

- (1) Visão Nacional (Brasil): No contexto brasileiro, a variação do preço do minério de ferro é de extrema importância devido ao país ser um dos principais produtores e exportadores mundiais desse recurso. O setor de mineração desempenha um papel significativo na economia brasileira, contribuindo para o PIB (Produto Interno Bruto) e gerando empregos. A flutuação dos preços do minério de ferro afeta diretamente a rentabilidade das empresas do setor e, conseqüentemente, a economia

nacional como um todo.

- (2) Visão Global: Globalmente, o preço do minério de ferro é um indicador crucial para a saúde da indústria siderúrgica e para a economia global. O minério de ferro é um dos principais ingredientes na produção de aço, utilizado em uma ampla gama de setores, incluindo construção, automotivo e manufatura. Portanto, as flutuações nos preços do minério de ferro têm ramificações significativas em diversas economias ao redor do mundo, afetando o investimento em infraestrutura, o custo de produção e a competitividade das indústrias.
- (3) Variáveis Externas: Além da variação do preço do minério de ferro, incorporamos duas variáveis externas relevantes em nosso modelo de *Machine Learning*: a produção de aço no Brasil e o consumo de aço no Brasil, ambas disponíveis em [3], com dados de janeiro de 2013 até outubro de 2023. A escolha dessas variáveis foi feita a critério do autor, a justificativa é trazer um olhar para o mercado brasileiro de construção civil, correlacionando a produção e o consumo de aço em no Brasil, com o preço do minério de ferro.

Ao integrar essas variáveis externas ao nosso modelo, buscamos capturar melhor a complexidade e os fatores que influenciam a variação do preço do minério de ferro, permitindo previsões mais precisas e robustas.

2.1.3 *Machine Learning*

Desde a sua concepção por Arthur Samuel em 1959 [8], o campo de *Machine Learning* (Aprendizado de Máquina) tem sido uma força transformadora na maneira como os computadores interagem com informações e resolvem problemas. Samuel definiu-o como o estudo que confere aos computadores a capacidade de aprender sem serem explicitamente programados para tal. Uma definição mais formal de Tom Mitchel [8], ampliou essa ideia, descrevendo o aprendizado de máquina como um programa de computador que aprende por experiência E em relação a uma tarefa T e mede sua performance P . Se a performance em T , medida por P , melhora com a experiência E , então o programa está de fato aprendendo.

Inserido no campo mais amplo da Inteligência Artificial, o *Machine Learning* tem como missão principal dotar as máquinas da capacidade de aprender e evoluir, adaptando-se a

novos padrões e informações sem a necessidade de reprogramação direta. Essa abordagem permite que sistemas computacionais se tornem autônomos na identificação de padrões complexos e soluções para uma ampla gama de problemas em diversas áreas.

Segundo [9], o objetivo fundamental do *Machine Learning* é desenvolver algoritmos capazes de aprender automaticamente associações em grandes conjuntos de dados, em tempo real. Ao estudar técnicas de modelagem de processos computacionais, o *Machine Learning* capacita as máquinas a realizar análises avançadas e oferecer *insights* em situações complexas que muitas vezes escapam à capacidade humana de processamento.

Este método de análise de dados representa um ramo essencial da Inteligência Artificial, buscando identificar padrões com base em dados brutos. Ao contrário de métodos convencionais, não depende de regras programadas explicitamente, mas sim da capacidade de aprender com experiências e ajustar-se dinamicamente a novas informações.

Assim, o *Machine Learning* emerge como uma ferramenta poderosa, transformando a abordagem tradicional de resolução de problemas e análise de dados. À medida que as máquinas aprendem e evoluem, elas se tornam parceiras inteligentes na busca por soluções eficazes em um mundo cada vez mais complexo e interconectado. Seu potencial para impulsionar a inovação e a eficiência em diversas indústrias é vasto, e seu papel na moldagem do futuro da tecnologia é inegável.

2.1.4 Análise de Séries Temporais

A análise de séries temporais é uma ferramenta fundamental para compreender e prever o comportamento de conjuntos de observações ao longo do tempo. Conforme [10], uma série temporal é essencialmente um conjunto de observações $Y(t)$, t pertencente a T , onde Y representa a variável de interesse e T é o conjunto de índices que denotam o tempo.

Segundo [11], o principal objetivo da análise de séries temporais é entender o mecanismo subjacente que gera os dados e, a partir disso, fazer previsões sobre o comportamento futuro da série. Em sua essência, uma série temporal é composta por observações ordenadas cronologicamente, permitindo a identificação de padrões e tendências ao longo do tempo.

Como visto em [11], uma das componentes mais importantes na análise de séries temporais é a tendência. A tendência revela o comportamento geral da série ao longo do tempo, indicando se a série está crescendo, decrescendo ou se mantendo estável.

Além da tendência, outro aspecto crucial é a sazonalidade. De acordo com [11], a

sazonalidade refere-se a padrões que se repetem em intervalos de tempo bem definidos, muitas vezes associados a eventos previsíveis, como estações do ano, feriados ou ciclos de produção.

Ao combinar análise estatística avançada e técnicas de modelagem, a análise de séries temporais permite fazer previsões precisas em uma ampla gama de contextos, desde finanças e economia até meteorologia e epidemiologia. Dominar essa metodologia é importante para estruturar políticas eficazes em um mundo onde a compreensão do tempo é essencial para o sucesso das corporações.

2.1.5 ARIMA

O modelo ARIMA, abreviação para *Autoregressive Integrated Moving Average*, é uma generalização do modelo ARMA (*Autoregressive Moving Average*), combinando os processos autorregressivo (AR) e de médias móveis (MA), com a adição de um componente de diferenciação (I), representado pelo termo (p, d, q) , termo esse que será definido a seguir.

A representação matemática do modelo ARIMA é dada por:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (2.1)$$

Nesta fórmula:

- Y_t representa a série temporal no tempo t .
- c é uma constante.
- $\phi_1, \phi_2, \dots, \phi_p$ são os parâmetros de autorregressão.
- $\theta_1, \theta_2, \dots, \theta_q$ são os parâmetros de média móvel.
- ϵ_t é o termo de erro aleatório no tempo t .

Definido em [1], este é o modelo ARIMA(p, d, q), onde p é a ordem da autorregressão, d é a ordem de diferenciação e q é a ordem da média móvel.

Como explicito em [10], o ciclo de inferência do modelo ARIMA pode ser dividido em quatro etapas:

- (1) Identificação – Nesta fase, o objetivo é identificar a estrutura apropriada do modelo ARIMA para uma dada série temporal. Em nosso caso, o modelo ARIMA é apropriado por estamos tratando uma série não-estacionária.

- (2) Estimação – uma vez que a estrutura do modelo ARIMA tenha sido identificada, os parâmetros do modelo devem ser estimados. Para realizar a estimação de parâmetros, observamos a autocorrelação e autocorrelação parcial dos dados, utilizando um código em *Python*, apresentado no capítulo 3.
- (3) Verificação – Após a estimação dos parâmetros do modelo, é igualmente importante verificar a adequação do modelo. Isso geralmente envolve a análise dos resíduos do modelo para determinar se há algum padrão significativo deixado para trás.
- (4) Previsão - Finalmente, após a identificação e estimativa do modelo e a verificação da previsão, o modelo ARIMA é utilizado para obter previsões futuras da série temporal.

Em suma, o modelo ARIMA é uma ferramenta estruturada e versátil para lidar com a modelagem e previsão em séries temporais. Ao seguir um processo de análise bem definido, os analistas conseguem extrair informações estratégicas com base em dados e realizar boas previsões em diversos cenários.

2.1.6 *Random Forest Regressor*

Segundo [5], o algoritmo *Random Forest Regressor* é uma extensão poderosa do método de árvores de decisão, que capitaliza a técnica de Dividir-para-Conquistar. Sobre as *Random Forest*, evidenciadas em [12], se destacam em relação a outras técnicas de modelagem, oferecendo uma série de vantagens distintas:

- Algoritmo Poderoso: Em comparação com uma única árvore de decisão, as *Random Forests* se mostram significativamente mais poderosas, capazes de lidar com uma variedade maior de problemas de modelagem.
- Boa Taxa de Acerto: Um dos pontos fortes das *Random Forest Regressor* é sua capacidade de obter boas taxas de acerto ao ser testada em diferentes conjuntos de dados. Isso as torna confiáveis e versáteis em uma variedade de contextos de aplicação.
- Técnica Exata: As *Random Forests* são conhecidas por sua precisão, fornecendo resultados confiáveis e consistentes ao lidar com problemas de regressão.

- Evita *Overfitting*: Um desafio comum em modelos de aprendizado de máquina é o *overfitting*, definido em [13], no entanto, as *Random Forest Regressor* são menos suscetíveis a esse problema, graças à sua abordagem de treinamento por meio de múltiplas árvores de decisão.
- Menos Sensíveis a Ruídos: Outra vantagem das *Random Forest Regressor* é sua robustez em relação a ruídos nos dados. Elas conseguem generalizar padrões mesmo em presença de informações irrelevantes ou ruidosas.
- Classificação Aleatória das Árvores: As árvores de decisão em uma *Random Forest* são construídas com diferentes subconjuntos de atributos selecionados aleatoriamente a partir do conjunto original. Isso introduz diversidade nas árvores e reduz a correlação entre elas, o que contribui para um modelo mais robusto e generalizável.

Em suma, o *Random Forest Regressor* é um algoritmo poderoso e flexível para modelagem de problemas de regressão. Sua capacidade de evitar *overfitting*, lidar com ruídos nos dados e produzir resultados precisos o torna uma escolha popular para uma ampla gama de aplicações em ciência de dados e aprendizado de máquina. Ao explorar a natureza aleatória e diversificada das *Random Forests*, os profissionais podem realizar previsões precisas em seus projetos de análise de dados.

2.2 Métricas

2.2.1 MAPE

Erro de previsão MAPE, definido em [14] (*Mean Absolute Percentual Error*)

O MAPE é, se não a mais, uma das mais importantes métricas para avaliar a precisão dos modelos de previsão. Calculado como a média dos erros percentuais absolutos entre os valores previstos e os valores reais, fornece uma medida direta da acurácia das previsões. No planejamento de demanda, o MAPE desempenha um papel vital ao ajudar a entender a confiabilidade das previsões e orientar decisões estratégicas sobre alocação de recursos, produção e estoque.

A representação matemática da métrica de erro MAPE é dada por:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|, \quad (2.2)$$

onde:

- Y_i são os valores observados.
- \hat{Y}_i são os valores previstos.
- n é o número total de observações.

2.2.2 MAE

Erro de previsão MAE, evidenciado em [14] (*Mean Absolute Error*)

O MAE é uma métrica essencial na avaliação da precisão dos modelos de previsão. Calculado como a média dos valores absolutos das diferenças entre os valores previstos e os valores reais, o MAE oferece uma medida direta da magnitude dos erros de previsão. Também desempenha um papel crucial ao fornecer uma avaliação clara da precisão das previsões, permitindo que as empresas identifiquem e corrijam imprecisões nos modelos de previsão e otimizem seus processos de planejamento.

A representação matemática da métrica de erro MAE é dada por:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, \quad (2.3)$$

onde:

- Y_i são os valores observados.
- \hat{Y}_i são os valores previstos.
- n é o número total de observações.

CAPÍTULO 3

EXPERIMENTOS COMPUTACIONAIS

3.1 Códigos

3.1.1 *Python* - Previsão estatística com ARIMA

Nas figuras 3.1 e 3.2, há os códigos desenvolvidos pelo autor, na linguagem *Python*, responsáveis pela construção da previsão estatística com ARIMA.

O código em *Python* em questão desempenha um papel crucial na análise e previsão dos dados, prevendo os meses de novembro de 2023 até março de 2024. Sua funcionalidade primordial consiste em ler e processar os dados disponíveis, preparando-os para aplicação no algoritmo ARIMA. O código é capaz de gerar previsões para os preços do minério de ferro nos períodos desejados. Além disso, a inclusão das métricas MAPE e MAE proporciona uma avaliação da precisão das previsões, permitindo uma análise do desempenho do modelo em relação aos dados reais. Os valores de p,d,q para o algoritmo ARIMA foram escolhidos com base nas autocorrelações e autocorrelações parciais, citadas em [16] e desenvolvidas pelo código.

Em suma, o código não apenas prevê os valores futuros, mas também oferece uma avaliação da qualidade dessas previsões para decisões futuras no contexto do mercado de minério de ferro.

Figura 3.1: Código ARIMA PT.1

```

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

from statsmodels.tsa.arima.model import ARIMA
df_ARIMA = r"C:\Users\giuli\OneDrive\Área de Trabalho\TCC\Dataset\ironprice_ARIMA.csv"
try:
    dados = pd.read_csv(df_ARIMA, sep=';', encoding='utf-8-sig')
    dados['Data'] = pd.to_datetime(dados['Data'])
    dados = dados.set_index('Data')
    print("\nDados carregados com sucesso:")
    print(dados.head())
except Exception as e:
    print("\nErro ao carregar os dados do arquivo CSV:", e)

from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt

fig, ax = plt.subplots(2, 1, figsize=(12, 8))
plot_acf(dados, ax=ax[0])
plot_pacf(dados, ax=ax[1])
plt.show()

p = 1
q = 0
d = 1

modelo = ARIMA(dados, order=(p, d, q))
resultado = modelo.fit()

previsao = resultado.forecast(steps=12)
print("Previsão para os próximos 12 meses:")
print(previsao)

dados_reais_nov = [126.15]
previsao_nov = [118.206874]

dados_reais_dez = [130.46]
previsao_dez = [118.408967]

dados_reais_jan = [136.37]
previsao_jan = [118.611060]

dados_reais_fev = [128.63]
previsao_fev = [118.813153]

dados_reais_mar = [116.94]
previsao_mar = [119.015247]

from sklearn.metrics import mean_absolute_error, r2_score

dados_reais_nov = np.array(dados_reais_nov)
previsao_nov = np.array(previsao_nov)

mape_nov = np.mean(np.abs((dados_reais_nov - previsao_nov) / dados_reais_nov)) * 100
mae_nov = mean_absolute_error(dados_reais_nov, previsao_nov)

print("Métricas de Erro para Novembro:")
print("MAPE:", mape_nov)
print("MAE:", mae_nov)

dados_reais_dez = np.array(dados_reais_dez)
previsao_dez = np.array(previsao_dez)

mape_dez = np.mean(np.abs((dados_reais_dez - previsao_dez) / dados_reais_dez)) * 100
mae_dez = mean_absolute_error(dados_reais_dez, previsao_dez)

print("Métricas de Erro para Dezembro:")
print("MAPE:", mape_dez)
print("MAE:", mae_dez)

dados_reais_jan = np.array(dados_reais_jan)
previsao_jan = np.array(previsao_jan)

mape_jan = np.mean(np.abs((dados_reais_jan - previsao_jan) / dados_reais_jan)) * 100
mae_jan = mean_absolute_error(dados_reais_jan, previsao_jan)

```

Fonte: O autor (2024)

Figura 3.2: Código ARIMA PT.2

```
print("Métricas de Erro para Janeiro:")
print("MAPE:", mape_jan)
print("MAE:", mae_jan)

dados_reais_fev = np.array(dados_reais_fev)
previsao_fev = np.array(previsao_fev)

mape_fev = np.mean(np.abs((dados_reais_fev - previsao_fev) / dados_reais_fev)) * 100
mae_fev = mean_absolute_error(dados_reais_fev, previsao_fev)

print("Métricas de Erro para Fevereiro:")
print("MAPE:", mape_fev)
print("MAE:", mae_fev)

dados_reais_mar = np.array(dados_reais_mar)
previsao_mar = np.array(previsao_mar)

mape_mar = np.mean(np.abs((dados_reais_mar - previsao_mar) / dados_reais_mar)) * 100
mae_mar = mean_absolute_error(dados_reais_mar, previsao_mar)

print("Métricas de Erro para Março:")
print("MAPE:", mape_mar)
print("MAE:", mae_mar)

erros_nov = {'MAPE': mape_nov, 'MAE': mae_nov}
erros_dez = {'MAPE': mape_dez, 'MAE': mae_dez}
erros_jan = {'MAPE': mape_jan, 'MAE': mae_jan}
erros_fev = {'MAPE': mape_fev, 'MAE': mae_fev}
erros_mar = {'MAPE': mape_mar, 'MAE': mae_mar}

tabela_erros = pd.DataFrame([erros_nov, erros_dez, erros_jan, erros_fev, erros_mar],
                           index=['Novembro', 'Dezembro', 'Janeiro', 'Fevereiro', 'Março'])

print("Métricas de Erro mensal, Modelo ARIMA:")
print(tabela_erros)

mape_erro = [mape_nov, mape_dez, mape_jan, mape_fev, mape_mar]
mae_erro = [mae_nov, mae_dez, mae_jan, mae_fev, mae_mar]

media_mape_erro = sum(mape_erro) / len(mape_erro)
media_mae_erro = sum(mae_erro) / len(mae_erro)

print("Média dos Erros:", media_mape_erro, media_mae_erro)
```

Fonte: O autor (2024)

3.1.2 *Python* - Previsão estatística com *Random Forest Regressor*

A seguir, nas figuras 3.3, 3.4, 3.5 e 3.6 temos os códigos, desenvolvidos pelo autor em *Python*, responsáveis pela construção da previsão estatística com *Random Forest Regressor*.

Figura 3.3: Código RFR PT.1

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.preprocessing import StandardScaler

ironore = r"C:\Users\giuli\OneDrive\Área de Trabalho\TCC\Dataset\ironprice_ARIMA.csv"

try:
    df = pd.read_csv(ironore, sep=';', encoding='utf-8-sig')
    print("
Dados carregados com sucesso:")
    print(df.head())
except Exception as e:
    print("
Erro ao carregar os dados do arquivo CSV:", e)

df['Data'] = pd.to_datetime(df['Data'])
df = df.set_index('Data')
df.head()

df['year'] = df.index.year
df['month'] = df.index.month
df["ProdAço"] = df["ProdAço"].str.replace(',','.')
df["ConsumAço"] = df["ConsumAço"].str.replace(',','.')

X = df[['year', 'month', 'ProdAço', 'ConsumAço']]
y = df['Price']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train_scaled, y_train)

y_pred = rf_model.predict(X_test_scaled)
mae = mean_absolute_error(y_test, y_pred)
print(f'Mean Absolute Error: {mae}')

input1_nov23 = {'year': [2023], 'month': [11], 'ProdAço': [7567.84], 'ConsumAço': [0]}
input1_dez23 = {'year': [2023], 'month': [12], 'ProdAço': [7198.06], 'ConsumAço': [0]}
input1_jan24 = {'year': [2024], 'month': [1], 'ProdAço': [7551.29], 'ConsumAço': [0]}
input1_fev24 = {'year': [2024], 'month': [2], 'ProdAço': [7701.58], 'ConsumAço': [0]}
input1_mar24 = {'year': [2024], 'month': [3], 'ProdAço': [7703.34], 'ConsumAço': [0]}

x1_nov23 = pd.DataFrame(input1_nov23)
x1_dez23 = pd.DataFrame(input1_dez23)
x1_jan24 = pd.DataFrame(input1_jan24)
x1_fev24 = pd.DataFrame(input1_fev24)
x1_mar24 = pd.DataFrame(input1_mar24)

def print_dataframe_as_table(df):
    print(df.to_string(index=False))

# Exibindo os dataframes como tabelas
print("x1_nov23:")
print_dataframe_as_table(x1_nov23)
print("
x1_dez23:")
print_dataframe_as_table(x1_dez23)
print("
x1_jan24:")
print_dataframe_as_table(x1_jan24)
print("
x1_fev24:")
print_dataframe_as_table(x1_fev24)
print("
x1_mar24:")
print_dataframe_as_table(x1_mar24)

```

Fonte: O autor (2024)

Figura 3.4: Código RFR PT.2

```

ypred1_nov23 = rf_model.predict(x1_nov23)
ypred1_dez23 = rf_model.predict(x1_dez23)
ypred1_jan24 = rf_model.predict(x1_jan24)
ypred1_fev24 = rf_model.predict(x1_fev24)
ypred1_mar24 = rf_model.predict(x1_mar24)

df1_nov23 = pd.DataFrame({'Novembro 2023': ypred1_nov23})
df1_dez23 = pd.DataFrame({'Dezembro 2023': ypred1_dez23})
df1_jan24 = pd.DataFrame({'Janeiro 2024': ypred1_jan24})
df1_fev24 = pd.DataFrame({'Fevereiro 2024': ypred1_fev24})
df1_mar24 = pd.DataFrame({'Março 2024': ypred1_mar24})

df1_concatenado = pd.concat([df1_nov23, df1_dez23, df1_jan24, df1_fev24, df1_mar24], axis=1)
print(df1_concatenado)

from sklearn.metrics import mean_absolute_error

dados_reais_nov = [126.15]
pred1_nov23 = ypred1_nov23

dados_reais_dez = [130.46]
pred1_dez23 = ypred1_dez23

dados_reais_jan = [136.37]
pred1_jan24 = ypred1_jan24

dados_reais_fev = [128.63]
pred1_fev24 = ypred1_fev24

dados_reais_mar = [116.94]
pred1_mar24 = ypred1_mar24

dados_reais_nov = np.array(dados_reais_nov)
pred1_nov23 = np.array(pred1_nov23)
mapel_nov23 = np.mean(np.abs((dados_reais_nov - pred1_nov23) / dados_reais_nov)) * 100
mael_nov23 = mean_absolute_error(dados_reais_nov, pred1_nov23)
print("[CENÁRIO 1] Métricas de Erro para Novembro:")
print("MAPE:", mapel_nov23)
print("MAE:", mael_nov23)

dados_reais_dez = np.array(dados_reais_dez)
pred1_dez23 = np.array(pred1_dez23)
mapel_dez23 = np.mean(np.abs((dados_reais_dez - pred1_dez23) / dados_reais_dez)) * 100
mael_dez23 = mean_absolute_error(dados_reais_dez, pred1_dez23)
print("[CENÁRIO 1] Métricas de Erro para Dezembro:")
print("MAPE:", mapel_dez23)
print("MAE:", mael_dez23)

dados_reais_jan = np.array(dados_reais_jan)
pred1_jan24 = np.array(pred1_jan24)
mapel_jan24 = np.mean(np.abs((dados_reais_jan - pred1_jan24) / dados_reais_jan)) * 100
mael_jan24 = mean_absolute_error(dados_reais_jan, pred1_jan24)
print("[CENÁRIO 1] Métricas de Erro para Janeiro:")
print("MAPE:", mapel_jan24)
print("MAE:", mael_jan24)

dados_reais_fev = np.array(dados_reais_fev)
pred1_fev24 = np.array(pred1_fev24)
mapel_fev24 = np.mean(np.abs((dados_reais_fev - pred1_fev24) / dados_reais_fev)) * 100
mael_fev24 = mean_absolute_error(dados_reais_fev, pred1_fev24)
print("[CENÁRIO 1] Métricas de Erro para Fevereiro:")
print("MAPE:", mapel_fev24)
print("MAE:", mael_fev24)

dados_reais_mar = np.array(dados_reais_mar)
pred1_mar24 = np.array(pred1_mar24)
mapel_mar24 = np.mean(np.abs((dados_reais_mar - pred1_mar24) / dados_reais_mar)) * 100
mael_mar24 = mean_absolute_error(dados_reais_mar, pred1_mar24)
print("[CENÁRIO 1] Métricas de Erro para Março:")
print("MAPE:", mapel_mar24)
print("MAE:", mael_mar24)

erros1_nov = {'MAPE': mapel_nov23, 'MAE': mael_nov23}
erros1_dez = {'MAPE': mapel_dez23, 'MAE': mael_dez23}
erros1_jan = {'MAPE': mapel_jan24, 'MAE': mael_jan24}
erros1_fev = {'MAPE': mapel_fev24, 'MAE': mael_fev24}
erros1_mar = {'MAPE': mapel_mar24, 'MAE': mael_mar24}
tabela_erros = pd.DataFrame([erros1_nov, erros1_dez, erros1_jan, erros1_fev, erros1_mar],
                           index=['Novembro', 'Dezembro', 'Janeiro', 'Fevereiro', 'Março'])

```

Figura 3.5: Código RFR PT.3

```

input2_nov23 = {'year': [2023], 'month': [11], 'ProdAço': [0], 'ConsuAço': [1963.96]}
input2_dez23 = {'year': [2023], 'month': [12], 'ProdAço': [0], 'ConsuAço': [1916.18]}
input2_jan24 = {'year': [2024], 'month': [1], 'ProdAço': [0], 'ConsuAço': [1946.82]}
input2_fev24 = {'year': [2024], 'month': [2], 'ProdAço': [0], 'ConsuAço': [1896.92]}
input2_mar24 = {'year': [2024], 'month': [3], 'ProdAço': [0], 'ConsuAço': [1894.89]}

x2_nov23 = pd.DataFrame(input2_nov23)
x2_dez23 = pd.DataFrame(input2_dez23)
x2_jan24 = pd.DataFrame(input2_jan24)
x2_fev24 = pd.DataFrame(input2_fev24)
x2_mar24 = pd.DataFrame(input2_mar24)

def print_dataframe_as_table(df):
    print(df.to_string(index=False))

# Exibindo os dataframes como tabelas
print("x2_nov23:")
print_dataframe_as_table(x2_nov23)
print("x2_dez23:")
print_dataframe_as_table(x2_dez23)
print("x2_jan24:")
print_dataframe_as_table(x2_jan24)
print("x2_fev24:")
print_dataframe_as_table(x2_fev24)
print("x2_mar24:")
print_dataframe_as_table(x2_mar24)

ypred2_nov23 = rf_model.predict(x2_nov23)
ypred2_dez23 = rf_model.predict(x2_dez23)
ypred2_jan24 = rf_model.predict(x2_jan24)
ypred2_fev24 = rf_model.predict(x2_fev24)
ypred2_mar24 = rf_model.predict(x2_mar24)

df2_nov23 = pd.DataFrame({'Novembro 2023': ypred2_nov23})
df2_dez23 = pd.DataFrame({'Dezembro 2023': ypred2_dez23})
df2_jan24 = pd.DataFrame({'Janeiro 2024': ypred2_jan24})
df2_fev24 = pd.DataFrame({'Fevereiro 2024': ypred2_fev24})
df2_mar24 = pd.DataFrame({'Março 2024': ypred2_mar24})
df2_concatenado = pd.concat([df2_nov23, df2_dez23, df2_jan24, df2_fev24, df2_mar24], axis=1)
print(df2_concatenado)

dados_reais_nov = [126.15]
pred2_nov23 = ypred2_nov23

dados_reais_dez = [130.46]
pred2_dez23 = ypred2_dez23

dados_reais_jan = [136.37]
pred2_jan24 = ypred2_jan24

dados_reais_fev = [128.63]
pred2_fev24 = ypred2_fev24

dados_reais_mar = [116.94]
pred2_mar24 = ypred2_mar24

dados_reais_nov = np.array(dados_reais_nov)
pred2_nov23 = np.array(pred2_nov23)
mape2_nov23 = np.mean(np.abs((dados_reais_nov - pred2_nov23) / dados_reais_nov)) * 100
mae2_nov23 = mean_absolute_error(dados_reais_nov, pred2_nov23)
print("[CENÁRIO 2] Métricas de Erro para Novembro:")
print("MAPE:", mape2_nov23)
print("MAE:", mae2_nov23)

dados_reais_dez = np.array(dados_reais_dez)
pred2_dez23 = np.array(pred2_dez23)
mape2_dez23 = np.mean(np.abs((dados_reais_dez - pred2_dez23) / dados_reais_dez)) * 100
mae2_dez23 = mean_absolute_error(dados_reais_dez, pred2_dez23)
print("[CENÁRIO 2] Métricas de Erro para Dezembro:")
print("MAPE:", mape2_dez23)
print("MAE:", mae2_dez23)

```

Fonte: O autor (2024)

Figura 3.6: Código RFR PT.4

```

dados_reais_jan = np.array(dados_reais_jan)
pred2_jan24 = np.array(pred2_jan24)
mape2_jan24 = np.mean(np.abs((dados_reais_jan - pred2_jan24) / dados_reais_jan)) * 100
mae2_jan24 = mean_absolute_error(dados_reais_jan, pred2_jan24)
print("[CENÁRIO 2] Métricas de Erro para Janeiro:")
print("MAPE:", mape2_jan24)
print("MAE:", mae2_jan24)

dados_reais_fev = np.array(dados_reais_fev)
pred2_fev24 = np.array(pred2_fev24)
mape2_fev24 = np.mean(np.abs((dados_reais_fev - pred2_fev24) / dados_reais_fev)) * 100
mae2_fev24 = mean_absolute_error(dados_reais_fev, pred2_fev24)
print("[CENÁRIO 2] Métricas de Erro para Fevereiro:")
print("MAPE:", mape2_fev24)
print("MAE:", mae2_fev24)

dados_reais_mar = np.array(dados_reais_mar)
pred2_mar24 = np.array(pred2_mar24)
mape2_mar24 = np.mean(np.abs((dados_reais_mar - pred2_mar24) / dados_reais_mar)) * 100
mae2_mar24 = mean_absolute_error(dados_reais_mar, pred2_mar24)
print("[CENÁRIO 2] Métricas de Erro para Março:")
print("MAPE:", mape2_mar24)
print("MAE:", mae2_mar24)

erros2_nov = {'MAPE': mape2_nov23, 'MAE': mae2_nov23, 'R²': r2_pred2}
erros2_dez = {'MAPE': mape2_dez23, 'MAE': mae2_dez23, 'R²': r2_pred2}
erros2_jan = {'MAPE': mape2_jan24, 'MAE': mae2_jan24, 'R²': r2_pred2}
erros2_fev = {'MAPE': mape2_fev24, 'MAE': mae2_fev24, 'R²': r2_pred2}
erros2_mar = {'MAPE': mape2_mar24, 'MAE': mae2_mar24, 'R²': r2_pred2}
tabela_erros2 = pd.DataFrame([erros2_nov, erros2_dez, erros2_jan, erros2_fev, erros2_mar],
                             index=['Novembro', 'Dezembro', 'Janeiro', 'Fevereiro', 'Março'])

mapel_erro = [mapel_nov23, mapel_dez23, mapel_jan24, mapel_fev24, mapel_mar24]
mael_erro = [mael_nov23, mael_dez23, mael_jan24, mael_fev24, mael_mar24]
media_mapel_erro = sum(mapel_erro) / len(mapel_erro)
media_mael_erro = sum(mael_erro) / len(mael_erro)
print("Média dos Erros:", media_mapel_erro, media_mael_erro)

mape2_erro = [mape2_nov23, mape2_dez23, mape2_jan24, mape2_fev24, mape2_mar24]
mae2_erro = [mae2_nov23, mae2_dez23, mae2_jan24, mae2_fev24, mae2_mar24]
media_mape2_erro = sum(mape2_erro) / len(mape2_erro)
media_mae2_erro = sum(mae2_erro) / len(mae2_erro)
print("Média dos Erros:", media_mape2_erro, media_mae2_erro)

```

Fonte: O autor (2024)

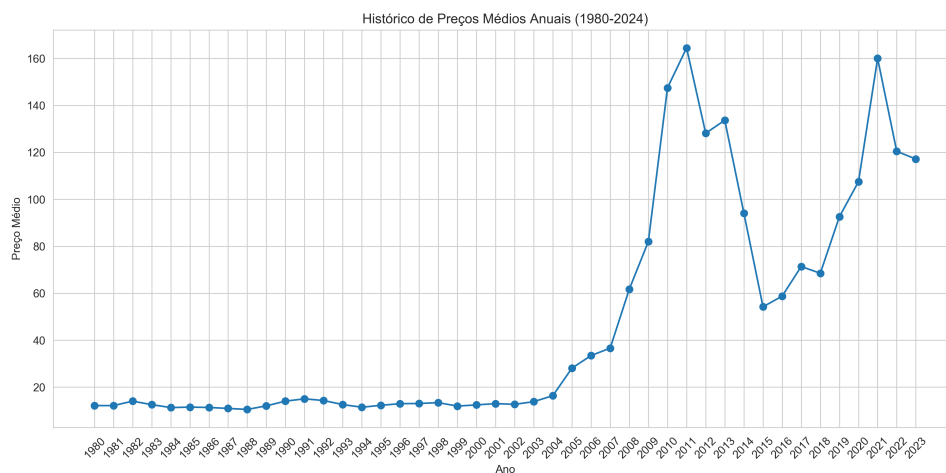
Novamente, o código *Python* realiza previsão do preço para os meses de novembro de 2023 até março de 2024, construindo dois cenários distintos, designados como Cenário 1 e Cenário 2. Sua função é ler, processar e prever os dados utilizando o algoritmo *Random Forest Regressor*, conhecido por sua capacidade de lidar com conjuntos de dados complexos e variáveis explicativas. Por meio dessa abordagem, o código é capaz de gerar previsões para os preços do minério de ferro no futuro especificado. Adicionalmente, ao mensurar os valores das métricas MAPE e MAE, o código proporciona uma avaliação objetiva da qualidade das previsões em ambos os cenários, permitindo uma comparação entre diferentes abordagens e fornecendo boas análises para o planejamento estratégico no mercado de minério de ferro.

3.2 Análise do Histórico de Preços

3.2.1 Histórico

A figura 3.7 abaixo, desenvolvida pelo autor, representa a distribuição da base de dados que estamos trabalhando, disponível em [7]. Temos o histórico de preços do minério de ferro desde janeiro de 1980 até outubro de 2023, que nos proporciona analisar as tendências de longo prazo, flutuações sazonais e eventos significativos que impactaram o mercado ao longo das décadas.

Figura 3.7: Preço Médio Anual.



Fonte: O autor (2024)

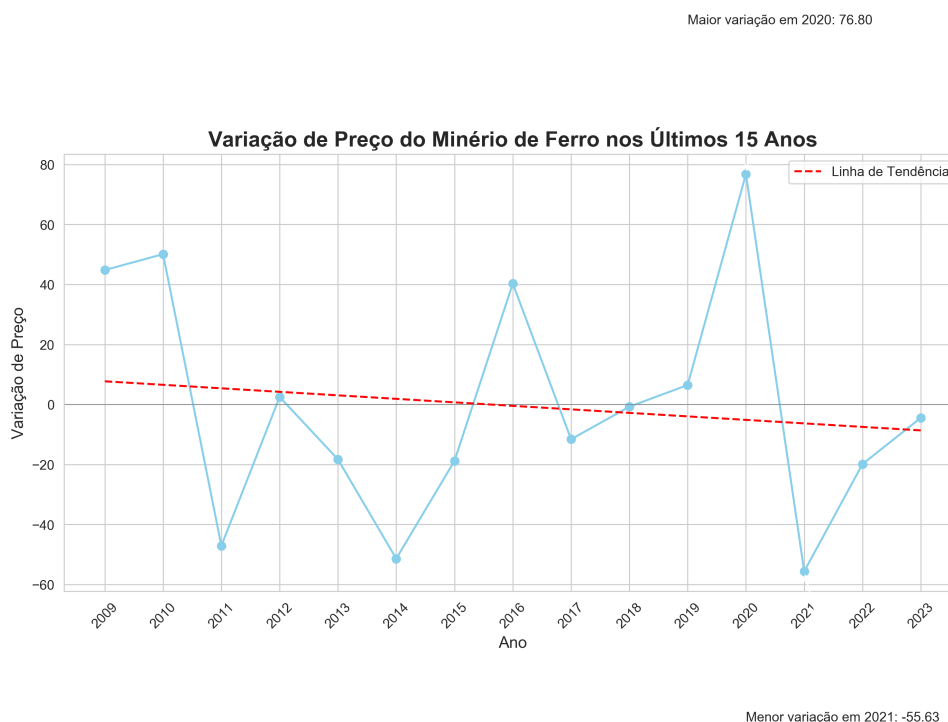
Ao analisar o histórico de preços do minério de ferro ao longo dos anos, podemos concluir que houve uma variação considerável nos preços, com tendência geral de aumento ao longo do tempo. Observamos flutuações sazonais e alguns picos e vales, refletindo volatilidade no mercado. Esses dados sugerem a influência de fatores econômicos e sazonais na determinação dos preços, que nos influenciou a buscar variáveis externas relacionadas ao aço para otimizar o processo preditivo.

Variação nos Últimos 15 anos

Na figura 3.8, apresentada abaixo, temos um recorte mais recente do histórico de preços do minério de ferro, desenvolvido pelo autor, apresentando a análise da variação

dos preços nos últimos 15 anos. A escolha do período de análise se deu a critério do autor, nos permite uma compreensão mais abrangente das tendências e flutuações desse importante indicador econômico.

Figura 3.8: Variação Últimos 15 anos.



Fonte: O autor (2024)

Ao analisar o recorte recente, podemos concluir que existe uma tendência de queda da variação do preço do minério de ferro, especialmente nos anos onde a pandemia do coronavírus estava em alta. Neste período, tivemos uma alta dos preços do minério de ferro, mas uma baixa variação. Essa observação sugere uma possível interferência direta da pandemia na alta dos preços e em sua constância, tendo em vista a queda da variação de preços desta commodity.

Seguiremos analisando o histórico.

3.2.2 Identificação de *Outliers*

O método utilizado para identificação de *outliers* é conhecido como “método do inter-

valo interquartil” (IQR), disponível em [15]. Este método utiliza os quartis Q1 e Q3 para calcular o intervalo interquartil (IQR), que é a diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1).

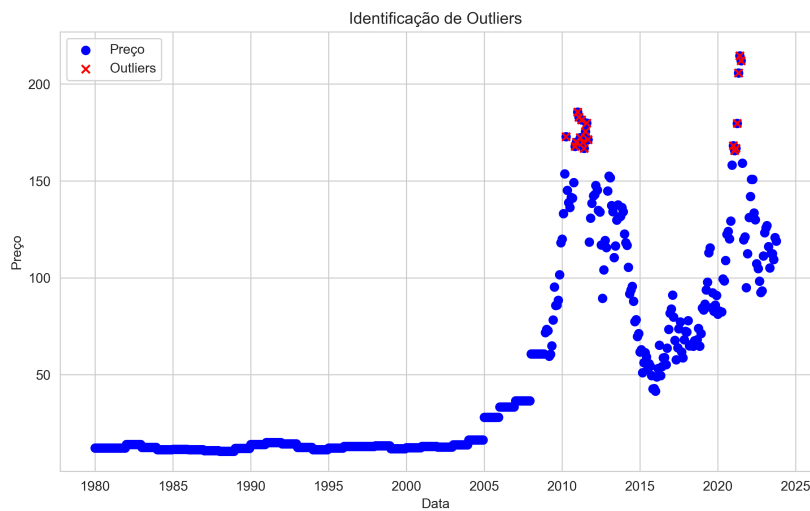
Os *outliers* são então identificados como valores que caem fora do intervalo definido pelos limites inferior e superior, onde:

Limite inferior: $Q1 - 1.5 \times IQR$

Limite superior: $Q3 + 1.5 \times IQR$

Conforme pode ser visto na figura 3.9, os valores que estão abaixo do limite inferior ou acima do limite superior são considerados outliers.

Figura 3.9: *Outliers*.



Fonte: O autor (2024)

Como observado no histórico de preços do minério de ferro, foram identificados *outliers* que ultrapassaram apenas o limite superior estabelecido pelo método interquartil. Isso sugere que, ao longo do período analisado, houve momentos em que os preços do minério de ferro apresentaram variações significativas para cima, indicando possíveis eventos ou condições excepcionais no mercado. Esses *outliers* podem ter sido influenciados por fatores como mudanças na demanda global, interrupções na oferta, políticas governamentais ou flutuações macroeconômicas. Concluimos por aqui a análise do histórico de preços.

3.3 Análise de Correlação

Neste estudo, realizamos uma análise de correlação das variáveis selecionadas para integrar o modelo *Random Forest Regressor* com o histórico de preços do minério de ferro. As correlações foram calculadas com as funções: Função de Autocorrelação - ACF (*Autocorrelation Function*) e Função de Autocorrelação Parcial - PACF (*Partial Autocorrelation Function*). Ambas as funções são citadas em [16]. Estas, são as mais indicadas para fazermos correlação de variáveis em séries temporais, sendo esse o motivo de sua escolha por parte do autor.

As variáveis em análise são “ProdAço”, representando a Produção de Aço no Brasil, e “ConsuAço”, representando o consumo de aço no Brasil. O período analisado compreende os anos de 2013 a 2023. Ambas variáveis estão disponíveis em [3].

Lag

Definido em [16], um “lag”, no contexto de série temporais, é simplesmente um intervalo de tempo entre observações. Como estamos trabalhando com dados mensais, os períodos atribuídos as funções são meses. Para esta análise de correlação, $lag = 10$, então os períodos representados seriam os 10 meses anteriores ao período de interesse. Se estamos prevendo novembro de 2023 até março de 2024, os períodos representados por $lag = 10$ seriam janeiro de 2023 até outubro de 2023. A escolha de 10 meses se deu a critério do autor, para que possamos ter uma visão do ano de 2023, tendo uma análise de correlação recente. Esta interpretação de lag é crucial para o entendimento das figuras abaixo, desenvolvidas pelo autor, onde apresentamos os valores calculado de correlação entre o histórico de preços do minério de ferro e as variáveis escolhidas.

3.3.1 Correlação do Preço do Minério de Ferro com Produção de Aço no Brasil

A figura 3.10, exibida abaixo e desenvolvida pelo autor, nos apresenta uma tabela com os valores de ACF e PACF para a variável ‘ProdAço’ e o preço do minério de ferro.

Figura 3.10: Correlação de Produção e Preço

Correlação da Variável ProdAço com o Preço

Lag	ACF	PACF
1.0	0.614722	0.619487
2.0	0.511019	0.219461
3.0	0.380005	0.003533
4.0	0.237255	-0.099225
5.0	0.226735	0.101956
6.0	0.124964	-0.068233
7.0	0.165898	0.12823
8.0	0.054282	-0.156841
9.0	0.101242	0.136076
10.0	0.08539	-0.017056

Fonte: O autor (2024)

Com base nos valores de correlação apresentados e lembrando do $lag = 10$ citado anteriormente, podemos concluir que a correlação diminui significativamente nos 10 meses anteriores a nossa previsão.

3.3.2 Correlação do Preço do Minério de Ferro com Consumo de Aço no Brasil

Na figura 3.11, exibida abaixo e desenvolvida pelo autor, são apresentados os valores de correlação da variável 'ConsuAço' com o preço do minério de ferro.

Figura 3.11: Correlação de Consumo e Preço

Correlação da Variável ConsuAço com o Preço

Lag	ACF	PACF
1.0	0.810111	0.816391
2.0	0.757108	0.307173
3.0	0.653367	-0.055942
4.0	0.607144	0.072346
5.0	0.577626	0.131166
6.0	0.522251	-0.0496
7.0	0.450224	-0.132035
8.0	0.391973	-0.017082
9.0	0.337328	-0.005661
10.0	0.349518	0.17531

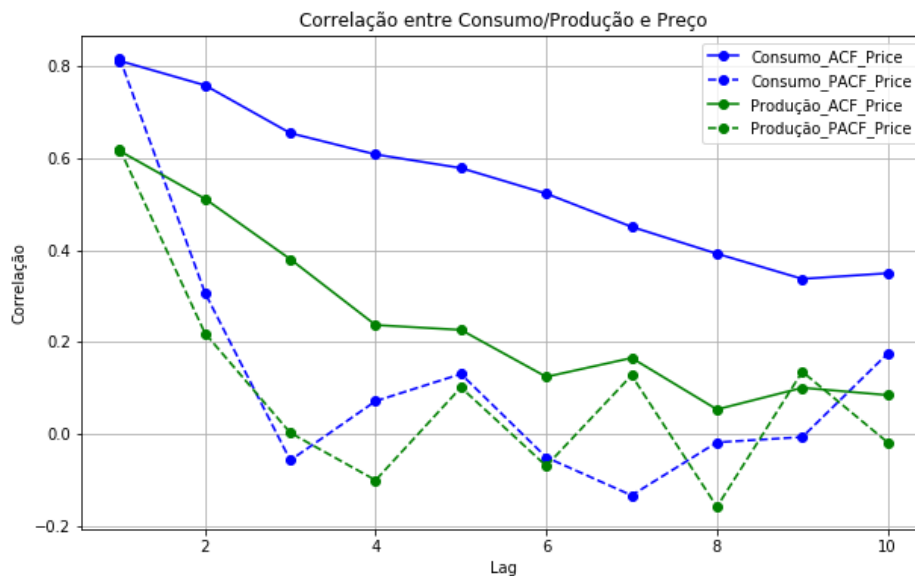
Fonte: O autor (2024)

Analisando os valores de ACF e PACF da figura, concluímos que assim como a variável de 'ProdAço', os valores de correlação também diminuem conforme nos aproximamos do mês previsto.

3.3.3 Comparação de Correlações

Abaixo, na figura 3.12, são comparados os valores de ACF e PACF das variáveis de Produção e Consumo de aço no Brasil com o histórico de preços do minério de ferro, gerados em nosso cálculo.

Figura 3.12: Comparação de Correlações



Fonte: O autor (2024)

Com base nos resultados da análise de correlação, observamos que a variável “ConsumoAço” apresenta uma autocorrelação (ACF) mais forte com o histórico de preços de minério de ferro, especialmente nos períodos inicialmente analisados, em comparação com a variável “ProdAço”. Quando analisados os valores de autocorrelação parcial (PACF), temos resultados com maior variação.

Os valores calculados sugerem que as flutuações no consumo de aço no Brasil, têm uma influência mais significativa nos preços do minério de ferro do que a produção de aço, sendo um dos fatores para que o modelo atinja uma maior precisão, como veremos nos resultados.

Com base nessas observações, podemos concluir que o consumo de aço é uma variável importante a ser considerada ao analisar e prever os preços do minério de ferro, e seu impacto mostrará ser mais significativo do que a produção de aço.

3.4 Resultados dos Modelos

3.4.1 Métricas Modelo ARIMA

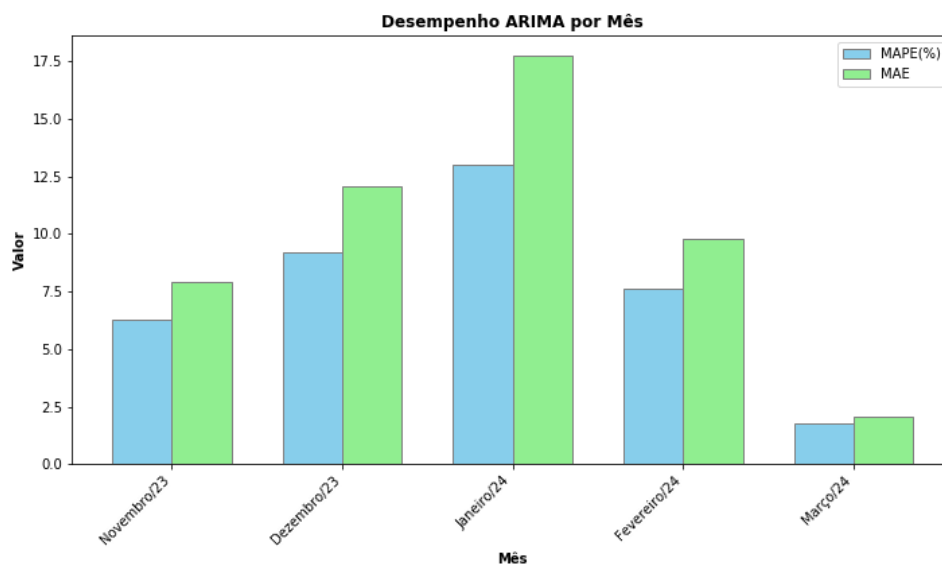
A seguir, são apresentadas a tabela 3.1 e a figura 3.13 exibindo as métricas de erro selecionadas para avaliação do desempenho do modelo ARIMA ao longo dos meses.

Tabela 3.1: Desempenho Mensal do Modelo ARIMA.

	MAPE(%)	MAE
Novembro/23	6,29	7,94
Dezembro/23	9,23	12,05
Janeiro/24	13,02	17,75
Fevereiro/24	7,63	9,81
Março/24	1,77	2,07

Fonte: O autor (2024)

Figura 3.13: Métricas Modelo ARIMA.



Fonte: O autor (2024)

Análise dos Resultados Obtidos

MAPE (Erro Percentual Médio Absoluto): Os valores de MAPE variaram consideravelmente, atingindo seu ponto mais baixo em março de 2024 (1,77%) e seu ponto mais alto em janeiro de 2024 (13,02%). Isso sugere uma variação significativa na precisão do modelo ao longo do tempo.

MAE (Erro Absoluto Médio): Similarmente ao MAPE, o MAE exibiu uma ampla variação ao longo dos meses, com os valores mais altos em janeiro de 2024 (17,75) e os mais baixos em março de 2024 (2,07).

Com base na análise dos valores de MAPE e MAE ao longo dos meses, é possível concluir que o desempenho do modelo ARIMA apresentou flutuações consideráveis durante o período analisado.

A variação significativa observada nos valores de MAPE e MAE sugere que o modelo teve dificuldades em manter uma consistência em sua precisão ao longo do tempo. Essa inconsistência pode ser atribuída a uma série de fatores, como mudanças na natureza dos dados, variações sazonais ou mesmo inadequações no próprio modelo.

3.4.2 Métricas Modelo *Random Forest Regressor*

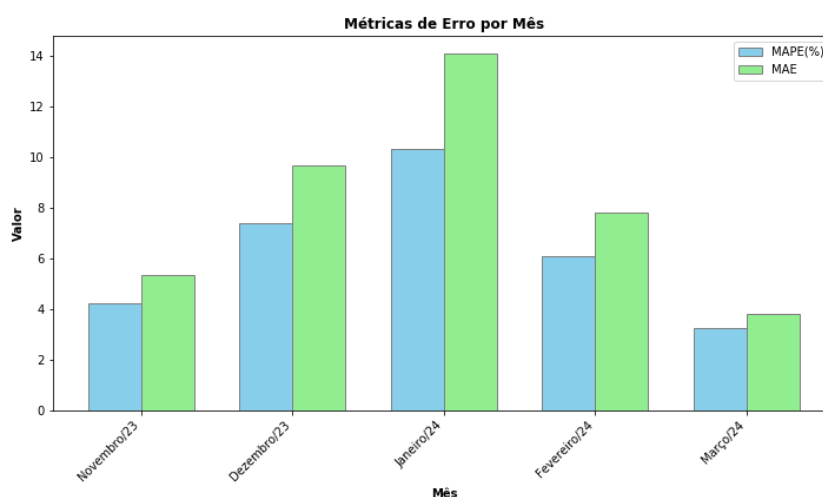
Cenário 1

No primeiro cenário, emprega-se a Produção de Aço no Brasil como variável independente no modelo *Random Forest Regressor*. Esta abordagem é adotada para prever o preço do minério de ferro, onde a relação entre a produção de aço e o preço do minério de ferro é explorada como um fator-chave na modelagem preditiva. Os valores das métricas de erro para este cenário são apresentados pela tabela 3.2 e figura 3.14 abaixo.

Tabela 3.2: Desempenho Mensal do Modelo *Random Forest Regressor* - Cenário 1.

	MAPE(%)	MAE
Novembro/23	4,25	5,36
Dezembro/23	7,41	9,67
Janeiro/24	10,34	14,11
Fevereiro/24	6,09	7,84
Março/24	3,29	3,84

Fonte: O autor (2024)

Figura 3.14: Métricas Modelo *Random Forest Regressor* - Cenário 1.

Fonte: O autor (2024)

Análise dos Resultados Obtidos

Os resultados indicam que o modelo *Random Forest Regressor* possui um desempenho razoável na previsão do preço do minério de ferro, com a produção de aço no Brasil sendo adicionada como variável externa. A variação de 3,29% a 10,34% ao longo dos meses analisados, sugere que o modelo pode ter dificuldades em capturar variações sazonais ou eventos específicos que afetam o preço do minério de ferro nesse período. Por outro lado, o MAE mostra uma variação timidamente mais estável, indicando uma maior capacidade

do modelo em prever o preço do minério de ferro, apesar das flutuações na produção de aço. Em resumo, esses resultados destacam a importância de ajustes contínuos no modelo para melhorar sua capacidade de capturar fatores influentes no preço do minério de ferro, garantindo previsões mais precisas e confiáveis.

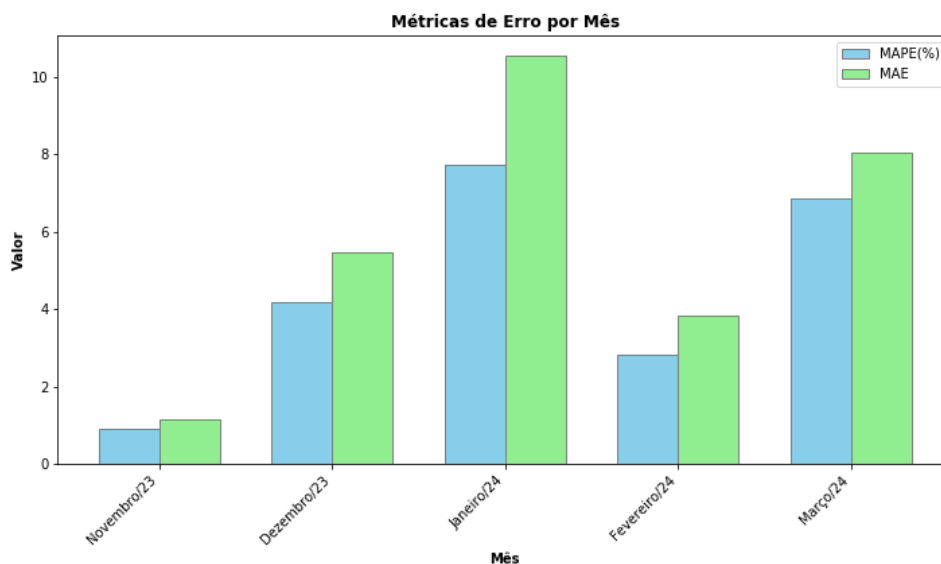
Cenário 2

O segundo cenário adota o Consumo de Aço no Brasil como variável independente, buscando oferecer uma perspectiva alternativa na previsão do preço do minério de ferro. Neste contexto, a análise das métricas de erro para o modelo *Random Forest Regressor* proporciona uma visão abrangente do desempenho preditivo, permitindo a comparação e avaliação da eficácia do modelo sob diferentes conjuntos de variáveis externas. Os erros são apresentados na tabela 3.3 e figura 3.15 a seguir.

Tabela 3.3: Desempenho Mensal do Modelo *Random Forest Regressor* - Cenário 2.

	MAPE(%)	MAE
Novembro/23	0,92	1,16
Dezembro/23	4,19	5,47
Janeiro/24	7,74	10,55
Fevereiro/24	2,83	3,84
Março/24	6,87	8,04

Fonte: O autor (2024)

Figura 3.15: Métricas Modelo *Random Forest Regressor* - Cenário 2.

Fonte: O autor (2024)

Análise dos Resultados Obtidos

MAPE (Erro Percentual Absoluto Médio): O MAPE varia de 0,92% a 7,74% ao longo dos diferentes períodos. Esses valores indicam uma precisão relativamente alta na previsão do preço do minério de ferro, com o menor erro percentual registrado em novembro de 2023 e o maior em janeiro de 2024. No geral, os resultados sugerem uma tendência de aumento do erro percentual dos meses iniciais, podendo ser refletida nos meses subsequentes, tendo em vista que, após uma queda no erro de fevereiro de 2024, temos um aumento em março de 2024.

MAE (Erro Absoluto Médio): O MAE varia de 1,16 a 10,55 unidades nos diferentes períodos avaliados. Assim como o MAPE, observa-se uma tendência de aumento do erro absoluto em relação aos meses seguintes, com o menor erro absoluto registrado em novembro de 2023 e o maior em janeiro de 2024. Os resultados indicam uma certa consistência na capacidade do modelo em prever o preço do minério de ferro, apesar das flutuações no consumo de aço no Brasil.

Comparação entre Cenário 1 e 2

Em comparação com o primeiro cenário, onde a Produção de Aço no Brasil foi considerada como variável externa, os resultados sugerem uma possível melhoria na precisão do modelo ao utilizar o Consumo de Aço como variável independente. Isso é evidenciado principalmente pela redução geral nos valores de MAPE e MAE, indicando uma maior capacidade do modelo em capturar os padrões subjacentes que influenciam o preço do minério de ferro.

Em suma, os resultados apresentados sugerem que o modelo *Random Forest Regressor*, quando alimentado com o Consumo de Aço no Brasil como variável independente, exibe uma boa capacidade de prever o preço do minério de ferro.

3.4.3 Comparação das Métricas em Média

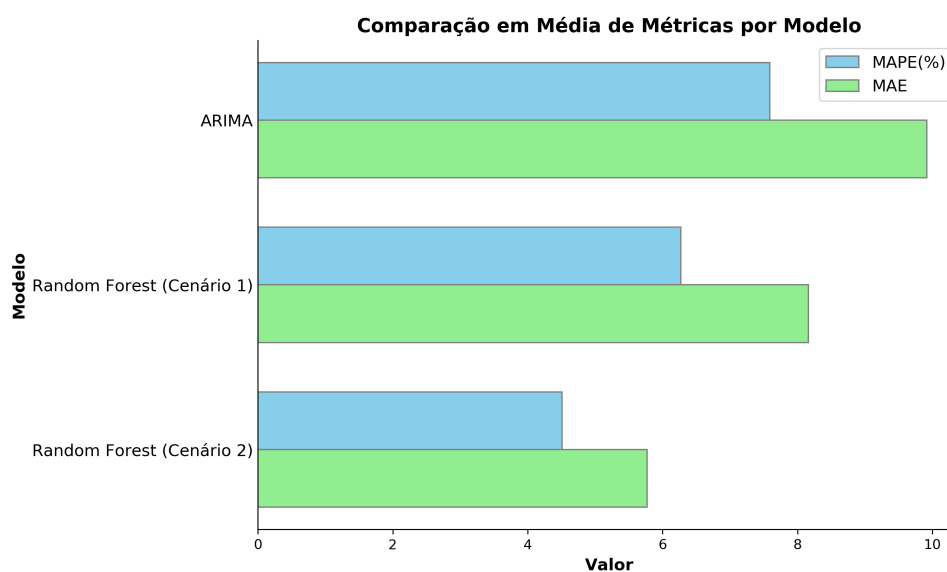
A tabela 3.4 e a figura 3.16 representam os valores médios, das métricas escolhidas para o trabalho, obtidos por cada um dos modelos de previsão construídos até aqui.

Tabela 3.4: Métricas de desempenho dos modelos.

	MAPE(%)	MAE
ARIMA	7,59	9,92
<i>Random Forest Regressor</i> (Cenário 1)	6,27	8,16
<i>Random Forest Regressor</i> (Cenário 2)	4,51	5,77

Fonte: O autor (2024)

Figura 3.16: Métricas Médias por Modelo.



Fonte: O autor (2024)

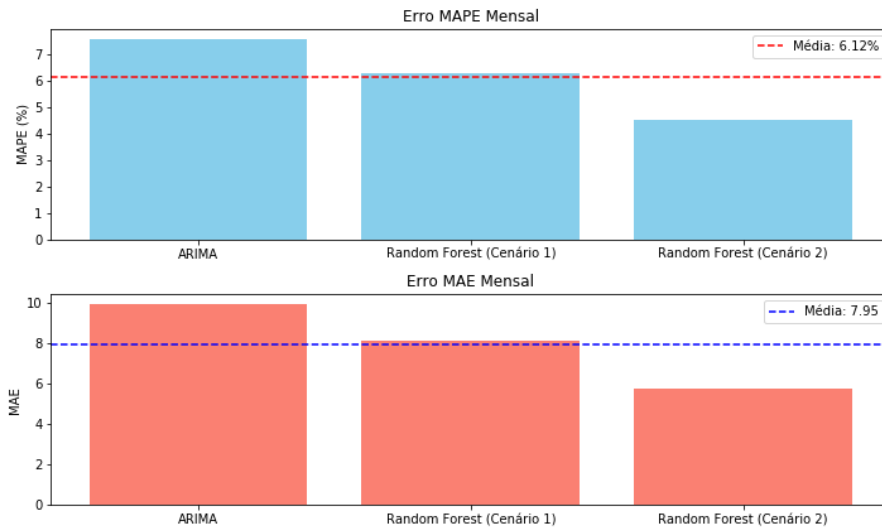
Ao analisar os erros médios de cada modelo, podemos concluir que o *Random Forest Regressor* no segundo cenário, utilizando o Consumo de Aço no Brasil como variável independente, apresenta o melhor desempenho, com um MAPE de 4,51% e um MAE de 5,77. Isso sugere uma maior precisão na previsão do preço do minério de ferro em comparação com o ARIMA e o *Random Forest Regressor* no primeiro cenário. Esses resultados destacam a importância de considerar variáveis relevantes e atualizadas para aprimorar a precisão dos modelos de previsão.

3.5 Assertividade dos Modelos

3.5.1 Gráfico de Erros

No gráfico apresentado na Figura 3.17 mostramos os valores em média das métricas de erro selecionadas para realizar a previsão. Além disso, a linha pontilhada em cada gráfico, mostra a média dos valores observados.

Figura 3.17: Erro Médio dos Modelos.



Fonte: O autor (2024)

Como podemos observar, o *Random Forest Regressor* obteve maior precisão, especialmente no cenário 2 (considerando o Consumo de Aço no Brasil), ao prever o preço do minério de ferro ao longo dos últimos meses.

Na conclusão a seguir, iremos realizar um fechamento do trabalho e da análise como um todo, explicitando os motivos pelos quais técnicas de *Machine Learning*, a exemplo do *Random Forest Regressor*, tem se tornado o centro das atenções quando falamos de algoritmos e previsões estatísticas.

CAPÍTULO 4

CONSIDERAÇÕES FINAIS

4.1 Conclusão

Ao encerrar este estudo, evidenciamos como o advento da Inteligência Artificial, por meio de técnicas de *Machine Learning*, como o *Random Forest Regressor*, desempenha um papel significativo no aprimoramento e precisão dos modelos estatísticos e matemáticos. Neste contexto específico, exploramos o mercado do ferro, utilizando as variáveis de Produção e Consumo de Aço no Brasil como guias para a construção de previsões pelo algoritmo *Random Forest Regressor*. Os resultados obtidos foram satisfatórios, conforme o esperado, a acuracidade do método *Random Forest Regressor* foi maior do que o tradicional método ARIMA. Ressaltamos que não buscamos, de maneira nenhuma, desmerecer o Método ARIMA, mas sim apresentar melhorias e benefícios tangíveis ao adotar abordagens mais recentes de previsão.

Com este trabalho de conclusão de curso, concluímos uma etapa importante, mas não final, de uma jornada acadêmica e científica. Reconhecemos a necessidade contínua de explorar novos métodos e técnicas, bem como a aplicação desses conhecimentos em diversas áreas do mercado.

Ao fechar estas páginas, deixamos espaço para reflexão e sugestões de futuras investigações. Abaixo, apresentamos ideias para trabalhos futuros, visando contribuir ainda mais para o avanço do conhecimento nesta área.

4.2 Trabalhos Futuros

Ao considerar as perspectivas para futuras investigações, é crucial manter uma abordagem flexível e aberta à exploração de diversas metodologias e técnicas. Além de repetir a abordagem deste estudo com algoritmos distintos e métodos tradicionais alternativos, como mencionado, também é promissor explorar outras técnicas de *Machine Learning* e estar atento a novos algoritmos emergentes que possam oferecer vantagens adicionais. O processo de previsão estatística de demanda é dinâmico e em constante evolução, requerendo uma mentalidade de melhoria contínua.

Outra consideração importante é o nível de granularidade das informações utilizadas. Enquanto este estudo adotou uma abordagem de granularidade simples, focada no preço do minério de ferro, há um potencial significativo para explorar outras áreas de negócio com granularidades mais complexas. Isso poderia envolver a análise de dados em diferentes escalas, segmentando o mercado de forma mais detalhada e incorporando uma variedade de variáveis para melhor compreender e prever o comportamento da demanda.

Portanto, os trabalhos futuros podem se concentrar não apenas em expandir as técnicas e algoritmos utilizados, mas também em explorar novos domínios de aplicação e níveis de detalhe nas análises, garantindo assim uma abordagem abrangente e robusta para a previsão de demanda em diferentes contextos de negócios.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] SCHLENDER, S.G.; CERETTA, P. S.; RIGHI, M. B. Previsão do preço da soja: uma comparação entre os modelos ARIMA e redes neurais artificiais. **Informações Econômicas**, São Paulo, v. 40, n. 9, p. 16-27, set. 2010.
- [2] TELOKEN, A. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão. In: SIMPÓSIO DE PESQUISA E DESENVOLVIMENTO EM COMPUTAÇÃO, v. 2, n. 1, 2016
- [3] ESTATÍSTICA Mensal. Disponível em: <https://acobrasil.org.br/site/estatistica-mensal/>. Acesso em: 12 mar. 2024.
- [4] VIRGILITTO, S. B.; FAMA, R. A análise estatística multivariada na previsão de insolvência de empresas. **Revista Administração em Diálogo - RAD**, v. 4, n. 1, 2002.
- [5] FERREIRA, A. et al. Um estudo sobre previsão da demanda de encomendas utilizando uma rede neural artificial. **Blucher Marine Engineering Proceedings**, v. 2, n. 1, p. 353-364, 2016
- [6] GUIMARAES, P. L. **Processo de Previsão de demanda para empresa têxtil**. 2008. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2008.

- [7] MARKET Index. Disponível em: <https://www.marketindex.com.au/iron-ore>. Acesso em: 12 mar. 2024.
- [8] SOUZA, A. J. T. S. **Aplicação de técnicas de Machine Learning para a previsão de processos industriais**. 2023. Dissertação (Mestrado integrado em Engenharia e Gestão de Sistemas de Informação) – Universidade do Minho, 2023.
- [9] ALMEIDA, I. N. **A aplicação de tecnologias disruptivas na capacidade preditiva da cadeia de suprimentos**. 2020. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) – Centro de Tecnologia, Departamento de Engenharia de Produção e Sistemas, Universidade Federal de Santa Maria, 2020. Disponível em: <http://repositorio.ufsm.br/handle/1/21494>. Acesso em: 12 mar. 2024.
- [10] MIGON, H. Análise de séries temporais. Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, 2007. Apresentação em slide. Disponível em: <https://www.dme.ufrj.br/dani/pdf/slidespartefrequentista.pdf>. Acesso em: 12 mar. 2024.
- [11] CRISPIM, C. M. **Comparação entre os Modelos SARIMA e os Métodos de Machine Learning para previsão de demanda**. 2022. Trabalho de Conclusão de Curso (Graduação em Estatística) – Instituto de Matemática e Estatística, Universidade Federal Fluminense, 2022.
- [12] JUNIOR, W. C. et al. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuária Brasileira**, v. 51, n. 9, p. 1428-1437, 2016.
- [13] HOMEM, W. L. **Apostila de machine learning**. PET Engenharia Mecânica, UFES, 2020
- [14] SAIGAL, S.; MEHROTRA, D. Performance comparison of time series data using predictive data mining techniques. **Advances in Information Mining**, v. 4., n. 1, p. 57–66, 2012.
- [15] SAGAR, B.M.; VINUTHA, H. P.; POORNIMA, B. Detection of outliers using interquartile range technique from intrusion dataset. In: SATAPATHY, S. C.; TAVARES, J. M. R. S.; BHATEJA, V.; MOHANTY, J. R. (ed.). **Information and decision**

sciences: Proceedings of the 6th international conference on FICTA. Springer, 2018.
p. 511-518.

- [16] HYNDMAN, R.J. Better acf and pacf plots, but no optimal linear prediction. **Electronic Journal of Statistics**, v. 8., n. 1, p. 2296-2300, 2014.