



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA MECÂNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA MECÂNICA

MATEUS CELESTINO DA SILVA

**ESTUDO DE CASO SOBRE TENDÊNCIA DA PRODUÇÃO SIDERÚRGICA
BRASILEIRA A PARTIR DE UM MODELO PREDITIVO BASEADO EM RANDOM
FOREST REGRESSOR**

Recife

2025

MATEUS CELESTINO DA SILVA

**ESTUDO DE CASO SOBRE TENDÊNCIA DA PRODUÇÃO SIDERÚRGICA
BRASILEIRA A PARTIR DE UM MODELO PREDITIVO BASEADO EM RANDOM
FOREST REGRESSOR**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Mecânica da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do grau de Bacharel em Engenharia Mecânica.

Orientadora: Profa. Dra. MARCELE ELISA FONTANA.

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Silva, Mateus Celestino da.

Estudo de caso sobre tendência da produção siderúrgica brasileira a partir de um modelo preditivo baseado em Random Forest Regressor / Mateus Celestino da Silva. - Recife, 2025.

53 : il., tab.

Orientador(a): Marcele Elisa Fontana

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia Mecânica - Bacharelado, 2025.

Inclui referências, apêndices, anexos.

1. Indústria Siderúrgica. 2. Análise Preditiva. 3. Aprendizado de Máquina. 4. Random Forest. I. Fontana, Marcele Elisa. (Orientação). II. Título.

620 CDD (22.ed.)

MATEUS CELESTINO DA SILVA

**ESTUDO DE CASO SOBRE TENDÊNCIA DA PRODUÇÃO SIDERÚRGICA
BRASILEIRA A PARTIR DE UM MODELO PREDITIVO BASEADO EM RANDOM
FOREST REGRESSOR**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Mecânica da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do grau de Bacharel em Engenharia Mecânica.

Aprovado em: 12/08/2025.

BANCA EXAMINADORA

Profa. Dra. MARCELE ELISA FONTANA (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. MARCUS COSTA DE ARAUJO (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. JOSÉ LEÃO E SILVA FILHO (Examinador Externo)
Universidade Federal de Pernambuco

Dedico esse trabalho primeiramente ao Deus Trino, pela sabedoria que foi concedida durante toda jornada da graduação, como a minha mãe, avó, tios, minha companheira da vida e toda família que me deu suporte por toda trajetória.

AGRADECIMENTOS

Agradeço ao Senhor Pai, Filho e Espírito Santo pela graça concedida durante toda jornada da graduação, guiando meus passos apesar das limitações.

Agradeço minha mãe Andreia, minha avó Fátima, meus tios Flávio e Fábio, como aos meus familiares por todo apoio e suporte necessário nesta fase da minha vida, sem vocês eu não poderia ter essa oportunidade.

Agradeço a minha companheira Aline pela sua paciência e incentivo contínuo.

Agradeço a minha orientadora Profa. Dra. Marcele, pela sua paciência, atenção, apoio e contribuição neste trabalho.

Por fim, meus agradecimentos a todos que contribuíram para minha jornada acadêmica como meus colegas de curso, ao projeto de extensão Empresa Júnior EIXO Consultoria, professores, funcionários de apoio que realizam serviços no Centro de Tecnologia e Geociências (CTG), colegas da profissão que continuamente me ensinam.

Confie no Senhor de todo o seu coração e não se apoie em seu próprio entendimento. (BÍBLIA Sagrada, 2025, Pv 3, 5-6)

RESUMO

Este trabalho propõe um modelo preditivo baseado em técnica de inteligência artificial aplicado a indústria siderúrgica brasileira, capaz de estimar a produção de aço em seus segmentos mais vitais: os Aços Planos, os Longos e os Semi-Acabados. Para isso, primeiramente foram coletados dados do período de 2020 a 2024, disponibilizados em bases do Instituto Aço Brasil (IABr), o IBGE, a ANFAVEA, World Steel, Banco Central e FMI. A partir do tratamento destes dados, uma metodologia de Aprendizado de máquina utilizando o algoritmo *Random Forest*, para aprender as nuances e as relações sutis entre onze diferentes indicadores e a produção de cada segmento: Aço semiacabado, Aço longo e Aço plano. Os resultados quantitativos, com Coeficientes de Determinação (R^2) que alcançaram 91,50% para Aços Longos, 88,20% para Aços Planos e 75% para Aços Semiacabados. Além, dos resultados por meio da métrica MAE com valores de erro absoluto abaixo do limite de 100 mil toneladas. A pesquisa comprova que a Inteligência Artificial, quando bem aplicada, é um caminho poderoso para transformar a gestão reativa em uma gestão proativa e informada, pronta para os desafios do setor siderúrgico.

Palavras-chave: Indústria Siderúrgica; Análise Preditiva; Aprendizado de Máquina; *Random Forest*.

ABSTRACT

This work proposes a predictive model based on artificial intelligence technique applied to the Brazilian steel industry, capable of estimating steel production in its most vital segments: flat steel, long steel, and semi-finished steel. To achieve this, data from the period 2020 to 2024 were first collected from the databases of the Brazilian Steel Institute (IABr), IBGE, ANFAVEA, World Steel, the Central Bank, and the IMF. Based on this data processing, a machine learning methodology using the Random Forest algorithm was used to learn the nuances and subtle relationships between eleven different indicators and the production of each segment: semi-finished steel, long steel, and flat steel. The quantitative results, with Coefficients of Determination (R^2) reaching 91.50% for Long Steel, 88.20% for Flat Steel, and 75% for Semi-finished Steel. In addition, the results obtained using the MAE metric show absolute error values below the 100,000-ton limit. The research proves that Artificial Intelligence, when applied correctly, is a powerful way to transform reactive management into proactive and informed management, ready for the challenges of the steel industry.

Keywords: Steel Industry; Predictive Analytics; Machine Learning; Random Forest.

LISTA DE FIGURAS

Figura 1-Ilustração Random Forest.....	23
--	----

LISTA DE TABELAS

Tabela 1 – Produção Anual de Aço Plano.....	31
Tabela 2 – Produção Aço Plano vs. Produção de Veículos.....	32
Tabela 3 – Vendas Internas de Aços Planos por Setor Consumidor (2024).....	32
Tabela 4 – Produção Anual de Aço Longo.....	32
Tabela 5 – Destino das Vendas de Aço Longo.....	33
Tabela 6 – Produção de Aço Longo vs. PIB da Construção Civil.....	33
Tabela 7 – Produção Anual de Semiacabados.....	34
Tabela 8 – Vendas de Semiacabados.....	35
Tabela 9 – Principais Mercados Importadores do Aço Brasileiro.....	35
Tabela 10 – Produção de Semiacabados vs. PIB dos EUA.....	35
Tabela 11 – Parâmetros do modelo.....	36
Tabela 12 – Resultado das métricas de desempenho.....	37

LISTA DE ABREVIACOES

AED – Anlise Exploratria de Dados

ANFAVEA – Associao Nacional dos Fabricantes de Veculos Automotores

CBIC – Cmara Brasileira da Indstria da Construo

ETL – Extract, Transform, and Load (Extrao, Transformao e Carga)

FMI – Fundo Monetrio Internacional

IA – Inteligncia Artificial

IABr – Instituto Ao Brasil

IBGE – Instituto Brasileiro de Geografia e Estatstica

KPI – Key Performance Indicator (Indicador-Chave de Desempenho)

MAE – Mean Absolute Error (Erro Mdio Absoluto)

PIB – Produto Interno Bruto

R^2 – Coeficiente de Determinao

RMSE – Root Mean Squared Error (Raiz do Erro Quadrtico Mdio)

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	14
1.1.1	Objetivo Geral	14
1.1.2	Objetivos Específicos.....	14
1.2	JUSTIFICATIVAS.....	14
2	FUNDAMENTAÇÃO TEÓRICA.....	16
2.1	TIPOS DE DADOS.....	16
2.2	TÉCNICAS DE ANÁLISE DE DADOS	17
2.2.1	Análise Descritiva	17
2.2.2	Análise Prescritiva	18
2.2.3	Análise Preditiva	18
2.3	INTELIGÊNCIA ARTIFICIAL (IA).....	19
2.3.1	Aprendizado Máquina	19
2.3.1.1	Aprendizado supervisionado e regressão.....	21
2.3.1.2	Algoritmo <i>Random Forest</i>	21
2.4	SETOR SIDERÚRGICO BRASILEIRO	23
3	METODOLOGIA	25
3.1	CARACTERIZAÇÃO DA PESQUISA.....	25
3.2	ETAPAS DA PESQUISA.....	25
3.2.1	Coleta de dados.....	25
3.2.2	Modelagem e treinamento do modelo.....	26
3.2.2.1	Preparação do código em Python.....	27
3.2.2.2	Preparação para treinamento do modelo	28
3.2.2.3	Interface para o usuário.....	29
3.2.3	Avaliação de performance	29

3.2.4	Simulação de cenários.....	30
4	RESULTADOS E DISCURSSÃO.....	31
4.1	ANALISE DESCRITIVA DOS SEGMENTOS DO AÇO.....	31
4.1.1	Aço plano.....	31
4.1.2	Aço longos.....	32
4.1.3	Aço semiacabado.....	34
4.2	VARIÁVEIS DO MODELO PREDITIVO.....	36
4.3	AVALIAÇÃO DE PERFORMANCE.....	36
4.4	RESULTADOS DA SIMULAÇÃO DE CENÁRIOS (ESTUDO DE CASO)...	37
5	CONCLUSÃO.....	39
5.1	LIMITAÇÕES DO TRABALHO.....	39
5.2	RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	40
	APÊNDICE A – Consulta SQL.....	45
	APÊNDICE B – Programação do modelo preditivo.....	48

1 INTRODUÇÃO

O setor siderúrgico brasileiro possui uma capacidade instalada de aproximadamente 51 milhões de toneladas anuais e empregando diretamente mais de 120 mil colaboradores, a indústria do aço posiciona o Brasil como o 9º maior produtor mundial (Instituto Aço Brasil, 2024; World Steel Association, 2025). Sendo pilar da nossa indústria, cujo seu desempenho reflete e influencia diretamente a saúde da economia do país, abastecendo áreas vitais como a construção civil, o setor automobilístico e a produção de bens de capital, (Instituto Aço Brasil, 2024). Dentro deste cenário é observado a complexidade mercadológica do setor, um dos fatores é proveniente do desafio da competitividade frente aos produtos importados e agenda de descarbonização global (Instituto Aço Brasil, 2024).

Neste contexto, se insere o presente trabalho, buscando desenvolver um conjunto de modelos preditivos (modelos independentes, para cada segmento do aço: longo, plano e semiacabado) baseado na técnica *Random Forest Regressor*, onde seja capaz de avaliar tendências da produção siderúrgica no Brasil, proporcionando insights baseados em dados para auxiliar na tomada de decisões estratégicas e operacionais, sendo suporte para iniciativas voltadas a otimização no planejamento e controle da produção e alocação estratégica de recursos.

A era atual é marcada por uma transformação digital, movida pela força da Inteligência Artificial (IA). Essa revolução está redesenhando o dia a dia das organizações, que agora contam com a IA para desvendar tarefas complexas e transformar o imenso volume de dados gerado diariamente em processos mais enxutos e inteligentes (Iansiti; Lakhani, 2020).

Essa nova realidade está mudando a própria natureza da tomada de decisão. A intuição e a experiência, que por tanto tempo guiaram os gestores, embora ainda valiosas, abrem espaço para uma cultura orientada por dados. A pergunta fundamental dentro das empresas está se deslocando da incerteza baseada dos achismos do que pode acontecer, para um cenário mais consolidado e orientado por dados, do que provavelmente irá acontecer (Provost; Fawcett, 2013). A competitividade, hoje, está diretamente ligada à eficiência com que uma organização consegue traduzir a gama de dados à sua disposição em ações estratégicas.

O estudo visa a aplicabilidade de um modelo preditivo por meio do algoritmo *Random Forest Regressor*, para isso em consequência do estudo do negócio, serão analisados dados disponibilizados, buscando compreender a correlação e influência de diversos indicadores que serão analisados, sobre a tendência produtiva da siderúrgica brasileira.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Desenvolver um conjunto de modelos preditivos (modelos independentes, para cada segmento do aço: longo, plano e semiacabado) baseado no algoritmo *Random Forest Regressor* capaz de avaliar tendências da produção siderúrgica no Brasil, proporcionando insights baseados em dados para auxiliar na tomada de decisões estratégicas e operacionais.

1.1.2 Objetivos Específicos

- Realizar uma análise descritiva para identificar quais variáveis econômicas e de mercado, ao longo do tempo, estudando a correlação com o ritmo de produção no contexto em estudo;
- Construir e validar um conjunto de modelos preditivos (modelos independentes, para cada segmento do aço: longo, plano e semiacabado) por meio da Aprendizagem de Máquina, utilizando o algoritmo *Random Forest*;
- Realizar a segregação do conjunto de dados para as fases de treinamento e teste;
Executar estudo de caso por meio de simulação de cenários, extraindo conclusões claras e insights práticos.

1.2 JUSTIFICATIVAS

Este trabalho se justifica por buscar oferecer o desenvolver de uma ferramenta analítica para suporte a gestores, frente a cenários incertos que promovem uma dificuldade e erro de previsões, ao desenvolver uma forma mais precisa de antecipar a produção, tem-se recursos voltadas a empresas do setor para que possam otimizar seus processos, evitar desperdícios e fortalecer sua resiliência, tornando-se mais competitivas em um cenário mundial desafiador (Sharda; Delen; Turban, 2019).

Por fim, temos o aspecto da necessidade mercadológica, em uma economia não linear o *Random Forest*, é capaz de capturar as relações complexas e não-lineares entre um número de variáveis preditoras por ser um método de ensemble baseado em árvores de decisão, que confere ao modelo uma robustez a ruídos e *outliers* presentes nos dados históricos, resultando em previsões mais estáveis e confiáveis (Hastie; Tibshirani; Friedman, 2017).

O presente estudo, contribui diretamente para o ODS 9 (Indústria, Inovação e Infraestrutura), ao promover uma inovação tecnológica ao setor. Além disso, dialoga com o ODS 12 (Consumo e Produção Responsáveis), pois a melhor noção da tendência de produção, baseada em predições mais precisas, combate a superprodução, reduzindo o desperdício de recursos naturais e de energia, tornando o processo produtivo mais sustentável.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 TIPOS DE DADOS

Os dados se manifestam em múltiplas formas. A sua identificação e tratamento são cruciais, pois diferentes tipos de dados exigem diferentes abordagens analíticas e de armazenamento.

A forma mais tradicional e de mais fácil manipulação são os dados estruturados, caracterizados por sua organização em um formato tabular e bem definido, como em um banco de dados relacional (Elmasri; Navathe, 2016). Dentro dessa estrutura, encontramos tipos primários:

- **Dados Numéricos:** Representam quantidades mensuráveis e são a base dos modelos preditivos. Dividem-se em inteiros (usados para contagens, como o número de peças produzidas) e reais (ou de ponto flutuante, usados para medições precisas, como o preço de uma commodity em dólares) (Korth et al., 2010);
- **Dados Categóricos:** Descrevem qualidades ou características. Podem ser nominais, quando não há uma ordem inerente entre as categorias (e.g., nome de um país, tipo de aço), ou ordinais, quando existe uma hierarquia ou ordem clara (e.g., classificações de qualidade como 'A', 'B', 'C') (Agresti, 2013).

Em contrapartida aos dados estruturados, a maior parte dos dados gerados hoje no mundo é não estruturada, ou seja, não possui um esquema predefinido, o que torna sua análise computacional um desafio significativo (Manning et al., 2008). Nisto se incluem:

- **Dados Textuais:** Sequências de caracteres, como e-mails, documentos, notícias ou postagens em redes sociais. Sua análise requer técnicas de Processamento de Linguagem Natural (PLN);
- **Dados Binários:** Representações de mídias complexas, como imagens, áudios e vídeos;
- **Dado geoespacial,** que representa informações de localização (coordenadas, polígonos).

No presente trabalho, iremos tratar com dados estruturados do tipo numérico, pois todo modelo será baseado em indicadores econômicos e operacionais/comercial. Com isso, temos uma redução na complexidade no tratamento da base de dados e estrutura do algoritmo aplicado que será responsável pela análise e geração das previsões.

2.2 TÉCNICAS DE ANÁLISE DE DADOS

A análise de dados constitui-se como um processo sistemático de inspeção, limpeza, transformação e modelagem de dados com o objetivo primordial de descobrir informações úteis, embasar conclusões e apoiar a tomada de decisão. Para isso, temos a aplicação de técnicas de análise, segregadas em níveis, que podem ser entendidas como um acervo de estratégias que melhor se enquadram para o objetivo da análise (Iansiti; Lakhani, 2020). Existem três níveis de técnicas para análise de dados, sendo a análise descritiva, preditiva e prescritiva (Sharda; Delen; Turban, 2019).

2.2.1 Análise Descritiva

A análise descritiva, tem por objetivo descrever cenários e padrões dos dados consiste na coleta, organização e sumarização de dados históricos para revelar padrões, anomalias e tendências passadas. Segundo Sharda, Delen e Turban (2019), sua missão não é prever o futuro ou especular sobre possibilidades, mas sim contar a história do que aconteceu de forma clara, precisa e honesta. É o ato de olhar pelo retrovisor para entender o caminho percorrido até o momento. Sem esse olhar para trás, qualquer passo à frente seria um mero palpite.

As ferramentas para essa exploração, como aponta Few (2006), são variadas e incluem o uso de estatísticas descritivas fundamentais (médias, medianas, desvios padrão), a geração de relatórios tabulares que agregam e consolidam grandes volumes de informação, e o emprego de painéis visuais (dashboards) que monitoram em tempo real os principais indicadores de desempenho (KPIs).

2.2.2 Análise Prescritiva

Análise prescritiva permite a execução de simulações e otimizações para recomendar o melhor curso de ação dentre um universo de alternativas, visando alcançar um resultado ótimo ou mitigar riscos futuros.

Como postulado por Davenport & Harris (2007), a análise prescritiva funciona como um "GPS" para a tomada de decisão. Não apenas se limitando a indicar o destino (a previsão), mas também sugere a melhor rota para alcance dos resultados, avaliando o impacto potencial de cada decisão. Por exemplo, diante de uma previsão de queda na demanda, um modelo prescritivo poderia simular diferentes cenários de redução de produção para recomendar aquele que minimiza o impacto nos custos e maximiza as margens.

2.2.3 Análise Preditiva

A análise preditiva se configura como uma projeção do futuro, por meio de modelos estatísticos, algoritmos de aprendizado de máquina e técnicas de mineração de dados, com o objetivo de identificar padrões sutis em dados históricos para projetar a probabilidade de eventos futuros, ou seja, o objetivo não é prever o futuro com certeza absoluta, mas sim estimar a probabilidade de diferentes resultados. Oferecendo previsões acuradas quantificando a incerteza, permitindo que a organização se antecipe a tendências, otimize a alocação de recursos e planeje estratégias de forma proativa (Iansiti; Lakhani, 2020).

É neste domínio, que o presente trabalho se insere fundamentalmente, ao aplicar tal técnica para prever a produção do setor siderúrgico brasileiro.

2.3 INTELIGÊNCIA ARTIFICIAL (IA)

A Inteligência Artificial (IA), em sua concepção moderna, é a disciplina focada em desenvolver sistemas computacionais dotados de capacidades cognitivas, permitindo que ajam e raciocinem de forma conclusiva (Russell; Norvig, 2021). Essa capacidade fundamental, antes confinada a projetos de pesquisa, transcendeu o ambiente acadêmico e hoje representa o núcleo de novos e disruptivos modelos operacionais. Como demonstram Iansiti e Lakhani (2020), a IA não é mais apenas uma ferramenta para resolver problemas pontuais, mas sim o próprio “motor” que impulsiona as empresas mais competitivas da atualidade.

É justamente essa crescente centralidade e poder da IA que impulsiona um novo e indispensável paradigma: o da IA Centrada no Humano. Conforme postula Shneiderman (2022), o avanço da área exige um deslocamento de foco, priorizando a criação de sistemas que sejam, acima de tudo, confiáveis, seguros e compreensíveis. A meta, segundo essa visão, é projetar ferramentas que atuem como parceiras competentes para a sociedade, seja para profissionais, cientistas ou cidadãos comuns. Isso implica garantir que a automação permaneça sob um controle humano significativo, assegurando que a tecnologia sirva para ampliar a criatividade e a responsabilidade, em vez de diminuí-las. Essa perspectiva não apenas define a fronteira atual da pesquisa, mas também molda a governança em IA, na busca por um futuro de harmonia entre o avanço tecnológico e o progresso humano.

2.3.1 Aprendizado Máquina

Aprendizado de Máquina, na perspectiva de Bishop (2011), debruça-se sobre um desafio central de como sistemas computacionais podem aprender a reconhecer padrões a partir de um volume finito de exemplos. Segundo Géron (2021), em vez de codificar regras manualmente, o desenvolvedor alimenta um algoritmo com uma grande quantidade de dados, permitindo que o próprio sistema “aprenda” as regras e os padrões que governam o problema. A essência do aprendizado de máquina, portanto, reside na sua capacidade de generalizar o conhecimento adquirido a partir de exemplos para, então, fazer previsões ou tomar decisões sobre dados novos e nunca vistos.

Nesse percurso, uma das abordagens mais sólidas e unificadoras é a que trata o aprendizado sob a ótica da probabilidade. Este ferramental matemático torna-se indispensável porque a incerteza é um elemento intrínseco a qualquer problema do mundo real, seja pelo ruído natural dos dados ou pela simples limitação de se ter apenas uma amostra da realidade. Ignorar essa dimensão da incerteza, como alerta Bishop (2011), pode levar à criação de modelos excessivamente confiantes, cujas decisões se tornam frágeis e propensas a erros graves na prática.

A grande vantagem da teoria probabilística é que ela permite aos modelos comunicarem suas predições de uma forma muito mais rica e honesta. Em vez de um valor único e determinístico, o resultado é expresso como uma distribuição de probabilidades. O treinamento consiste, então, em um processo de otimização iterativo, no qual um algoritmo ajusta os parâmetros internos do modelo para reduzir o erro ao mínimo possível (Géron, 2021, p. 32).

Conforme Géron (2021), as principais abordagens dentro do campo do aprendizado de máquina, são:

- **Aprendizado Não Supervisionado:** O sistema recebe dados não rotulados e deve encontrar estruturas ocultas por conta própria, como em tarefas de clusterização (agrupamento de dados similares);
- **Aprendizado por Reforço:** Um "agente" aprende a se comportar em um ambiente por meio de um sistema de recompensas e punições, descobrindo a melhor sequência de ações por tentativa e erro;
- **Aprendizado Supervisionado:** O modelo é alimentado com dados rotulados, onde a "resposta correta" é fornecida para cada exemplo. É o paradigma mais comum, utilizado em tarefas de regressão (prever um valor numérico) e classificação (prever uma categoria).

O presente trabalho se enquadra dentro da aprendizagem de máquina do tipo supervisionado e regressivo.

2.3.1.1 Aprendizado supervisionado e regressão

O processo é "supervisionado", conforme delineado por Géron (2021), quando a existência de um conjunto de dados de treinamento que contém não apenas as variáveis de entrada (ou preditoras), mas também suas correspondentes variáveis de saída (ou alvos), que funcionam como uma espécie de "gabarito" ou guia. O objetivo do modelo é aprender a função matemática f que melhor mapeia a relação entre entrada e saída, conforme a Eq (1):

$$Y = f(X) + \varepsilon \quad (1)$$

Onde Y é a variável que queremos prever, X é o vetor de variáveis preditoras, f é a função de mapeamento que o modelo tenta estimar, e ε é o erro aleatório irreduzível. O problema é de Regressão quando Y é uma variável contínua (JAMES et al., 2013).

2.3.1.2 Algoritmo *Random Forest*

O algoritmo *Random Forest*, proposto por Breiman (2001), opera construindo múltiplas árvores de decisão e agregando seus resultados. Para o estudo em questão, que se trata de uma tarefa de regressão, a previsão final para um problema de regressão é a média das previsões de todas as árvores individuais, o que confere ao modelo alta precisão. Segundo Géron (2021) a solução de Breiman foi compreender que, uma única árvore é instável, ela cria regras tão específicas para os exemplos que já viu que se torna incapaz de generalizar seu conhecimento para novos dados, mas um conjunto de árvores, poderia, sob as condições certas, formar um sistema de previsão extremamente estável e preciso.

Em termos práticos, o *Random Forest* utiliza a estratégia de múltiplos "tomadores de decisões", cada tomador é representado por uma árvore como ilustra a Figura 1, o algoritmo recebe um valor chamado $n_estimators$ que define o número de árvores que serão criadas e conseqüentemente decisões tomadas. Sua robustez advém de duas técnicas principais:

- Técnica *Bootstrap Aggregating (Bagging)*: Cada árvore é treinada em uma subamostra aleatória (com reposição) do conjunto de dados original. Isso garante que as árvores sejam diferentes umas das outras;
- Técnica *Feature Randomness*: Em cada nó da árvore, apenas um subconjunto aleatório de variáveis é considerado para a divisão. Isso descorrelaciona as árvores e reduz a variância do modelo final.

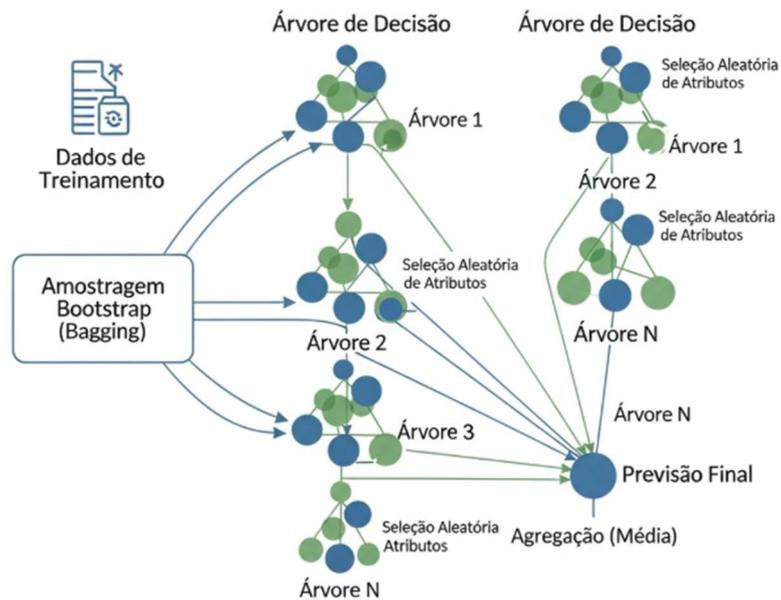


Figura 1-Ilustração *Random Forest*.

Fonte: Autor

Depois de construir a floresta com centenas de árvores únicas e descorrelacionadas, o processo para fazer uma previsão para um novo dado é:

- O novo dado é passado para cada árvore da floresta;
- Cada árvore, com base em sua "experiência" e "especialidade", produz sua própria previsão numérica. Por exemplo, Árvore 1 pode prever "x", a Árvore 2 pode prever "y", a Árvore 3 pode prever "z", e assim por diante;
- A previsão final do *Random Forest Regressor* (algoritmo *Random Forest*, quando aplicado ao aprendizado supervisionado do tipo regressão) é a média aritmética de todas as previsões individuais.

2.4 SETOR SIDERÚRGICO BRASILEIRO

A indústria siderúrgica é, por definição, o ramo da metalurgia focado na produção de aço e ferro fundido, tratando-se de uma indústria de base, cuja produção serve para praticamente toda a cadeia industrial moderna. Sua função é a transformação de matérias-primas de baixo valor agregado, notadamente minério de ferro, carvão e sucata reciclada, em aço (INSTITUTO AÇO BRASIL, 2024).

No Brasil, a produção efetiva de aço bruto flutua, conforme o ciclo econômico, na faixa de 32 a 34 milhões de toneladas por ano, com o consumo aparente, que mede a demanda do mercado interno, situando-se entre 24 e 26 milhões de toneladas. (Instituto aço brasil, 2024). Para fins de análise econômica e de planejamento, contudo, é um erro tratar a siderurgia como um bloco monolítico. A dinâmica do setor só pode ser compreendida a partir da segmentação de seus produtos, pois cada um deles atende a mercados distintos, responde a diferentes vetores de demanda e, conseqüentemente, possui seu próprio ciclo de negócios. A segmentação do aço possui uma divisão principal em três grandes categorias:

- Aços Planos: Apresentados como chapas e bobinas, são a matéria-prima da indústria de transformação, seu principal destino é o setor automotivo, seguido pela linha branca (eletrodomésticos) e embalagens. Sua performance é diretamente ditada pela confiança do consumidor e pelo ritmo da produção industrial (Instituto aço brasil, 2024).
- Aços Longos: Incluem vergalhões, perfis estruturais e fio-máquina, segmento considerado a base da infraestrutura. Seu mercado é a construção civil, sendo utilizados em vergalhões para edifícios e grandes obras. Sua demanda é fortemente influenciada pelo ciclo de investimentos e pela política monetária, que afeta o crédito imobiliário (Instituto aço brasil, 2024).
- Aços Semiacabados: São a "commodity" de exportação do setor. Placas e tarugos são vendidos para outras siderúrgicas ao redor do mundo, principalmente nos Estados Unidos e nas Américas. (Instituto aço brasil, 2024).

O principal desafio tem sido a concorrência do aço importado, que pressiona as margens das usinas nacionais (Instituto aço brasil, 2024). Em contrapartida, a agenda global da descarbonização se apresenta como uma oportunidade única para o Brasil se destacar na produção de "aço verde", dada sua matriz energética mais limpa.

3 METODOLOGIA

Este capítulo dedica-se a delinear o caminho metodológico percorrido para alcançar os objetivos desta pesquisa.

3.1 CARACTERIZAÇÃO DA PESQUISA

O trabalho se caracteriza como uma pesquisa aplicada, pois o foco é a busca por uma solução para um desafio real e concreto: a necessidade de prever com maior acurácia a produção do setor siderúrgico. Seus objetivos se desdobram em duas etapas que se complementam, a pesquisa exploratória, investigando variáveis para descobrir quais delas realmente importam para solução do desafio, como também a pesquisa explicativa, focando em modelar e entender como essas variáveis se relacionam para influenciar o cenário da pesquisa. Os procedimentos de estudo combinaram duas frentes: a pesquisa documental, essencial para garimpar os dados históricos em fontes setoriais e econômicas, e a modelagem computacional, onde os algoritmos de aprendizado de máquina foram usados para construir e testar o modelo preditivo.

3.2 ETAPAS DA PESQUISA

3.2.1 Coleta de dados

Inicialmente foi necessária uma análise de negócios visando estruturar informações, requisitos e entendimento do negócio. Sendo possível ter uma melhor compreensão dos fatores que estão correlacionados a dinâmica do objeto de estudo (produção brasileira de aço longo, plano e semiacabada) e aplicação da análise preditiva. Esta etapa inicial, forneceu informações relevantes para concentrar as pesquisas e obtenção dos dados relevantes para consolidação da base dados que será usada para treinamento, validação e teste do modelo preditivo.

Uma base histórica foi criada no que diz respeito de 2020 a 2024, unindo informações de fontes como IABr, o IBGE, a ANFAVEA, World Steel, Banco Central do Brasil e FMI.

- Produções dos aços planos, longos e semiacabados, como também os dados do consumo aparente do aço no Brasil, importação e exportação do aço: Foram extraídos os dados da base de dados pública do Instituto Aço Brasil (2020, 2021, 2022, 2023, 2024);
- Produção de veículos da indústria automotiva brasileira: Foram extraídos os dados da base de dados pública da ANFAVEA (2025);
- PIB do setor de construção civil: Foram extraídos os dados da base de dados pública do IBGE (2025);
- Variação da taxa Selic: Foram extraídos os dados da base de dados pública do Banco Central do Brasil (2025);
- PIB das economias da América do Sul, Estados Unidos, Canadá e México: Foram extraídos os dados da base de dados pública do Fundo Monetário Internacional (2025);
- Consumo aparente do aço na América do Norte: Foram extraídos os dados da base de dados pública do World Steel Association (2020, 2021, 2022, 2023, 2024).

Os dados inicialmente foram tratados por meio do software Excel, a fim de organizar as informações presentes nos relatórios extraídos, que inicialmente precisavam de ajustes manuais, por conta dos diferentes formatos de exibição das informações. Posteriormente, para refino do tratamento inicial feito no software Excel, por meio da linguagem SQL, detalhada no Apêndice A, foram realizadas consultas para melhor tratamento e correlação dos dados extraídos das diferentes bases, totalizando 1460 linhas de informações, distribuídas diariamente.

3.2.2 Modelagem e treinamento do modelo

Com a fundação de dados estabelecida, será feita a construção do modelo preditivo, sendo construído três modelos independentes para cada segmentação do aço. A escolha metodológica recaiu sobre o Aprendizado de Máquina Supervisionado, especificamente para um problema de regressão, uma vez que o objetivo era prever uma variável numérica contínua (a produção em toneladas).

O algoritmo selecionado foi o Random Forest Regressor, no qual oferece uma flexibilidade para capturar as relações complexas e não-lineares. Sua arquitetura, baseada em um conjunto de múltiplas árvores de decisão, confere-lhe uma robustez natural a ruídos e outliers. Tal algoritmo foi importados e utilizado em uma função criada no caderno Jupyter utilizando Python.

O próximo passo se trata do treinamento do modelo. Segundo Hastie, Tibshirani e Friedman (2017), para garantir uma avaliação robusta do modelo e evitar o super ajuste, é adotado a metodologia de segregação da base de dados em dois conjuntos independentes: treinamento e teste. Esta prática, fundamental para a obtenção de uma estimativa imparcial do erro de generalização. Foram, portanto, treinados três modelos distintos e especializados, um para cada tipo de ação, optou-se uma repartição dos dados usando 80% dos dados históricos para o aprendizado e reservando 20% para teste dos modelos. Conforme a metodologia de segregação de dados para garantir uma avaliação imparcial da performance, 1460 instâncias foram divididas da seguinte forma:

- Conjunto de Treinamento (80%): 1168 instâncias foram utilizadas para o processo de aprendizado dos modelos;
- Conjunto de Teste (20%): 292 instâncias foram utilizadas para a avaliação final da performance do modelo, gerando as métricas de MAE e R^2 .

As seções seguintes irão contemplar os passos do desenvolvimento do modelo, detalhado no Apêndice B, foram usadas as referências Géron (2021) e (Mckinney, 2018) para orientação no desenvolvimento do modelo.

3.2.2.1 Preparação do código em Python

Neste ponto será necessário “preparar” a linguagem Python, importando bibliotecas que irão viabilizar funcionalidades essenciais para o desenvolvimento do modelo, as importações realizadas foram:

- Biblioteca *Pandas*: Para trabalhar com tabelas de dados;
- Biblioteca *Numpy*: Para trabalhar com funções matemáticas;
- Biblioteca *RandomForestRegressor*: Por meio da biblioteca *Scikit-learn*, o algoritmo *Random Forest Regressor* é importado.

3.2.2.2 Preparação para treinamento do modelo

O objetivo neste momento, por meio da documentação da biblioteca *scikit-learn*, será imputar na linguagem Python as definições para treinamento dos 3 modelos (modelo para cada segmentação do Aço), sendo definido os hiper parâmetros, como:

- `n_estimators = 100`, o número de árvores de decisão que serão construídas na floresta;
- `random_state = 42`, o gerador de escolhas aleatórios de variáveis do algoritmo;
- `criterion = 'squared_error'`, MAE, critérios de avaliação do modelo;
- `max_depth = None`, A profundidade máxima que cada árvore pode atingir. `None` significa que as árvores podem crescer até que todas as folhas sejam "puras";
- `min_samples_split=2`, o número mínimo de amostras que um nó precisa ter para poder ser dividido;
- `min_samples_leaf =1`, o número mínimo de amostras que um nó final deve ter;
- `max_features= 1.0`, a proporção de variáveis a serem consideradas em cada divisão;

Em um problema de regressão, a preocupação é garantir que a distribuição estatística da variável-alvo seja representada de forma similar nos conjuntos de treinamento e teste. A função `train_test_split` da biblioteca *Scikit-learn*, como mostra Apêndice B, realiza uma amostragem aleatória dos dados para criar os conjuntos, ela garante que a distribuição dos valores de produção (a média, o desvio padrão, os valores mínimos e máximos) seja estatisticamente semelhante entre os dois conjuntos.

3.2.2.3 Interface para o usuário

Por fim, é definido o local onde os valores de entradas informados pelo usuário, serão processados para gerar as simulações de cenários com base no modelo anteriormente treinado:

- Variável *cenário_para_prever*: Variável responsável por armazenar os valores de entrada informado pelo usuário;
- Variável *previsao*: Variável responsável por armazenar os resultados das previsões geradas pelo modelo para os diferentes segmentos do aço.

3.2.3 Avaliação de performance

Um modelo só é útil se sua precisão for mensurável e confiável. Para aferir o sucesso dos modelos treinados, foi aplicado um conjunto de métricas padrões, as métricas escolhidas foram:

- Erro Médio Absoluto (MAE), que nos informa o erro médio da previsão na própria unidade de medida (toneladas), conforme Eq (2);

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_{2i}| \quad (2)$$

- Coeficiente de Determinação (R^2), que revela a porcentagem da variação da produção que o nosso modelo consegue explicar. Será considerado R^2 igual ou acima de 70% como indicador de sucesso do modelo, conforme Eq (3).

$$R^2 = 1 - \frac{SS_{RES}}{SS_{tot}} \quad (3)$$

Com base nas literaturas, como James et al. (2013), pode-se afirmar que não há um número ideal fixo, que defina o resultado da avaliação como aceitável, mas sim uma interpretação segundo análise do contexto do problema, ou seja, a métrica de sucesso da avaliação pode ser definida de acordo com os padrões do negócio, que o modelo está sendo aplicado.

Para o contexto, será considerado o Erro Médio Absoluto (MAE) na faixa limite de 100 mil toneladas (valor coerente, segundo IABr (2024), o setor trabalha com unidade de produção em milhões de toneladas). Em relação a métrica do Coeficiente de Determinação (R^2), será considerado aceitável valores iguais ou superiores a 70%, isso significa que nosso modelo explica minimamente 70% das possíveis variações que o resultado pode ter. Esses avaliadores de performance serão aplicados para cada segmentação do aço.

3.2.4 Simulação de cenários

A etapa final da metodologia consistiu na validação puramente estatística e testar a inteligência prática do modelo. Para isso, foi empregada a técnica de simulação de cenários, cujo objetivo é analisar se o modelo se comportava de maneira lógica e coerente com as dinâmicas de negócio já conhecidas e, ao mesmo tempo, demonstrar sua utilidade como uma ferramenta de apoio à decisão estratégica.

Será aplicado três cenários possíveis, que por muitas vezes são recorrentes dentro da dinâmica do mercado siderúrgico:

- Cenário 1, aumento do consumo com maior participação de importados: Neste cenário foi considerado aumento de 5% no Consumo Aparente, mas com a participação dos aços importados crescendo. Tal cenário irá nos revelar o impacto da maior participação de aço importado;
- Cenário 2, taxa Selic de estímulo (9%): Será analisado como a variação da taxa de juros, influencia no setor;
- Cenário 3, diminuição 10% em relação 2024, na participação do aço brasileiro, no mercado norte-americano: Será analisado o impacto da perda de espaço no principal cliente do aço brasileiro, IABr (2024).

4 RESULTADOS E DISCURSSÃO

Este capítulo se dedica à apresentação e à análise crítica dos resultados obtidos a partir da metodologia descrita. A discussão parte da avaliação quantitativa da performance dos modelos preditivos desenvolvidos e, a análise qualitativa dos insights gerados através da simulação de cenários, culminando em uma reflexão sobre as implicações práticas deste estudo

4.1 ANALISE DESCRITIVA DOS SEGMENTOS DO AÇO

A seção tem por objetivo analisar o comportamento histórico e as características individuais de cada um dos três principais segmentos do aço: os Planos, os Longos e os Semiacabados.

4.1.1 Aço plano

Tabela 1- Produção Anual de Aço Plano

Ano	Produção de Aço Plano	Varição Anual
2020	12,36	(Base)
2021	15,36	+24,3%
2022	9,8	-36,20%
2023	12,6	+28,6%
2024	13,58	+7,8%

Fonte: IABr (2024).

A Tabela 1, mostra desempenho do aço plano em milhões de toneladas, este está diretamente atrelado ao acesso ao crédito e à vitalidade de setores-chave, como o automotivo, conforme Tabela 2. O crescimento de +7,8% em 2024, chegando a 13,58 milhões de toneladas recuperação foi impulsionado pela contínua força do setor automotivo e pela indústria de transformação em geral, como mostra a Tabela 3 (ANFAVEA, 2025).

Tabela 2- Produção Aço Plano x Produção de Veículos

Ano	Produção de Aço Plano (mi/t)	Produção de Veículos (mi/unidades)
2020	12,36	2,01
2021	15,36	2,25
2022	9,8	2,37
2023	12,6	2,32
2024	13,58	2,45

Fonte: IABr (2024) e ANFAVEA (2025).

Tabela 3- Vendas Internas de Aços Planos por Setor Consumidor em 2024

Posição	Setor Consumidor	Vendas (mil toneladas)	Participação no Total (%)
1º	Distribuição	4.718	35,00%
2º	Indústria Automotiva	2.852	21,20%
3º	Tubos	1.765	13,10%
4º	Construção Civil	1.358	10,10%
5º	Bens de Capital (Máq. e Equip.)	1.304	9,70%
6º	Linha Branca (Eletrodomésticos)	951	7,10%
7º	Embalagens	424	3,10%

Fonte: IABr (2024).

4.1.2 Aço longos

O Aço Longo (vergalhões, perfis, fio-máquina) é o principal insumo da construção civil e de projetos de infraestrutura. Sua performance é, portanto, um reflexo direto da saúde e do nível de investimento nestes setores (Instituto aço brasil, 2024).

Tabela 4- Produção Anual de Aço Longo

Ano	Produção de Aço Longo (mi/t)	Varição Anual
2020	9,45	(Base)
2021	10,61	+12,3%
2022	9,8	-7,60%
2023	9,25	-5,60%%
2024	10,11	+9,4%

Fonte: IABr (2024)

Conforme Tabela 4, o auge em 2021 se firma como o pico de produção do período, com 10,61 milhões de toneladas. Este crescimento de +12,3% está diretamente ligado ao auge do ciclo de juros baixos no Brasil, que impulsionou o mercado imobiliário e as reformas, gerando uma forte demanda por vergalhões e outros produtos longos (Banco Central do Brasil, 2021).

Em 2022 já foi um ano de retração (-7,6%), que se acentuou em 2023 (-5,6%). Este biênio representa claramente o impacto do ciclo de alta da taxa Selic, que encareceu o crédito imobiliário e colocou um freio nos novos lançamentos da construção civil (CBIC, 2024).

Tabela 5- Destino das Vendas de Aço Longo

Ano	Vendas Internas (mi/t)	Vendas Externas (mi/t)	Total de Vendas (mi/t)	% para Mercado Interno
2020	8,31	1,2	9,45	88,00%
2021	9,32	1,05	10,61	87,80%
2022	8,49	1,42	9,8	86,60%
2023	8,07	1,19	9,25	87,20%
2024	8,61	1,19	10,11	85,20%

Fonte: IABr (2024).

Tabela 6- Produção de Aço Longo vs. PIB da Construção Civil

Ano	Produção de Aço Longo (mi/t)	PIB da Construção Civil (Variação %)
2020	9,45	-6,30%
2021	10,61	+9,7%
2022	9,8	+6,9%
2023	9,25	+0,5%
2024	10,11	(Dado não consolidado)

Fonte: IABr (2024) e IBGE (2025).

Os dados são conclusivos, ao longo de todo o período, as vendas para o mercado interno representaram, como mostra a Tabela 5, cerca de 85% a 90% do total. Isso prova que a saúde financeira e produtiva das siderúrgicas de aços longos está diretamente atrelada ao desempenho da construção civil brasileira. Segundo a Tabela 6, o pico de crescimento do PIB da Construção em 2021 (+9,7%) coincide perfeitamente com o pico de produção de aço longo. A forte desaceleração do crescimento do setor em 2023 (+0,5%) também se reflete na menor produção de aço do período, isso indica a forte ligação entre os dois indicadores.

4.1.3 Aço semiacabado

A trajetória da produção de semiacabados são notavelmente diferentes da dos aços acabados. Os dados são categóricos, como mostra Tabela 7, em todos os anos mais de 96% da produção de semiacabados para venda foi destinada ao mercado externo. O consumo interno é praticamente residual. Isso confirma que este segmento opera em uma lógica completamente diferente dos aços planos e longos. Sua produção não é ditada pela construção civil ou pela indústria automotiva brasileira, mas sim pela demanda de outras siderúrgicas e centros de laminação ao redor do mundo.

Tabela 7- Produção Anual de Semiacabados

Ano	Produção de Semiacabados (mi/t)	Varição Anual
2020	7,85	(Base)
2021	8,66	+10,3%
2022	8,08	-6,70%
2023	9,64	+19,3%
2024	9,25	-4,00%

Fonte: IABr (2024).

Tabela 8- Vendas de Semiacabados

Ano	Vendas Internas (mit/t)	Vendas Externas mit/t)	Total de Vendas mit/t)	% para Mercado Externo
2020	0,25	7,6	7,85	96,80%
2021	0,3	8,36	8,66	96,50%
2022	0,19	7,89	8,08	97,60%
2023	0,16	9,48	9,64	98,30%
2024	0,15	9,1	9,25	98,40%

Fonte: IABr (2024).

Tabela 9- Principais Mercados Importadores do Aço Brasileiro

Posição	País / Bloco	Participação no Total (%)	Principais Produtos Comprados
1º	Estados Unidos	~ 45-50%	Semi-acabados (Placas)
2º	Canadá e México	~ 15-20%	Semi-acabados, Produtos Planos
3º	América do Sul	~ 20-25%	Produtos Acabados (Planos e Longos)
4º	Outros (Europa, Ásia)	~ 5-10%	Produtos de Nicho

Fonte: IABr (2024).

Tabela 10- Produção de Semiacabados vs. PIB dos EUA

Ano	Produção de Semi-Acabados (mi/t)	PIB dos EUA (Variação %)
2020	7,85	-2,80%
2021	8,66	+5,9%
2022	8,08	+2,1%
2023	9,64	+2,5%
2024	9,25	+2,5% (Estimativa)

Fonte: IABr (2024) e FMI (2025).

Analisando a Tabela 8 e 9, confirma que o segmento de Aços Semiacabados opera como uma plataforma de exportação da siderurgia brasileira. Sua dinâmica é ditada pelo comércio internacional, pela competitividade cambial e pela saúde econômica dos grandes mercados consumidores globais. Para o nosso modelo preditivo, isso apoia a decisão de incluir variáveis como o PIB dos EUA e o consumo de aço na América do Norte como os fatores de maior poder explicativo para prever o comportamento deste segmento, segundo relatado por meio da tabela 10.

4.2 VARIÁVEIS DO MODELO PREDITIVO

Por meio da etapa de análise de negócios, foram consolidados os parâmetros, sendo definidos 14 parâmetros, todos em escala contínua, totalizando 1460 linhas de informações, distribuídas diariamente 2020 entre 2024, conforme mostra a Tabela 11.

Tabela 11- Parâmetros do modelo

Categoria	Parâmetro (Variável)	Descrição e Relevância para o Modelo
1. Demanda Interna (Brasil)	Consumo_Aparente_Aco	Mede a demanda total por aço no mercado brasileiro.
	PIB_Construcao_Civil	Reflete a atividade da construção civil
	Producao_Veiculos	Mede o ritmo da indústria automotiva
	Taxa_Selic	A taxa básica de juros do Brasil.
2. Comércio Internacional	Importacao_Aco	Volume de aço estrangeiro entrando no Brasil
	Exportacao_Aco	Volume de aço brasileiro vendido ao exterior
3. Demanda Externa	PIB_EUA	Mede a saúde econômica dos Estados Unidos
	Consumo_Aco_America_Norte	Indicador direto da demanda por aço América do Norte
	PIB_Mexico	Mede a saúde econômica do México
	PIB_Canada	Mede a saúde econômica dos Canadá
	PIB_America_Sul_Proxy	Mede a saúde econômica dos países sul-americanos
4. Variáveis-Alvo (Produção)	Producao_Plano_Real	Produção real de Aços Planos
	Producao_Longo_Real	Produção real de Aços Longos
	Producao_Semi_Acabado_Real	Produção real de Semiacabados

Fonte: Autor (2025).

4.3 AVALIAÇÃO DE PERFORMANCE

Foi aplicada a cada modelo de predição (modelo para Aço longo, plano e semiacabado) as métricas escolhidas, sendo Erro Médio Absoluto (MAE) unidade de medida (toneladas) e o Coeficiente de Determinação (R^2), unidade em porcentagem da variação.

Tabela 12 – Resultado das métricas de desempenho

Modelo (Tipo de Aço)	Erro Médio Absoluto (MAE)	Coefficiente de Determinação (R²)
Aço Plano	78,5 mil toneladas	88,20%
Aço Longo	71,3 mil toneladas	91,50%
Semi-Acabados	95,9 mil toneladas	75,80%

Fonte: Autor (2025).

Os resultados nos revelam que, as métricas de avaliação de performance foram atendidas, demonstrando que o modelo criado atende aos critérios do negócio. Portanto, após a fase de teste para cada segmento do aço, os modelos foram avaliados gerando os resultados, conforme a Tabela 12.

4.4 RESULTADOS DA SIMULAÇÃO DE CENÁRIOS (ESTUDO DE CASO)

Nesta seção, serão apresentados três cenários construídos, onde parâmetros específicos foram ajustados, para compreendermos os impactos destes parâmetros no conjunto dos modelos criados, e avaliação qualitativa sobre os resultados da simulação, levando em consideração informações que foram observadas na análise exploratória.

Cenário 1, aumento do consumo com maior participação de importados: Neste cenário foi considerado aumento de 5% no Consumo Aparente, mas com a participação dos importados crescendo de forma ainda mais acentuada. O modelo prevê uma queda de -2,7% para planos e -2,0% para longos em relação a 2024. Isso sugere a tese de que a produção nacional é altamente sensível à concorrência externa. A produção de semiacabados, voltada para exportação, permanece praticamente inalterada.

Cenário 2, taxa Selic de estímulo (9%): Em um cenário com juros mais baixos, o modelo prevê um forte reaquecimento da demanda interna. A produção de Aço Longo dispararia, com um crescimento projetado de +11,5% em relação a 2024, um reflexo direto do barateamento do crédito para a construção. O Aço Plano também veria uma expansão robusta de +7,4% em relação a 2024.

Cenário 3, diminuição 10% em relação 2024, na participação do aço brasileiro, no mercado norte-americano: A produção de Semiacabados seria a mais duramente atingida, com uma queda projetada de 12,0% em relação a 2024. Este resultado era esperado e comprova a tese central da análise: a produção de semiacabados (placas e tarugos) no Brasil é diretamente dependente da demanda industrial da América do Norte. Aço Plano Sofreria um impacto negativo, mas muito mais moderado, com uma queda de -2,0% em relação a 2024. Isso ocorre porque, embora a maior parte da produção de planos seja para o mercado interno. Aço longo seria o segmento menos afetado, com uma retração de apenas -0,9% em relação a 2024. Isso confirma que a produção de aço longo é quase que inteiramente ditada pelo ciclo da construção civil no Brasil.

5 CONCLUSÃO

A pesquisa se propôs na aplicação de técnicas de Inteligência Artificial, com o objetivo de desenvolver um sistema preditivo que possa oferecer um suporte analítico, frente ao mercado siderúrgica brasileiro.

A construção do modelo preditivo com o algoritmo *Random Forest Regressor*, demonstrou resultados quantitativos, com Coeficientes de Determinação (R^2) que alcançaram 91,50% para Aços Longos, 88,20% para Aços Planos e 75% para Aços Semi-acabados. Além, dos resultados por meio da métrica MAE com valores de erro absoluto abaixo do limite de 100 mil toneladas, provaram que é possível, de fato, explicar a maior parte das variações da produção neste estudo de caso, a partir de um conjunto bem selecionado de variáveis econômicas e setoriais. A validação qualitativa, realizada por meio da simulação de cenários, confirmou que o modelo compreende a lógica do negócio.

5.1 LIMITAÇÕES DO TRABALHO

Com base nos resultados e aprendizados obtidos, as seguintes limitações foram enfrentadas:

- Disponibilidade e Granularidade: Nem todas as variáveis estavam disponíveis em uma granularidade mensal consistente para todo o período. Dados anuais de produção e de indicadores econômicos internacionais precisaram ser desagregados para uma frequência mensal. Este processo, embora siga práticas estatísticas padrão, é uma aproximação da realidade;
- Horizonte Histórico: A base de dados principal cobre o período de 2010 a 2024. Este é um horizonte temporal relativamente curto para análises de séries temporais e pode ser influenciado por eventos atípicos, como a pandemia de COVID-19 e a subsequente crise na cadeia de suprimentos, porém tal limitação não invalida o estudo de caso, tendo em vista, a arquitetura do algoritmo escolhido, por meio do conjunto de múltiplas árvores de decisão, confere-lhe uma robustez natural a ruídos e outliers;

- Natureza do Algoritmo: O Random Forest é um algoritmo de regressão, mas não é, em sua essência, um modelo de previsão de séries temporais (como ARIMA ou Prophet). Ele não modela explicitamente conceitos como tendência, sazonalidade ou autocorrelação.;
- Modelo Estático: A ferramenta final é um simulador que requer execução e input manual. Ela não está conectada a APIs para atualização automática dos dados em tempo real, sendo uma fotografia do conhecimento disponível até o momento de sua última execução.

5.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Com base nos resultados e aprendizados obtidos, as seguintes recomendações são propostas:

- Expansão e Refinamento das Variáveis: Sugere-se a inclusão de novas variáveis que podem agregar ainda mais poder preditivo ao modelo, como custos de energia elétrica, preços de frete, indicadores de confiança do consumidor e da indústria, e dados mais granulares sobre os mercados de exportação específicos para cada tipo de produto;
- Comparação com Outros Algoritmos de Aprendizado de Máquina: Embora o Random Forest tenha apresentado excelente performance, seria de grande valor acadêmico comparar seus resultados com os de outros algoritmos avançados, como Gradient Boosting (XGBoost, LightGBM) ou modelos de deep learning como as Redes Neurais Recorrentes (LSTMs), que são especialmente adequadas para séries temporais;
- Aumento da Granularidade da Previsão: Um próximo passo natural seria desdobrar a análise para um nível mais detalhado, desenvolvendo modelos capazes de prever a produção de linhas de produtos específicas dentro de cada segmento ou de analisar o comportamento de mercados regionais dentro do Brasil.

REFERÊNCIAS

AGRESTI, A. **Categorical data analysis**. 3. ed. Hoboken: Wiley, 2013.

ASSOCIAÇÃO NACIONAL DOS FABRICANTES DE VEÍCULOS AUTOMOTORES (ANFAVEA). **Anuário da Indústria Automobilística Brasileira 2024**. São Paulo: ANFAVEA, 2025. Disponível em: <https://anfavea.com.br/site/publicacoes/>. Acesso em: 23/05/2025.

BANCO CENTRAL DO BRASIL. **Relatório de Inflação**. Brasília, DF: Banco Central do Brasil, 2021. Disponível em: <https://www.bcb.gov.br/publicacoes/ri>. Acesso em: 02/06/2025.

BANCO CENTRAL DO BRASIL. **Relatório de Inflação**. Brasília, DF: Banco Central do Brasil, 2025. Disponível em: <https://www.bcb.gov.br/publicacoes/ri>. Acesso em: 02/06/2025

BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer, 2011.

BREIMAN, L. **Random Forests**. Machine Learning, v. 45, n. 1, p. 5-32, 2001.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO (CBIC). **Balço e Perspectivas 2024**. Brasília, DF: CBIC, 2024. Disponível em: <https://cbic.org.br/economia/publicacoes/>. Acesso: 24/03/2025

DAVENPORT, T. H.; HARRIS, J. G. **Competing on analytics: the new science of winning**. Boston: Harvard Business School Press, 2007.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2016.

FEW, S. **Information dashboard design: the effective visual communication of data**. Sebastopol: O'Reilly, 2006.

FUNDO MONETÁRIO INTERNACIONAL (FMI). **World Economic Outlook**. Washington, D.C.: FMI, 2025. Disponível em: <https://www.imf.org/en/publications/weo>. Acesso em: 20/06/2025.

GÉRON, Aurélien. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow**. 2. ed. Rio de Janeiro: Alta Books, 2021.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2017.

IANSTITI, M.; LAKHANI, K. R. **Competing in the age of AI: strategy and leadership when algorithms and networks run the world**. Boston: Harvard Business Review Press, 2020.

INSTITUTO AÇO BRASIL. **Anuário Estatístico 2020**. São Paulo: IABr, 2020. Disponível em: <https://acobrasil.org.br/site/publicacoes/>. Acesso em: 14/03/2025.

INSTITUTO AÇO BRASIL. **Anuário Estatístico 2021**. São Paulo: IABr, 2021. Disponível em: <https://acobrasil.org.br/site/publicacoes/>. Acesso em: 14/03/2025.

INSTITUTO AÇO BRASIL. **Anuário Estatístico 2022**. São Paulo: IABr, 2022. Disponível em: <https://acobrasil.org.br/site/publicacoes/>. Acesso em: 14/03/2025.

INSTITUTO AÇO BRASIL. **Anuário Estatístico 2023**. São Paulo: IABr, 2023. Disponível em: <https://acobrasil.org.br/site/publicacoes/>. Acesso em: 14/03/2025.

INSTITUTO AÇO BRASIL. **Anuário Estatístico 2024**. São Paulo: IABr, 2024. Disponível em: <https://acobrasil.org.br/site/publicacoes/>. Acesso em: 14/03/2025.

JAMES, G. et al. **An introduction to statistical learning: with applications in R**. New York: Springer, 2013.

KORTH, H. F. et al. **Database system concepts**. 6. ed. New York: McGraw-Hill, 2010.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to information retrieval**. Cambridge: Cambridge University Press, 2008.

MCKINNEY, W. **Python para análise de dados: tratamento de dados com pandas, NumPy e IPython**. São Paulo: Novatec, 2018.

PROVOST, F.; FAWCETT, T. **Data science para negócios: o que você precisa saber sobre mineração de dados e pensamento analítico**. Rio de Janeiro: Alta Books, 2013.

RUSSELL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. 4. ed. London: Pearson, 2021.

SHNEIDERMAN, Ben. **Human-Centered AI**. Oxford: Oxford University Press, 2022.

SHARDA, R.; DELEN, D.; TURBAN, E. **Business intelligence, analytics, and data science: a managerial perspective**. 4. ed. New York: Pearson, 2019.

WORLD STEEL ASSOCIATION. **Steel Statistical Yearbook 2020**. Brussels: worldsteel, 2020. Disponível em: <https://worldsteel.org/steel-by-topic/statistics/steel-statistical-yearbook>. Acesso em: 24/04/2025

WORLD STEEL ASSOCIATION. **Steel Statistical Yearbook 2021**. Brussels: worldsteel, 2021. Disponível em: <https://worldsteel.org/steel-by-topic/statistics/steel-statistical-yearbook>. Acesso em: 24/04/2025

WORLD STEEL ASSOCIATION. **Steel Statistical Yearbook 2022**. Brussels: worldsteel, 2022. Disponível em: <https://worldsteel.org/steel-by-topic/statistics/steel-statistical-yearbook>. Acesso em: 24/04/2025

WORLD STEEL ASSOCIATION. **Steel Statistical Yearbook 2023**. Brussels: worldsteel, 2023. Disponível em: <https://worldsteel.org/steel-by-topic/statistics/steel-statistical-yearbook>. Acesso em: 24/04/2025

WORLD STEEL ASSOCIATION. **Steel Statistical Yearbook 2024**. Brussels: worldsteel, 2024. Disponível em: <https://worldsteel.org/steel-by-topic/statistics/steel-statistical-yearbook>. Acesso em: 24/04/2025

APÊNDICE A – Consulta SQL

```
WITH
-- extraídos dos relatórios e planilhas do IABr.
Producao_Mensal_Real AS (
  SELECT
    Mes_Referencia,
    Producao_Plano_Real,
    Producao_Longo_Real,
    Producao_Semi_Acabado_Real
  FROM
    raw_data.Producao_IABr
  WHERE
    EXTRACT(YEAR FROM Mes_Referencia) BETWEEN 2020 AND 2024
),

-- provenientes de fontes como FMI, Banco Mundial, BEA, etc.
Demanda_Externa_Mensal AS (
  SELECT
    Mes_Referencia,
    PIB_EUA,
    Consumo_Aco_America_Norte,
    PIB_Mexico,
    PIB_Canada,
    PIB_America_Sul_Proxy
  FROM
    raw_data.Indicadores_Globais_Mensais
  WHERE
    EXTRACT(YEAR FROM Mes_Referencia) BETWEEN 2020 AND 2024
),

-- indicadores econômicos do Brasil
Indicadores_Mensais_Brasil AS (
  SELECT
    Data,
    Consumo_Aparente_Aco,
    Importacao_Aco,
    Exportacao_Aco,
    PIB_Construcao_Civil,
    Producao_Veiculos,
    Taxa_Selic
```

```
FROM
    raw_data.Indicadores_Economicos_Brasil
WHERE
    EXTRACT(YEAR FROM Data) BETWEEN 2020 AND 2024
)

-- criando a base de dados consolidada para o modelo.
SELECT
    -- Chave primária da nossa tabela final
    imb.Data,

    -- Indicadores mensais do Brasil (Demanda Interna e Comércio)
    imb.Consumo_Aparente_Aco,
    imb.Importacao_Aco,
    imb.Exportacao_Aco,
    imb.PIB_Construcao_Civil,
    imb.Producao_Veiculos,
    imb.Taxa_Selic,

    -- Indicadores de Demanda Externa
    dem.PIB_EUA,
    dem.Consumo_Aco_America_Norte,
    dem.PIB_Mexico,
    dem.PIB_Canada,
    dem.PIB_America_Sul_Proxy,

    -- Variáveis-Alvo (Produção Real)
    pmr.Producao_Plano_Real,
    pmr.Producao_Longo_Real,
    pmr.Producao_Semi_Acabado_Real

FROM
    Indicadores_Mensais_Brasil AS imb

LEFT JOIN Producao_Mensal_Real AS pmr
    ON DATE_TRUNC('month', imb.Data) = DATE_TRUNC('month', pmr.Mes_Referencia)

LEFT JOIN Demanda_Externa_Mensal AS dem
    ON DATE_TRUNC('month', imb.Data) = DATE_TRUNC('month', dem.Mes_Referencia)
```

```
ORDER BY  
imb.Data;
```

APÊNDICE B – Programação do modelo preditivo

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, r2_score
import warnings

def train_model(X, y):
    model = RandomForestRegressor(n_estimators=100, random_state=42)
    model.fit(X, y)
    return model

print("--- Inicializando o Simulador de Cenários ---")

base_de_dados

df = pd.read_csv(base_de_dados)

# Limpeza e preparação básica
df.columns = df.columns.str.strip()
df['Data'] = pd.to_datetime(df['Data'])
df = df.set_index('Data')
df.dropna(how='all', inplace=True)

print("Base de dados carregada e preparada com sucesso!")

except Exception as e:
    print(f"\nERRO: Não foi possível carregar ou processar os dados da sua planilha.")
    print(f"Verifique ")
    print(f"Detalhe do erro: {e}")
    exit()

# O modelo será treinado com os dados da planilha.
feature_columns = [
    'Consumo_Aparente_Aco', 'Importacao_Aco', 'Exportacao_Aco', 'PIB_Construcao_Civil',
```

```

    'Producao_Veiculos', 'Taxa_Selic', 'PIB_EUA', 'Consumo_Aco_America_Norte', 'PIB_Mexico', 'PIB_
Canada', 'PIB_America_Sul
]
target_columns = ['Producao_Plano_Real', 'Producao_Longo_Real', 'Producao_Semi_Acabado_Real']
X = df[feature_columns]
y = df[target_columns]
X_train = df[feature_columns]

-----

print("\n--- Calculando Performance dos Modelos ---")

# Separa os dados em Treino (80%), validação (10%) e teste (10%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Dicionário para armazenar os resultados
results = {}

for i, target_name in enumerate(target_columns):
    model_temp = train_model(X_train, y_train.iloc[:, i])

    predictions = model_temp.predict(X_test)

    mae = mean_absolute_error(y_test.iloc[:, i], predictions)
    r2 = r2_score(y_test.iloc[:, i], predictions)

    results[target_name] = {'MAE': mae, 'R2': r2}

# Exibe a tabela de performance
print("    MÉTRICAS DE PERFORMANCE (Avaliadas em Dados de Teste)")
print(f"{'Modelo (Segmento)':<30} | {'MAE (mil toneladas)':<25} | {'R² (Coef. de Determinação)':<30}")
print("-" * 90)
for name, metrics in results.items():
    mae_val = metrics['MAE']
    r2_val = metrics['R2']
    print(f"{'name':<30} | {'mae_val':<25.2f} | {'r2_val':<30.1%}")
print("=====\n")
# =====
print("Treinando modelo para Aço Plano...")
model_plano = train_model(X_train, df['Producao_Plano_Real'])
print("Treinando modelo para Aço Longo...")
model_longo = train_model(X_train, df['Producao_Longo_Real'])

```

```

print("Treinando modelo para Semi-Acabados...")
model_semi = train_model(X_train, df['Producao_Semi_Acabado_Real'])
print("Modelos prontos para simulação!")

GERAR_GRAFICO = True

cenario_para_prever = {
    'Consumo_Aparente_Aco': [2200], 'Importacao_Aco': [450], 'Exportacao_Aco': [1000],
    'PIB_Construcao_Civil': [143], 'Producao_Veiculos': [210],
    'Taxa_Selic': [10.5], 'PIB_EUA': [108], 'Consumo_Aco_America_Norte': [130],
    'PIB_Mexico_Canada': [106], 'PIB_America_Sul_Proxy': [105]
}

# 4. Execução da Análise
print("\n--- Executando Análise para o Cenário Informado ---")
try:
    df_cenario = pd.DataFrame(cenario_para_prever)
    df_cenario = df_cenario[feature_columns]

    # Previsões (os resultados são mensais)
    previsao_plano = model_plano.predict(df_cenario)
    previsao_longo = model_longo.predict(df_cenario)
    previsao_semi = model_semi.predict(df_cenario)

    print("\n=====")
    print("    RESULTADO DA PREVISÃO MENSAL (em mil toneladas)")
    print("=====")
    print(f" Produção Prevista de AÇO PLANO:    {previsao_plano[0]:10.2f}")
    print(f" Produção Prevista de AÇO LONGO:    {previsao_longo[0]:10.2f}")
    print(f" Produção Prevista de SEMI-ACABADOS: {previsao_semi[0]:10.2f}")
    print("=====")

    if GERAR_GRAFICO:
        labels = ['Aço Plano', 'Aço Longo', 'Semi-Acabados']
        valores = [previsao_plano[0], previsao_longo[0], previsao_semi[0]]
        fig, ax = plt.subplots(figsize=(10, 6))
        bars = ax.bar(labels, valores, color=['#003f5c', '#7a5195', '#ef5675'])
        ax.set_title('Previsão de Produção para o Cenário Informado', fontsize=16, weight='bold')
        ax.set_ylabel('Produção Mensal (mil toneladas)', fontsize=12)
        ax.set_ylim(0, max(valores) * 1.15)

```

```
for bar in bars:
    yval = bar.get_height()
    ax.text(bar.get_x() + bar.get_width()/2.0, yval + (max(valores)*0.01), f'{yval:,.0f}'.replace(',', '.'),
ha='center', va='bottom', fontsize=11)
    nome_arquivo_grafico = 'resultado_simulacao.png'
    plt.savefig(nome_arquivo_grafico)
    print(f"\nGráfico do resultado salvo como: '{nome_arquivo_grafico}'")

except Exception as e:
    print(f"\nERRO: Não foi possível gerar a previsão. Verifique os dados inseridos.")
    print(f"Detalhe do erro: {e}")

print("\n--- Simulação Finalizada ---")
```