

INSTITUTO FEDERAL DE SANTA CATARINA

LUIS GUSTAVO FREITAS E SILVA

PREVISÃO DE DEMANDA UTILIZANDO *MACHINE LEARNING* EM UMA LOJA DE
ESPORTES DE PEQUENO PORTE

Caçador

2025

LUIS GUSTAVO FREITAS E SILVA

PREVISÃO DE DEMANDA UTILIZANDO *MACHINE LEARNING* EM UMA LOJA DE
ESPORTES DE PEQUENO PORTE

Monografia apresentada ao curso de Engenharia de Produção do Câmpus Caçador do Instituto Federal de Santa Catarina para a obtenção do diploma de Bacharel em Engenharia de Produção.

Orientador: Prof. Dr. Eric Costa Carvalho

Coorientador: Prof. Dr. Cristiano Mesquita Garcia

Caçador
2025

S586p Silva, Luis Gustavo Freitas e.
Previsão de demanda utilizando *machine learning* em uma loja de esportes de pequeno porte / Luis Gustavo Freitas e Silva ; orientadores: Eric Costa Carvalho, Cristiano Mesquita Garcia. -- 2025.
81 f.

Trabalho de Conclusão de Curso (Graduação)-Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, Caçador, 2025.
Inclui bibliografias.

1. Previsão de demanda. 2. Aprendizado de máquina. 3. Gestão de estoques. 4. Varejo esportivo. I. Carvalho, Eric Costa. II. Garcia, Cristiano Mesquita. III. Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina – Graduação em Engenharia de Produção. IV. Título.

CDD 658.5

LUIS GUSTAVO FREITAS E SILVA

PREVISÃO DE DEMANDA UTILIZANDO *MACHINE LEARNING* EM UMA LOJA DE
ESPORTES DE PEQUENO PORTE

Este trabalho foi julgado adequado para obtenção do título em Engenharia de
Produção, pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa
Catarina, e aprovado na sua forma final pela comissão avaliadora
abaixo indicada.

Caçador, 16 de Dezembro de 2025.

Prof. Eric Costa Carvalho, Dr.

Orientador

Instituto Federal de Santa Catarina, Câmpus Caçador

Prof. Cristiano Mesquita Garcia, Dr.

Coorientador

Instituto Federal de Santa Catarina, Câmpus Caçador

Prof. Sebastian Johann Batista Perini, Me.

Instituto Federal de Santa Catarina, Câmpus Caçador

Prof. Paulo Roberto Córdova, Dr.

Instituto Federal de Santa Catarina, Câmpus Caçador

AGRADECIMENTO

Agradeço, primeiramente, à minha família, em especial a minha mãe, Claudia Martins Freitas e ao meu pai, José Antônio Brandão e Silva (in memorian) por terem acreditado em mim, mesmo nos momentos que pensei que não seria capaz, por compartilharem comigo não apenas as vitórias, mas também por serem minha força e suporte desde o início. Ao meu irmão, Jonas Freitas e Silva e a minha irmã, Eduarda Freitas e Silva que mesmo de longe, sempre fizeram esforços para estarem perto e me apoiado nas decisões da vida.

A minha filha, Clara Lautert Freitas que me mostrou recentemente o que é ser um pai e ter esse amor genuíno.

Aos meus amigos, pelo apoio e incentivo ao longo de todos esses anos e por terem tornado tudo mais leve.

Aos professores, especialmente ao meu orientador, Eric Costa Carvalho e coorientador, Cristiano Mesquita Garcia por toda a paciência, por acreditarem neste trabalho e em mim. E aos demais professores, que compartilharam do seu conhecimento para nos tornarmos excelentes profissionais e pessoas melhores.

A todos que, de alguma forma, participaram desse longo processo, contribuindo não apenas para o meu crescimento acadêmico, mas principalmente pessoal, o meu muito obrigado!

RESUMO

No âmbito empresarial, a gestão de estoques representa um dos principais pilares e de maior influência nos resultados. Uma empresa que possui um controle de estoque bem organizado, tende a apresentar uma maior vantagem competitiva e a administrar melhor seus custos. Um dos principais fatores da administração de estoques é a previsão de demanda, sendo este o tema deste estudo. O objetivo foi implementar e avaliar modelos de *Machine Learning* para previsão de demanda em uma loja de esportes de pequeno porte, com base em seus dados históricos de vendas e variáveis sazonais, visando maior precisão e eficiência na gestão de estoques.

Neste contexto, investigou-se a aplicação de técnicas de ML em um setor marcado por sazonalidade acentuada e recursos limitados. Para o desenvolvimento deste trabalho, realizou-se inicialmente a revisão da literatura para definição dos conceitos teóricos relacionados ao tema. Foi realizado o levantamento dos dados de vendas extraídos do software de gerenciamento da loja. Os dados foram explorados e testou-se em qual algoritmo de ML foi mais adequado. Para avaliação e escolha dos modelos, a partir de uma análise comparativa entre eles, considerou-se o desempenho preditivo, facilidade de interpretação e implementação e os resultados das previsões de cada um em diferentes cenários. Os resultados demonstraram a superioridade da Regressão Linear na maioria dos casos, desafiando a noção de que modelos complexos são mais eficazes, e revelaram que a adoção de abordagens quantitativas baseadas em ML pode gerar recomendações operacionais acionáveis, otimizando os níveis de estoque e reduzindo incertezas na gestão do pequeno varejo esportivo.

Palavras-Chave: Previsão de Demanda. Aprendizado de Máquina. Gestão de Estoques. Varejo Esportivo.

ABSTRACT

In the business context, inventory management represents one of the main pillars and has the greatest influence on results. A company that has a well-organized inventory control tends to have a greater competitive advantage and manage its costs better. One of the main factors in inventory management is demand forecasting, which is the subject of this study. The objective was to implement and evaluate Machine Learning models for demand forecasting in a small sports store, based on its historical sales data and seasonal variables, aiming for greater accuracy and efficiency in inventory management.

In this context, the application of ML techniques in a sector marked by pronounced seasonality and limited resources was investigated. For the development of this work, a literature review was initially carried out to define the theoretical concepts related to the topic. Sales data extracted from the store's management software were collected. The data were explored and the most suitable ML algorithm was tested. For the evaluation and selection of models, based on a comparative analysis between them, predictive performance, ease of interpretation and implementation, and the results of each model's predictions in different scenarios were considered. The results demonstrated the superiority of Linear Regression in most cases, challenging the notion that complex models are more effective, and revealed that the adoption of quantitative approaches based on ML can generate actionable operational recommendations, optimizing inventory levels and reducing uncertainties in the management of small sporting goods retailers.

Keywords: Demand Forecasting. Machine Learning. Inventory Management. Sports Retail.

LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas de um modelo de previsão de demanda	13
Figura 2 - Hierarquia do ML	21
Figura 3 - Árvore de decisão para classificação do conjunto de dados Iris	25
Figura 4 - Funcionamento do XGBoost	27
Figura 5 - Procedimentos adotados na realização do estudo	30
Figura 6 - Os dez produtos mais vendidos	43
Figura 7 - Representação visual da performance do Meião Pinheiros Juvenil Verde	48
Figura 8 - Representação visual da performance da Camisa Pinheiros Verde	48
Figura 9 - Representação visual da performance do Calção Pinheiros Verde	49
Figura 10 - Representação visual da performance do Chinelo Havaianas Top	50
Figura 11 - Representação visual da performance do Chinelo Havaianas Brasil	50
Figura 12 - Representação visual da performance da Caneleira Poker	51
Figura 13 - Representação visual da performance do Meião Penalty Matis Branco	52
Figura 14 - Representação visual da performance do Meião Penalty Storm Preto	52
Figura 15 - Representação visual da performance do Kit 3 Meias Lupo Invisível	53
Figura 16 - Representação visual da performance da Bolsa Pinheiros Verde	54
Figura 17 - Mapa de calor comparativo MAE (Com e Sem Prophet)	55
Figura 18 - Mapa de calor comparativo RMSE (Com e Sem Prophet)	56
Figura 19 - Mapa de calor comparativo R ² (Com e Sem Prophet)	57
Figura 20 - Performance média dos modelos	58

LISTA DE QUADROS

Quadro 1 – <i>Features</i> temporais implementadas na modelagem	34
Quadro 2 - Aplicação dos algoritmos de ML	38
Quadro 3 - Melhor modelo de ML por produto	59
Quadro 4 - Insights de gestão de estoque	60

LISTA DE ABREVIATURAS E SIGLAS

ML - *Machine Learning*

ERP - *Enterprise Resource Planning*

WMS - *Warehouse Management System*

MMS - Média Móvel Simples

MMP - Média Móvel Ponderada

CNNs - Redes Neurais Convolucionais

IoT - *Internet of Things*

MAE - Erro Absoluto Médio

R^2 - Coeficiente de Determinação

RMSE - Raiz do Erro Quadrático Médio

SUMÁRIO

1 INTRODUÇÃO	7
1.1 Definição do problema	8
1.2 Justificativa	8
1.3 Delimitação do estudo	9
1.4 Objetivos	9
1.4.1 Objetivo Geral	9
1.4.2 Objetivo Específico	9
2 REVISÃO DE LITERATURA	10
2.1 Gestão de Estoque	10
2.1.1 Inventário	11
2.1.2 Estoque Mínimo	12
2.1.3 <i>Enterprise Resource Planning</i> (ERP)	12
2.2 Previsão de Demanda	13
2.3 Métodos de Previsão	15
2.3.1 Métodos Qualitativos	15
2.3.2 Métodos Quantitativos	16
2.3.2.1 Séries Temporais	17
2.3.2.2 Média Móvel Simples (MMS)	18
2.3.2.3 Média Móvel Ponderada (MMP)	18
2.3.2.4 Suavização Exponencial Simples	19
2.4 <i>Machine Learning</i>	20
2.4.1 Aprendizado Supervisionado	21
2.4.2 Aprendizado Não-Supervisionado	21
2.4.3 Aprendizado Semi-supervisionado	22
2.4.4 Engenharia de Atributos	22
2.4.5 Modelos de ML	23
2.4.5.1 Regressão Linear	23
2.4.5.2 Árvores de Decisão	24
2.4.5.3 Floresta Aleatória	25
2.4.5.4 XGBoost	26
2.4.5.5 Prophet	27
2.5 ML para Gestão de Estoque	28

3 METODOLOGIA	30
3.1 Classificação da pesquisa	30
3.2 A Empresa	32
3.3 Procedimentos metodológicos	32
3.3.1 Coleta e preparação dos dados	32
3.3.2 Análise exploratória e seleção dos produtos	33
3.3.3 Engenharia de <i>Features</i> para Séries Temporais	33
3.3.4 Implementação e comparação dos algoritmos	35
3.3.5 Avaliação e desempenho dos modelos	37
3.3.6 Análise comparativa dos resultados	39
3.3.7 Validação da viabilidade	40
4 RESULTADOS E DISCUSSÃO	42
4.1 Caracterização do conjunto de dados analisados	42
4.2 Análise comparativa do desempenho dos modelos de ML	44
4.2.1 Performance dos modelos no período de teste	44
4.2.2 Comparativo das métricas por modelos	54
4.3 Avaliação dos modelos de ML	58
4.4 Validação técnica e operacional	60
4.5 Discussões	60
5 CONCLUSÃO	65
REFERÊNCIAS	68

1 INTRODUÇÃO

De acordo com Passari (2003), a gestão eficaz de qualquer empresa, seja no âmbito público ou privado, industrial ou varejista, requer um processo cuidadoso de planejamento. Para que esse planejamento seja efetivo, é fundamental ter a compreensão precisa das condições futuras nas quais a empresa irá atuar, bem como entender a interação dos elementos que condicionam essas expectativas.

A humanidade, desde os primórdios da civilização, utiliza a previsão de demanda para fins de subsistência e planejamento. Egípcios, romanos e gregos já realizavam estimativas empíricas de colheitas e estoques de alimentos, ainda que sem métodos científicos estruturados (Chiavenato, 2003; Wren, 2005).

Segundo Martins e Laugeni (2005), a previsão de demanda é um método que é baseado em dados estatísticos, matemáticos ou econômicos para projetar cenários futuros. Contudo, as empresas de pequeno porte costumam recorrer como processo manual e subjetivo, a intuição do gestor, na qual enfrenta dificuldades em realizar a previsão de demanda devido às variáveis que se encontram presentes no mercado. Além disso, a previsão de demanda realizada de forma intuitiva e empírica pode resultar em excesso de estoque, o que acarreta em aumento de custos de armazenamento; ou em escassez de produtos, levando à perda de vendas. Este cenário pode ser ainda mais extremo ao envolver produtos perecíveis, que possuem data de validade e assim, não podem se manter em estoque por muito tempo.

Nos dias atuais, com o avanço da tecnologia, métodos como o *Machine Learning* (ML) revolucionaram essa prática, permitindo previsões mais precisas e adaptáveis a diferentes setores, inclusive no varejo esportivo de pequeno porte. Conforme Mitchell (1997, p.2), "*Machine Learning* é definido como o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados". Esta característica é importante para a previsão de demanda, pois capacita modelos que se ajustam dinamicamente a padrões complexos de vendas.

Tem-se como exemplo a Amazon, que utiliza sistemas de ML para a previsão de demanda de milhões de produtos globalmente em segundos (Forbes, 2021). Ainda, no varejo da moda, a Zara aplica técnicas de ML para prever tendências do consumidor, coletando informações em tempo real de suas lojas sobre estilos, tamanhos e preferências regionais populares, e esses dados permitem que a Zara preveja a demanda dos clientes com precisão (Sandra, 2022).

Neste contexto, o uso de modelos de ML se apresenta como uma alternativa promissora para otimizar a previsão de demanda da Loja de Esporte com o objetivo de aprimorar a gestão do estoque, diminuir incertezas associadas a oscilações de mercado e aumentar a eficiência operacional. Desse modo, este trabalho demonstra como as técnicas de ML podem ser aplicadas de forma viável no varejo esportivo, contribuindo para a sustentabilidade financeira e operacional do negócio.

1.1 Definição do problema

A empresa em questão não utiliza métodos quantitativos para a previsão de demanda, tendo como base apenas métodos qualitativos. Sendo assim, pautados na experiência do gestor da loja, realiza os pedidos dos itens de forma subjetiva. Portanto, resulta no acúmulo ou escassez de produtos que influenciam diretamente na gestão do estoque e na eficiência operacional. Diante disso, como a implementação de modelos de Machine Learning pode substituir ou complementar o método manual e subjetivo de previsão de demanda utilizado em uma loja de esportes de pequeno porte, com base em dados históricos de vendas e variáveis sazonais, a fim de alcançar maior precisão, eficiência operacional e sustentabilidade financeira na gestão de estoques?

1.2 Justificativa

A previsão de demanda é um elemento essencial para a gestão eficiente de estoques e a sustentabilidade econômica das empresas, especialmente no setor varejista. Porém, pequenos negócios, como as lojas de esportes, frequentemente enfrentam problemas por falta de recursos e conhecimentos técnicos para utilizar métodos quantitativos avançados, dependendo de abordagens subjetivas e empíricas. Este cenário pode resultar em duas situações críticas: excesso de estoque, que eleva custos de armazenamento e prejuízos (principalmente para produtos perecíveis e sazonais), ou a falta de produtos, ocasionando perda de vendas e insatisfação dos clientes.

Dessa forma, o ML surge como uma alternativa viável e acessível, mesmo para empresas de pequeno porte, devido ao avanço de ferramentas computacionais mais baratas e acessíveis. Ao contrário dos métodos convencionais, que exigem

formulações matemáticas complexas e análises manuais, os algoritmos de ML tem a capacidade de assimilar padrões históricos de vendas, integrar variáveis sazonais, tendências de mercado e até mesmo fatores externos, aprimorando significativamente a precisão das previsões.

Portanto, este estudo busca preencher uma lacuna prática ao evidenciar que o ML não é exclusivo apenas a grandes corporações, mas sim uma ferramenta viável e transformadora também para pequenos negócios, promovendo uma gestão mais inteligente, redução de incertezas e maior sustentabilidade financeira.

1.3 Delimitação do estudo

O presente trabalho delimita-se a estudar modelos de previsão de demanda de ML com o intuito de definir aquele que seja apropriado para aplicação em uma Loja de Esporte de pequeno porte, no qual, tem como foco na análise de dados do segmento varejo, no qual a empresa atua, provendo serviços para atender às necessidades no varejo esportivo para clientes do tipo Pessoa Física na cidade de Taquari/RS.

1.4 Objetivos

1.4.1 Objetivo Geral

Implementar modelos para previsão de demanda de uma loja de esportes de pequeno porte utilizando ML, com base em seus dados históricos de vendas e variáveis sazonais.

1.4.2 Objetivo Específico

- Coletar e analisar dados históricos de vendas, sazonalidades e promoções;
- Selecionar algoritmos de ML para implementação do modelo;
- Avaliar e testar os modelos de previsão de demanda utilizando métricas apropriadas;
- Analisar a viabilidade operacional da implementação do modelo na realidade da Loja.

2 REVISÃO DE LITERATURA

Neste capítulo constam as referências bibliográficas utilizadas para a conceituação dos principais temas relacionados a este trabalho, sendo eles: Gestão de Estoque, Previsão de Demanda, Métodos de previsão, *Machine Learning* e ML para Gestão de Estoque.

2.1 Gestão de Estoque

Segundo Dias (2018), a gestão de estoques consiste em um processo organizacional que abrange o planejamento, controle e monitoramento dos recursos materiais, com o objetivo de manter um equilíbrio entre a disponibilidade de produtos e os custos associados à sua armazenagem. Ampliando essa perspectiva operacional, Ballou (2019) destaca que a gestão de estoques é o conjunto de métodos e abordagens estratégicas voltadas para a manutenção de níveis ideais de produtos, garantindo o atendimento às demandas dos clientes enquanto minimiza custos operacionais excessivos e evita falhas na cadeia de suprimentos.

Slack *et al.* (2018), complementa com a definição de estoque sendo o acúmulo de materiais, recursos, informações e clientes. A forma de administrar esse acúmulo é denominada “Gestão de Estoque”, que é fundamental para a saúde da empresa. Reduzir o estoque pode liberar mais capital de giro, resultando em mais liquidez, mas, por outro lado, uma redução excessiva pode gerar problemas no atendimento ao cliente, como atrasos ou falta de produtos.

Nesse contexto, Bowersox (2014) reforça que a falta de produtos disponíveis pode levar não apenas à perda de vendas, mas também a danos à reputação da empresa. Por isso, é imprescindível ter um gerenciamento estratégico que conecte eficiência operacional e satisfação do consumidor. Segundo Vago *et al.* (2013), as etapas básicas para controle de estoque são:

- Estabelecer o que se deve ter em estoque;
- Determinar quando e o quanto comprar;
- Informar ao setor de compras a intenção da aquisição;
- Receber, armazenar e controlar os materiais em estoque;
- Identificar itens obsoletos ou danificados e removê-los do estoque.

Conforme os autores Facchini *et al.* (2019), um dos principais fatores a ser

considerado dentro da concepção do gerenciamento é a determinação de um estoque mínimo, também conhecido como estoque de segurança. Sua aplicação consiste em definir uma quantidade mínima necessária de estoque para que se consiga proporcionar o nível de atendimento exigido, ou seja, o estoque adicional visa absorver flutuações de demanda e variações no sistema de reposição sem comprometer o atendimento ao consumidor.

2.1.1 Inventário

Conforme Martins (2009) menciona a importância do uso de ferramentas para a administração do estoque considerando a relevância financeira que ele representa para organização, sendo essas ferramentas os principais componentes para uma gestão eficiente. Dentre essas ferramentas, o inventário se destaca. Fonseca (2014) retrata o inventário como um processo de verificação de extrema importância para a empresa. A sua realização tem como objetivo coletar os dados das entradas e saídas de estoque dos produtos, comparando saldos iniciais e finais e, analisam a quantidade de vendas realizadas no período.

Contudo, quando mal gerenciado, esse processo pode ocasionar em falhas. Conforme Arnold (1999), a imprecisão dos registros de estoque pode gerar uma série de efeitos indesejáveis para as organizações, dentre eles o autor destaca:

- Baixa produtividade;
- Baixo nível de serviço;
- Expedição excessiva: envios emergenciais com frequência;
- Excesso de estoque;
- Falta de material e reprogramações constantes;
- Perda de vendas.

O autor ainda ressalta que os efeitos apresentados podem ser vitais para a boa performance das operações das organizações. Isso se deve ao fato de que, além de reduzir a receita, podem resultar no aumento dos custos. Um dos custos diretamente afetados refere-se ao desempenho operacional, influenciado pela dificuldade em planejar materiais e em programar a produção sem a certeza de um saldo preciso dos estoques. Isso gera, muitas vezes, solicitações urgentes para os fornecedores e alterações frequentes nos cronogramas de produção, ocasionando, por exemplo, um número superior de trocas de ferramentas do que o inicialmente programado.

2.1.2 Estoque Mínimo

Pozo (2007) afirma que o estoque de segurança, também conhecido como estoque mínimo, consiste em uma determinada quantidade de itens de determinado produto disponível no estoque, com o objetivo de suprir eventuais variações no sistema, que podem ser atrasos na entrega por parte dos fornecedores, rejeição do lote de compra ou aumento na demanda.

A importância do estoque de segurança é o elemento chave para a definição do ponto de pedido adequado. Dias (2018) também considera o estoque mínimo como um aspecto essencial para a gestão de estoques, ressaltando a menor quantidade de material estocado que a empresa deve ter disponível para enfrentar qualquer tipo de atraso na produção, entrega de produtos e reposição. Portanto, o estoque mínimo é o equilíbrio entre garantia de continuidade e otimização de custos, exigindo análise constante para adaptar-se a dinâmicas de mercado.

2.1.3 *Enterprise Resource Planning* (ERP)

Conforme Bowersox, Closs e Cooper (2006), *Enterprise Resource Planning* (ERP) é um sistema de informação que possui um banco de dados capaz de auxiliar grande parte das atividades da organização, como acompanhar, monitorar e gerar relatórios. Para Nettstrater *et al.* (2015), Programas ERP facilitam as atividades relacionadas aos estoques, desde o recebimento até a expedição, como também contagem e realização de inventários através de módulos que fornecem as informações e mecanismos necessários.

Segundo Banzato (2005) a administração dos estoques, WMS (*Warehouse Management System*) é um sistema que integra o ERP, otimizando todas as atividades relacionadas ao estoque, desde o nível operacional com processos de recebimento, estocagem e expedição até o nível gerencial com emissão de documentos e relatórios.

Neste sentido, Ackerman (2004) diz que o WMS faz a utilização de *hardware* (ferramentas) e *software* (programas) para facilitar as atividades dos colaboradores envolvidos com armazenagem. Martins (2005) complementa afirmando que o uso de softwares pode trazer melhoria nos processos, diminuindo o retrabalho enquanto

aumenta a qualidade dos processos.

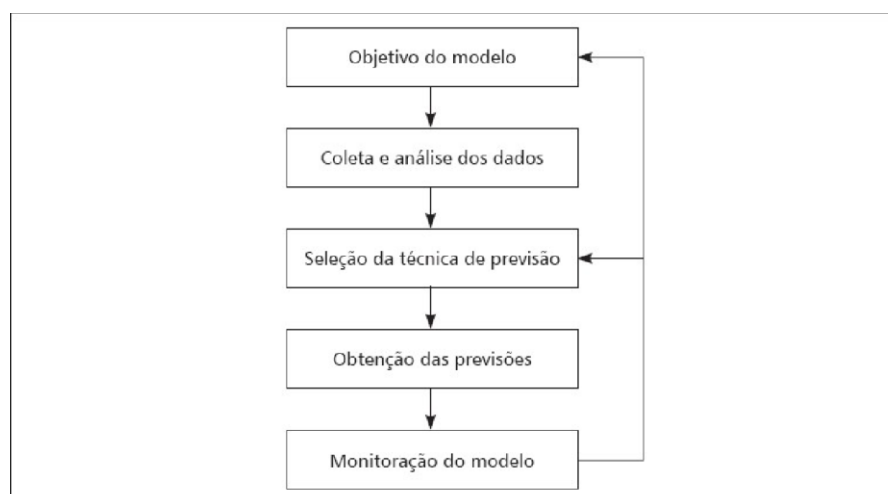
2.2 Previsão de Demanda

Para Tubino (2007, p.25), “a previsão de demanda é a base para o planejamento estratégico da produção, vendas e finanças de qualquer empresa”, destacando seu papel central na gestão empresarial. Essa visão é complementada por Da Veiga *et al.* (2013), que enfatiza o caráter analítico da previsão, ao defini-la como um processo de determinar e prever cenários por meio da ponderação de dados comerciais e mercadológicos.

Moreira (2011), afirma que para todo planejamento, a base comum é a previsão de demanda. Através das análises de previsão de demanda, a empresa deve estabelecer aonde quer chegar, por meio de metas. A partir do momento em que se houve a previsão de demanda, pode-se planejar as vendas futuras, é o ponto de partida para estimar o potencial de mercado.

Segundo Tubino (2007), um modelo de previsão de demanda pode ser dividido em cinco etapas, sendo essas apresentadas na Figura 1:

Figura 1 - Etapas de um modelo de previsão de demanda



Fonte: Tubino (2007)

A primeira etapa consiste na definição da razão pela qual há necessidade de realizar previsões. Que produto ou família de produto será prevista, qual nível de assertividade e detalhamento será aplicado à previsão. Em seguida, conforme Tubino (2007), é feita a coleta e análise dos dados históricos do item em questão, para que

seja possível identificar a técnica de previsão que melhor se adaptará ao estudo em questão. Tubino (2007) ainda destaca itens importantes a serem considerados, sendo eles:

A) Quanto mais extensos os dados históricos, maior tendência à confiabilidade tem a previsão;

B) Dados devem ser selecionados de modo a caracterizar a demanda da empresa, que nem sempre corresponde às vendas anteriores, pois podem haver faltas de produtos, atrasando entregas ou impossibilitando o atendimento;

C) Variações extraordinárias da demanda, como por exemplo, decorrentes de promoções, devem ser analisadas e substituídas por valores mais prováveis, compatíveis com o comportamento normal da demanda;

D) O tamanho do período de consolidação dos dados (semanal, mensal, anual, etc.) tem influência direta na escolha do método de previsão adequado.

A terceira etapa compreende a decisão da técnica de previsão que seja mais adequada ao caso estudado. Essa técnica pode ser do tipo qualitativa ou quantitativa, cada uma delas tendo seu próprio campo de atuação e aplicabilidade. Não existe uma técnica que seja adequada a todas as situações.

De acordo com Tubino (2007), a escolha passa por fatores como custo e assertividade, visto que técnicas mais precisas normalmente demandam maior investimento em sua formulação, seja pelo tempo que demandam ou pela ferramenta a ser empregada. Ainda são definidos outros fatores que devem ser observados na escolha da técnica de previsão:

- Disponibilidade de dados históricos;
- Experiência anterior com a aplicação de determinada técnica;
- Disponibilidade de tempo para a coleta, análise e preparação dos dados e previsão;
- Período de planejamento para qual se necessita da previsão.

A quarta etapa do processo tem como objetivo a obtenção das projeções futuras da demanda. A próxima e última etapa consiste na análise das previsões, avaliando a extensão do erro entre a demanda real e a prevista, com o objetivo de verificar se a técnica empregada é válida. Caso negativo, pode ser necessária uma nova elaboração desde a primeira etapa, optando por uma nova técnica de previsão.

2.3 Métodos de Previsão

Para Moreira (2011) e Corrêa e Corrêa (2012), é possível classificar os métodos de previsões, com o tipo de instrumentos e conceitos, a partir disto forma-se a previsão. Os principais métodos citados pelos autores são:

a) Qualitativos (ou baseados no julgamento): são métodos que repousam basicamente no julgamento de pessoas que, de forma direta ou indireta, tenham condições de opinar sobre a demanda futura, tais como gerentes, vendedores, clientes, fornecedores, etc.

b) Quantitativos (ou Matemáticos): são aqueles que utilizam modelos matemáticos para chegar aos valores previstos. Permitem controle do erro, mas exigem informações quantitativas preliminares. Neste método se subdivide em:

- Métodos causais: a demanda de um item ou conjunto de itens está associada a uma ou mais variáveis internas ou externas à empresa.
- Séries temporais: a análise de séries temporais nada exige além do conhecimento de valores passados da demanda (ou, de forma geral, da variável que se quer prever).
- Regressão simples: é o caso em que considera a demanda ligada a apenas uma variável causal.
- Regressão múltipla: é o caso em que são consideradas duas ou mais variáveis causais supostamente ligadas à demanda.

2.3.1 Métodos Qualitativos

Corrêa e Corrêa (2012) descreve os métodos qualitativos como um método mais dificultoso a ser julgado, já que precisa ser analisado por opiniões de especialistas. Para Moreira (2011, p. 295) “Os métodos qualitativos são baseados no julgamento e na experiência de pessoas que possam, por suas próprias características e conhecimentos, emitir opiniões sobre eventos futuros de interesse”

Moreira (2011), afirma que o método qualitativo não restringe apenas para a previsão de demanda, podendo ser utilizado para análise do cotidiano como: movimentos do comércio internacional, tendência de novos produtos, entre outros.

Temos alguns exemplos de métodos qualitativos, como as Opiniões de Executivos, de acordo com Moreira (2011, p.295) se define como “Um grupo

(geralmente pequeno) de altos executivos da empresa reúne-se para desenvolver em conjunto uma previsão. O grupo é formado por executivos vindos de áreas diversas, como Marketing, Finanças, Produção, etc.”

Como também, temos o exemplo a Opinião da Força de Vendas, na qual, segundo Krajewski, Ritzman e Malhotra (2009, p.439) “As estimativas de força de vendas são previsões compiladas de estimativas feitas periodicamente por membros da força de vendas da empresa”.

Outro método conhecido, temos a Pesquisa de Mercado, conforme autores Corrêa e Corrêa (2012, p.250) “Esse método solicita diretamente dos possíveis clientes ou consumidores sua intenção de compra futura [...]”. Para Moreira (2011), a opinião dos consumidores é o que de fato determinam a demanda, ficar atento sobre as intenções de consumo, se são positivas ou negativas, determinam a previsão de demanda

Por fim, o Método Delphi, segundo Moreira (2011, p.298) “[...] consiste na reunião de um grupo de pessoas que deve opinar sobre um certo assunto, dentro de regras determinadas para a coleta e a depuração das opiniões”. Conforme Corrêa e Corrêa (2012), contextualiza afirmando que envolvem de 6 a 12 especialistas, o coordenador do grupo faz perguntas, resultando em várias opiniões sobre qual variável queiram prever, em seguida coletam-se as várias opiniões de forma individual, computando em tratamento estatístico estas opiniões, refazem suas estimativas até chegar o resultado desejado, com o objetivo de gerar boas previsões.

2.3.2 Métodos Quantitativos

Conforme Corrêa e Corrêa (2012), os modelos quantitativos são baseados em modelos matemáticos que utilizam dados históricos disponíveis, supõem-se, que dados passados serão relevantes para o futuro. Com o uso de métodos quantitativos de previsão presume-se que o comportamento identificado no passado permanecerá no futuro, ou seja, os padrões reconhecidos terão influência na previsão futura.

Para Tubino (2007, p.18) “As previsões baseadas em séries temporais partem do princípio de que a demanda futura será uma projeção dos seus valores passados, não sofrendo influência de outras variáveis”. Com a disponibilidade de dados é possível plotar os gráficos e reconhecer suas características obtidas através das curvas. Pode-se identificar em uma curva temporal a tendência, sazonalidade, ciclos

e variáveis aleatórias.

Neste contexto os autores Gaither e Fraizer (2007) definem a tendência como sendo determinada pelo movimento progressivo a longo prazo, onde a tendência pode ser representada pela elevação dos valores ou um declínio. A sazonalidade possui um padrão que se repete de tempos em tempos, e está relacionada a questões que não se pode controlar ou prever como mudanças de temperatura.

2.3.2.1 Séries Temporais

De acordo com Moreira (2011, p.307), “[...] uma série temporal é a sequência de observações da demanda (no caso mais geral, de uma variável qualquer) ao longo do tempo”. O autor também afirma que as observações podem ocorrer ao longo de um intervalo de tempo, como dias, semanas, meses, trimestres, semestres, anos, entre outros. Para Krajewski, Ritzman e Malhotra (2009, p.442) “Em vez de usar variáveis independentes para a previsão, como fazem os modelos de regressão, os métodos de séries temporais usam informações históricas a respeito apenas da variável dependente”.

Para Peinado e Graeml (2007) as séries temporais podem apresentar quatro características. São elas:

- Nível: O nível da demanda refere-se a um patamar do volume de vendas das demandas passadas, desconsiderando variações de sazonalidade e variações aleatórias;

- Tendência: Os dados históricos podem apresentar uma tendência crescente, estabilizada ou decrescente, podendo ainda ser classificada como linear ou não linear. Esses dados são oriundos do longo prazo;

- Sazonalidade: Situação onde a demanda apresenta um padrão de variação ao longo do tempo, podendo este ser interpretado e previsto, por seguirem um padrão repetitivo. Pode estar ligado a influência de alterações climáticas ou férias escolares, por exemplo.

- Aleatoriedade: variações de demanda que ocorrem devido a fatores excepcionais, tais como catástrofes ou greves, e que portanto, não podem ser previstos.

Os autores afirmam que métodos que utilizam apenas dados de desempenho passado devem ser combinados com outros métodos de previsão, uma vez que o

futuro é incerto e mutável.

2.3.2.2 Média Móvel Simples (MMS)

Conforme os autores Fernandes *et al.* (2011) a média móvel simples é o método mais elementar dentre os métodos quantitativos e deve ser aplicado apenas para previsões de curto prazo e sem sazonalidade. Segundo Tubino (2007), para tal cálculo é usado dados de um determinado número de períodos, preferencialmente os mais recentes, para então gerar sua previsão. A cada novo período de previsão calculado se substitui o dado mais antigo pelo mais recente, e assim sucessivamente. A média móvel é determinada a partir da Equação 1.

$$Mm_n = \frac{\sum_{i=1}^n D_i}{n} \quad (1)$$

Onde:

Mm_n = média móvel de n períodos

D_i = demanda ocorrida no período i

n = número de períodos

i = índice do período ($i = 1, 2, 3, \dots$)

Conforme Tubino (2007) a vantagem do uso do método da média móvel está em sua simplicidade operacional e facilidade de entendimento, mas tem sua limitação na necessidade de armazenar um grande volume de dados. Outro fator a ser considerado é que ela fornece apenas a previsão para o período imediatamente posterior, sendo que para os períodos futuros não é possível ter uma estimativa. Chase (2013) complementa, salientando que em casos onde a variação da demanda for repentina, a média móvel simples é incapaz de alcançar o nível de mudança ideal dentro de um período de tempo razoável.

2.3.2.3 Média Móvel Ponderada (MMP)

Moreira (2011) afirma que a média móvel ponderada possui em comum com a MMS o fato de tomar n valores reais anteriores da demanda para a composição média.

Diferentemente da MMS, porém, os valores recebem pesos diferentes, geralmente refletindo uma maior importância dada aos valores mais recentes da demanda.

Conforme Krajewski, Ritzman, Malhotra (2009), citam que para calcular a média ponderada deve-se multiplicar os pesos de cada período pelo o valor desse período, somando-os no final. Os autores ainda afirmam que o método de média móvel ponderada tem as mesmas deficiências que o método de média móvel simples: os dados devem ser repetidos por n períodos de demanda para permitir o cálculo de média para cada período.

O cálculo utilizado na MMP é obtido pela Equação 2:

$$P_j = (D_1PE_1) + (D_2PE_2) + (D_3PE_3) + \dots + (D_nPE_n) \quad (2)$$

Onde:

P_j = previsão para o período j

PE_i = peso atribuído ao período i

D_i = demanda do período i

n = número de períodos

i = índice do período ($i = 1, 2, 3, \dots$)

Peinado e Graeml (2007) ainda sugerem a utilização dos pesos de forma que sua soma seja igual a um, para que não seja necessária uma etapa extra ao final, que é a divisão do resultado pelos pesos adotados.

2.3.2.4 Suavização Exponencial Simples

Para os autores Krajewski, Ritzman, Malhotra (2009), média móvel com suavização exponencial é um procedimento avançado que computa uma série temporal utilizando as demandas recentes, atribuindo pesos às observações anteriores. Os autores ressaltam que é o método de previsão formal mais usado por causa de sua simplicidade e da pequena quantidade de dados necessários para sustentá-lo.

Moreira (2011) cita a Equação 3 para a média móvel com suavização exponencial.

$$D_t = D_{t-1} + \alpha(Y_{t-1} - D_{t-1}) \quad (3)$$

Onde:

D_t = previsão para o período t

D_{t-1} = previsão para o período $(t - 1)$

α = constante de suavização ou de alisamento ($0 \leq \alpha \leq 1$)

Y_{t-1} = demanda real para o período $(t - 1)$

Para Peinado e Graeml (2007) o valor da constante de suavização (α) deve variar entre zero e um. Se o valor de α se aproximar de um, menor será a influência da demanda real do último período na previsão de demanda. A parte da equação representada por $(1 - \alpha)$ refere-se à taxa exponencial com que vai cair a influência dos dados históricos de demanda.

2.4 Machine Learning

A Inteligência Artificial, produz um vasto conjunto de técnicas, dentre as quais se destaca o *Machine Learning* (ML). No que se refere a origem desta expressão, o seu termo foi pronunciado pela primeira vez pelo pesquisador norte-americano Arthur Samuel em 1959. Samuel (1959) conceitua o ML como o campo de estudo que possibilita que máquinas adquiram conhecimento sem que sejam codificadas de maneira explícita. O pesquisador desenvolveu um sistema pioneiro, o *Game of Checkers*, que simula partidas de damas entre humano e computador, demonstrando pela primeira vez a capacidade de aprendizado autônomo. Segundo Taulli (2020, p.64), este trabalho revolucionário para a época, demonstrou que um computador poderia aprender por meio do processamento de dados sem ter sido explicitamente programado para realizar tal tarefa.

Conforme os autores Faceli *et al.* (2019, p.341), o ML visa construir modelos computacionais capazes de descrever sistemas complexos a partir da observação de dados. Um exemplo de sua aplicação é a utilização de algoritmos em plataformas de streaming como a Netflix e o Prime Vídeo da Amazon para realizar sugestões de mídias baseadas no comportamento do usuário. Dessa forma, os autores Faceli *et al.* (2019, p. 3) também afirmam que os algoritmos programados com ML aprendem com

base na experiência passada, por meio de um princípio de inferência denominado indução, no qual se obtêm conclusões genéricas a partir de um conjunto particular de exemplos.

Conforme ilustrado na Figura 2, o ML pode ser classificado em abordagens distintas, conforme detalhado a seguir.

Figura 2: Hierarquia do ML



Fonte: Faceli *et al.* (2019, p. 6)

2.4.1 Aprendizado Supervisionado

O aprendizado supervisionado, conforme Haykin (2001) e Castro e Ferrari (2016, p. 16), baseia-se em conjuntos de dados rotulados, onde as respostas esperadas são conhecidas a priori, representando o comportamento desejado do sistema. Nessa abordagem, a figura do supervisor, seja um especialista humano ou um critério pré-definido, atua fornecendo os rótulos corretos durante o treinamento, permitindo que algoritmos de classificação (decisão binária) ou regressão (previsão de valores contínuos) aprendam padrões. Uma abordagem de aplicação é a filtragem de spams em e-mails, onde o algoritmo é treinado com mensagens já classificadas como 'spam' ou 'não spam', aprendendo a generalizar para novos dados.

2.4.2 Aprendizado Não-Supervisionado

O aprendizado não-supervisionado fundamenta-se, de acordo com Castro e Ferrari (2016, p. 16), apenas nos elementos da base, cujos rótulos são desconhecidos. Basicamente, o algoritmo deve aprimorar sua capacidade de

“categorizar” ou rotular os dados brutos, sem dispor da figura do supervisor. O aprendizado não supervisionado é aplicado com o intuito de identificar padrões em conjuntos de dados que, geralmente, se encontram desestruturados, sendo a abordagem mais comum, segundo Taulli (2020, p. 77), o agrupamento (*clustering*), que manipula dados não rotulados e utiliza algoritmos para agrupar itens semelhantes em categorias.

Nesse contexto, a aprendizagem não-supervisionado pode ser utilizada na extração de dados de mídias sociais, no controle de empréstimos de livros ao traçar o perfil dos usuários de uma biblioteca e na validação de transações bancárias. Por fim, é relevante mencionar um dado interessante, conforme Taulli (2020, p. 78), as aprendizagens humana e animal são, em grande parte, não supervisionadas, pois o mundo, sob a perspectiva desses dois indivíduos, é explorado por meio de observações.

2.4.3 Aprendizado Semi-supervisionado

O aprendizado semi-supervisionado, segundo Taulli (2020, p. 79) é uma mistura de aprendizado supervisionado e não-supervisionado que surge quando se tem uma pequena quantidade de dados não rotulados. Este tipo de aprendizado utiliza um conjunto de dados de treinamento rotulados e não rotulados com o objetivo de induzir um modelo preditivo. Para Chapelle, Schölkopf e Zien, (2006), os algoritmos do aprendizado semi-supervisionado realizam comparações entre os dados e classificam por alguma medida que identifique a semelhança entre esses dados para serem rotulados com base nos dados que já possuem rótulos. Temos como exemplo, do uso de algoritmos de aprendizado semi-supervisionado a identificação de fraudes em todas as transações feitas com cartão de crédito a partir de algumas transações que tem conhecimento que são fraudes ou não.

2.4.4 Engenharia de Atributos

Segundo Domingos (2012), o sucesso de um projeto de *machine learning* depende fundamentalmente das variáveis empregadas. Tais variáveis são produtos de um processo chamado engenharia de *features*, que inclui desde a coleta e limpeza até a transformação e análise dos dados.

Além das técnicas de aprendizado, conforme Grus (2018), outro aspecto importante no ML é a preparação dos dados, conhecida como Engenharia de Atributos ou de *Features*. Esse é um processo de transformação dos dados de um estado bruto para um estado mais adequado para a modelagem. Seu intuito é extrair o máximo de informações que serão utilizadas durante a criação do modelo preditivo. Para isso, é necessário manipular seus atributos e buscar informações que muitas vezes estão implícitas no conjunto de dados. As principais técnicas incluem modificar e criar atributos a partir dos já existentes.

O autor ainda afirma que os atributos dos dados impactam de maneira direta os modelos preditivos e os resultados que podem ser obtidos. Uma das abordagens empregadas na construção de novas variáveis consiste em gerar novos atributos a partir dos atributos existentes que apresentam alta correlação com o atributo alvo. Para o autor, algoritmos de modelagem preditiva mais complexos executam a importância e a seleção de atributos internamente enquanto constroem seu modelo. Geralmente, a relevância de atributos fornece pontuações que indicam quão útil cada atributo foi durante a construção das árvores (de decisão ou regressão).

2.4.5 Modelos de ML

Segundo Zhou (2021), ML pode ser definido como “técnica que melhora o desempenho de sistemas aprendendo através da experiência de métodos computacionais”. Essas experiências se manifestam em forma de dados. Assim, a principal função do ML é elaborar algoritmos aptos a criar modelos a partir do conjunto de dados. O conjunto de dados tem como um termo equivalente, mais frequentemente utilizado, na língua inglesa sendo *dataset*.

De acordo com a perspectiva apresentada por Lipton (2018), a interpretabilidade de um modelo está relacionada à habilidade de um ser humano entender o modelo como um todo, em uma única análise abrangente. Isso implica a exigência de uma perspectiva global sobre como o modelo treinado realiza suas decisões, considerando não apenas as características envolvidas, mas também cada elemento aprendido, como pesos, estrutura e outros parâmetros que determinam o funcionamento do modelo.

2.4.5.1 Regressão Linear

A regressão linear, segundo Russell e Norvig (2009), constitui um dos métodos mais simples de aprendizado de máquina supervisionado, representando um modelo frequentemente utilizado nas situações em que o atributo alvo y é contínuo. Esse modelo possibilita a elaboração de uma função linear que minimize o erro quadrático e, por conseguinte, se ajuste de maneira mais eficaz a todos os pontos. Para a previsão de qualquer novo valor, o domínio, que pode ser formado por uma ou mais variáveis, é aplicado à função. Segundo Angrist e Pischke (2009), as técnicas de regressão funcionam como ferramentas estatísticas que calculam diferenças entre grupos tratados e de controle, independentemente da inclusão ou não de variáveis de controle.

2.4.5.2 Árvores de Decisão

Conforme Russell e Norvig (2009) uma árvore de decisão representa uma função que tem como entrada um vetor de valores de atributos e retorna uma decisão (um valor de saída único). Os valores de entrada e saída podem ser discretos ou contínuos. Para valores de saída discretos, as árvores são chamadas de árvores de classificação, já para prever um valor contínuo de saída, é necessário uma árvore de regressão.

Russell e Norvig (2009) também ressaltam que uma árvore de decisão atinge sua conclusão realizando uma série de avaliações. Cada nó da árvore corresponde a um teste do valor de um dos atributos de entrada, já as ramificações dos nós são classificadas com as possíveis respostas do teste. Cada nó da árvore representa uma decisão que pode ser retornada pela função. Quanto mais um atributo é usado para tomar decisões importantes, maior é sua importância relativa.

Para exemplificar melhor, temos a Figura 3 que apresenta a estrutura de uma árvore de decisão, utilizada para classificar espécies de flores, tendo como base um conjunto de dados amplamente reconhecido, denominado Iris.

Figura 3 - Árvore de decisão para classificação do conjunto de dados Iris



Fonte: Scikit (2023)

De forma geral, os algoritmos de árvores de decisão usam a estratégia de dividir e conquistar. Dessa forma, o algoritmo divide o problema em subproblemas menores e mais simples, construindo uma árvore de decisão com base em regras lógicas que determinam como as decisões devem ser tomadas em cada nó da árvore, como pode ser visto no exemplo da Figura 3. O processo se inicia pelo nó raiz (primeiro nó da árvore), a partir dele são criadas as divisões denominadas nós, que na Figura 3 são representados por caixas. Essas caixas, contêm informações de cada teste, como número de amostras de treino do nó (*samples*), formas de se medir a impureza do nó (*gini*) e quantos elementos de cada classe o nó em questão contém valor (*value*).

2.4.5.3 Floresta Aleatória

Segundo Oshiro (2013), a floresta aleatória é um dos algoritmos mais utilizados

na aprendizagem supervisionada, devido à sua semelhança e ao fato de que pode ser aplicado tanto em tarefas de classificação quanto de regressão. Ele é um *ensemble*, ou seja, vários classificadores combinados para potencializar os resultados ao realizar a predição de novos exemplos.

Diante disso, o autor afirma que esse algoritmo separa amostras aleatórias a partir do conjunto de treino. Cada nova árvore é produzida utilizando uma amostra aleatória e um subconjunto de atributos aleatórios, dentre os atributos aleatórios é escolhido um atributo mais representativo para ser o atributo de decisão. Este procedimento cria uma grande diversidade de árvores, o que geralmente leva a criação de modelos melhores.

2.4.5.4 XGBoost

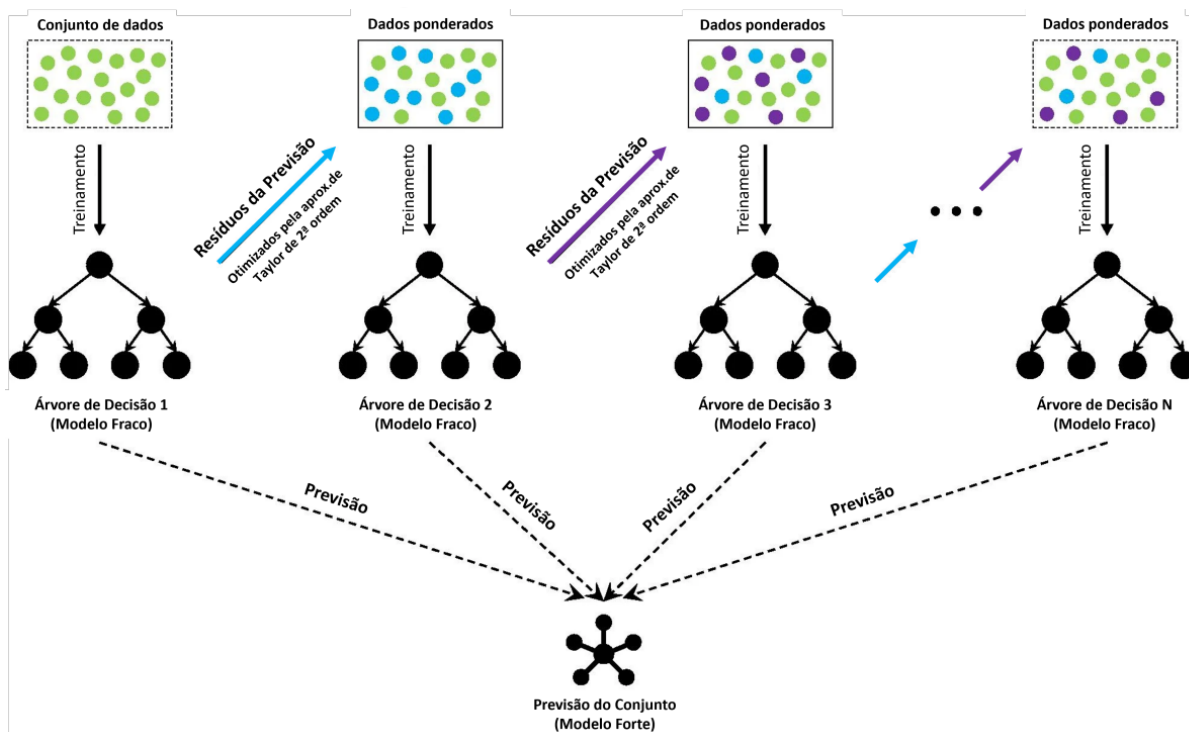
O algoritmo XGBoost foi proposto inicialmente em um trabalho seminal de Chen e Guestrin (2016). De acordo com os autores, sua popularidade é atribuída a vantagens-chave como alta escalabilidade, que confere ganhos significativos de velocidade em relação a implementações anteriores de *gradient boosting*, aliada a uma notável flexibilidade e a um desempenho consistente em diversos tipos de problema e conjuntos de dados.

Para Wang e Ross (2018), o XGBoost é um algoritmo de aprendizado de máquina fundamentado em árvores de decisão, sendo aplicável a problemas de regressão e classificação. Trata-se de uma evolução do método de *gradient boosting*, cuja principal inovação reside na expansão da função de custo com a incorporação de um termo de regularização. Conforme Chen e Guestrin (2016), esse termo tem como objetivo explicitamente penalizar a complexidade do modelo, mitigando a variância e combatendo o sobreajuste (*overfitting*). O algoritmo opera como um *ensemble* sequencial, no qual árvores de decisão de profundidade reduzida são construídas iterativamente para corrigir os erros dos modelos anteriores.

A Figura 4 demonstra o funcionamento do XGBoost. Conforme explica Grandó (2023, p. 27), o treinamento do algoritmo é um processo iterativo e sequencial, no qual cada nova árvore tem o objetivo específico de corrigir os desvios (resíduos) deixados pelas árvores anteriores. A inovação deste método está no uso de uma aproximação de Taylor de segunda ordem para otimizar a função de perda de forma mais eficiente e rápida a cada iteração, em vez de depender apenas do gradiente de primeira ordem,

como era comum em algoritmos de *boosting* tradicionais.

Figura 4 - Funcionamento do XGBoost



Fonte: Grando (2022)

2.4.5.5 Prophet

Conforme Taylor e Letham (2017), o Prophet é um *framework* de previsão de séries temporais de código aberto, originalmente desenvolvido pelo Facebook e disponibilizado em 2017 para Python e R. Os autores afirmam que sua criação foi motivada pela necessidade de simplificar a geração de previsões precisas e robustas. O algoritmo opera por meio da decomposição da série em elementos-chave: tendência, sazonalidade e efeito de feriados. Ainda de acordo com os autores, essa estrutura é ideal para séries influenciadas por comportamentos humanos, que tipicamente exibem sazonalidades sobrepostas de diferentes períodos. O modelo é capaz de identificar, de forma automática, padrões que se repetem semanalmente (como a queda de movimento em finais de semana) e anualmente (como picos sazonais em épocas de festividade), integrando-os em uma previsão unificada.

Corrêa e Vieira (2021) apontam que o Prophet destaca-se pela capacidade de acomodar múltiplas sazonalidades e pela não exigência de periodicidade regular nas

séries temporais, características que viabilizam sua aplicação em cenários complexos como a previsão de preços de ativos financeiros.

2.5 ML para Gestão de Estoque

Para Lingam (2018), o uso de ML tem impactado diversas áreas, incluindo a gestão de inventário. O ML pode ajudar em tomadas de decisões relacionadas à gestão de estoque em indústrias de comércio eletrônico, por exemplo, melhorando a precisão nas previsões de demanda.

Islam *et al.* (2024) ressalta que a eficácia da cadeia de suprimentos é um elemento que contribui para a competitividade das empresas, uma vez que a gestão de estoque está intimamente relacionada ao custo. Segundo Gonçalves (2022), com o surgimento da Indústria 4.0, que enfatiza o uso inteligente de informações em vez de ativos físicos, a tomada de decisões pelos gestores é significativamente aprimorada com as estruturas de ML.

Conforme García-Barrios (2021), a implementação de ferramentas de inteligência artificial representa uma nova oportunidade para processar grandes volumes de informações complexas em tempo real, o que não era viável anteriormente. Ferramentas de ML podem ser empregadas para análise e categorização do estoque, e, nos próximos anos, podem se destacar como uma das tecnologias mais eficazes para previsão de inventário devido ao seu potencial para análises complexas de dados. Ao combinar computação em nuvem, *big data* e algoritmos de aprendizado, é viável avaliar grandes volumes de dados considerando diversos fatores de demanda com alta precisão e mínima ocorrência de erros. Para Lingam (2018), essas ferramentas dispõem de capacidade de autocorreção, o que também proporciona maior eficiência em comparação às técnicas tradicionais de previsão, dado que é possível realizar análises repetidamente e comparar resultados.

Conforme Tang *et al.* (2023), podemos também explorar de forma ampla a aplicação de ML em técnicas de previsão de demanda e controle de inventário como forma de aprimorar parâmetros de reabastecimento ou compreender sazonalidade de demandas. Segundo Islam *et al.* (2024), financeiramente, foi observado que sistemas inteligentes são capazes de reduzir custos operacionais e melhorar a alocação de recursos, elevando a agilidade e a eficiência da cadeia logística. Além disso, Para Pasupuleti *et al.* (2024), o uso de ML pode elevar a sustentabilidade nas operações

de inventário diminuindo desperdícios.

Segundo Xu *et al.* (2023) um ponto muito interessante é a aplicação das redes neurais convolucionais (CNNs) na logística e gestão de inventário e como as CNNs estão sendo adaptadas para otimizar processos de logística e logística reversa, para ajustar automaticamente os níveis de estoque em armazéns inteligentes. Para Tang *et al.* (2023) essas redes ajudam a identificar padrões complexos em dados de inventário, trazendo a possibilidade de criar correlações mais complexas entre os padrões de entradas e saídas de itens e permitindo ajustes em tempo real.

Spirito (2024) afirma que as CNNs são extremamente eficazes para monitorar o fluxo de entrada e saída de produtos, facilitando a categorização automatizada e a organização do inventário com base em dados históricos e projeções futuras. Conforme Jokela (2024), quanto à aplicação olhando para a logística reversa, aumentamos a capacidade de prever quando e quais produtos podem ser devolvidos, e quando devolvidos, otimizar o processo de recolhimento e reintegração ao estoque. Isso não só reduz os custos associados à devolução de mercadorias, como também melhora a disponibilidade de produtos reciclados para venda, fazendo um uso mais sustentável dos recursos.

Conforme Liu *et al.* (2022) os armazéns inteligentes podem juntar as capacidades do ML para trabalhar em sinergia com sensores IoT (*Internet of Things*, Internet das Coisas) e sistemas de gestão de inventário (tipo WMS) para ajustar dinamicamente os níveis de estoque. Eles analisam dados em tempo real de diversas fontes e traqueiam a localização de itens das bases logísticas. Ajustam o nível de estoque imediatamente quando o item sai da localização, somado à informação de utilização do sistema, criam pedidos de reabastecimento para evitar falta de estoque e criam movimentações necessárias para maximizar a utilização do estoque.

3 METODOLOGIA

Este capítulo descreve a organização do estudo, iniciando pela classificação da pesquisa, pela caracterização da empresa objeto de investigação, por fim, detalhando os procedimentos metodológicos adotados.

3.1 Classificação da pesquisa

As pesquisas são classificadas de acordo com sua natureza. Conforme Gerhardt e Silveira (2009) e Gil (2017), a pesquisa básica trata-se de uma investigação sobre determinado tema que tem como objetivo descobrir, aprimorar ou expandir um conhecimento científico. A princípio, não há aplicação prevista ou imediata desse saber. Consiste no desenvolvimento do conhecimento por meio do conhecimento. Os autores ainda ressaltam sobre a pesquisa aplicada, na qual, visa encontrar respostas para problemas observados pelos pesquisadores. Constitui-se em uma pesquisa que busca respostas ou alternativas para problemas concretos. É elaborada com um interesse específico. Portanto, classifica-se a pesquisa realizada por este estudo como sendo de natureza aplicada.

Em relação à abordagem de pesquisa, Gerhardt e Silveira (2009) e Fonseca (2002) se referem a pesquisa qualitativa como sendo uma busca de entender a dinâmica e os significados de um fenômeno dentro de um grupo social. Por se basear na interpretação do pesquisador, possui um caráter subjetivo e tendo como foco principal a percepção e o sentido que as pessoas atribuem ao objeto ou situação estudada. Quando os autores falam sobre a pesquisa quantitativa, referem-se a uma abordagem que analisa um fenômeno por meio de dados quantitativos, utilizando métodos estatísticos e linguagem matemática para explicar suas causas e relação entre variáveis. Por ter base em medições numéricas, é centrada na precisão e objetividade. Com base nisso, este estudo se classifica como sendo uma abordagem quantitativa.

Quanto aos objetivos, Severino (2017) afirma que a pesquisa exploratória como o nome indica, busca explorar um tema específico, proporcionando maior compreensão e familiarização com o assunto e o problema de pesquisa. De acordo com Gil (2017), a maioria dos estudos acadêmicos, especialmente em sua fase inicial, apresenta características de pesquisa exploratória, o autor ainda comenta sobre a

pesquisa descritiva que visa descrever as características de um objeto, população ou fenômeno. Além da descrição, busca identificar e analisar possíveis relações entre as variáveis envolvidas no contexto estudado. Por fim, a pesquisa explicativa, segundo Gil (2017) e Severino (2017) tem como objetivo identificar os fatores que determinam ou influenciam o assunto em estudo. A finalidade é explicar as causas e os motivos por trás dos fatos observados. Neste contexto, este se caracteriza como uma pesquisa exploratória.

Dentre os métodos, existe o método dedutivo, que conforme Gil (2008) baseia-se em princípios aceitos como verdades incontestáveis, permitindo chegar a conclusões por meio de raciocínio puramente lógico e formal. No método indutivo, de acordo com Marconi e Lakatos (2003) permite alcançar conclusões mais abrangentes do que as premissas iniciais nas quais o raciocínio utilizou como base de pensamento. Este estudo define-se como um método indutivo.

Por fim, no que se tange aos procedimentos adotados, dentre os diversos tipos, destaca-se para o presente trabalho, a pesquisa-ação e o estudo de caso. Conforme Fonseca (2002), Gil (2017) e Severino (2017), a pesquisa-ação envolve a interação direta entre pesquisador e os membros onde a investigação está sendo desenvolvida, observa sistematicamente as respostas e analisa os efeitos dessa intervenção. E temos o estudo de caso que, segundo os autores, trata-se de uma pesquisa que se fundamenta na observação minuciosa de uma unidade de análise, que pode ser desde um indivíduo até um grupo, sala de aula, instituição ou comunidade. O pesquisador seleciona uma unidade relevante para conduzir uma investigação detalhada, realizando uma coleta e análise de dados detalhada.

Essas abordagens metodológicas alinham-se à Extensão Universitária que pode ser compreendida como um processo educativo, cultural e científico que combina ensino, pesquisa e extensão, promovendo saberes acadêmicos e demandas sociais através de uma relação benéfica entre a universidade e comunidade, com o propósito de gerar uma transformação social. Nesse intuito, o IFSC ao ofertar cursos de graduação tecnológica, demonstra potencial para desenvolver atividades de extensão em parceria com a comunidade externa, transformando o conhecimento acadêmico em ferramenta de desenvolvimento social e profissional. Portanto, este estudo se trata de uma combinação de estudo de caso e pesquisa-ação por envolver o discente com a sociedade, onde o conhecimento adquirido sobre previsão de demanda e gestão de estoque vem de encontro às necessidades e desafios empresariais.

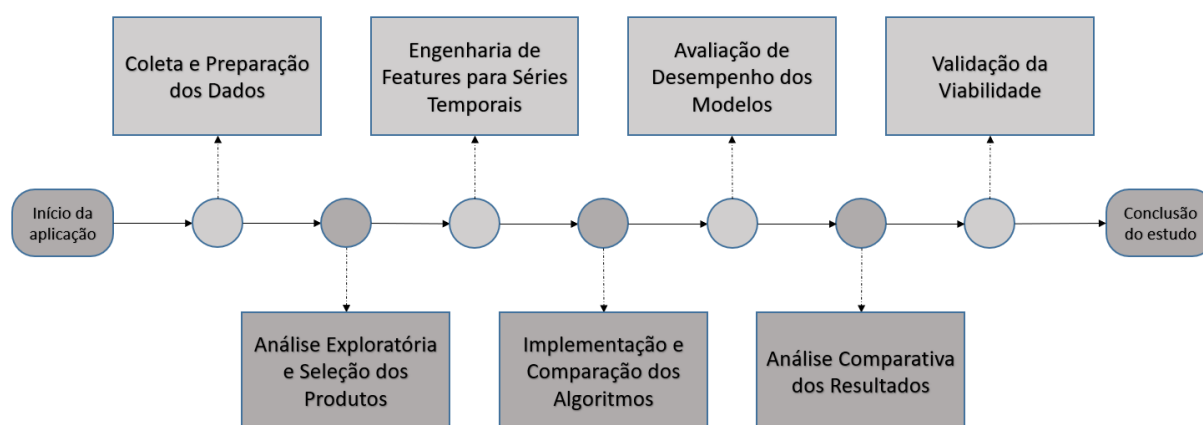
3.2 A Empresa

A empresa em estudo iniciou suas atividades em 1988 na cidade de Taquari/RS. A empresa está estabelecida no ramo do varejo esportivo, com um modelo de negócio focado em dois eixos principais: o atendimento ao consumidor final (pessoas físicas) e o fornecimento para entidades esportivas locais. Entre seus clientes institucionais, destaca-se escolas e escolinhas de diversas modalidades, como futebol, futevôlei, vôlei e beach tennis, às quais fornece artigos e equipamentos esportivos.

3.3 Procedimentos metodológicos

Conforme detalhado na Figura 5, temos os procedimentos adotados para realização do estudo:

Figura 5: Procedimentos adotados na realização do estudo



Fonte: Elaborado pelo autor (2025)

3.3.1 Coleta e preparação dos dados

O estudo iniciou com a coleta dos dados através do software empresarial utilizado pela loja de seu histórico de vendas e estoque dos produtos, abrangendo o período de janeiro de 2022 a outubro de 2025. Dada a realidade operacional de uma empresa de pequeno porte, onde o ERP é simples e sem muitos meios de exportação dos dados, a coleta foi realizada de forma manual. Os registros foram extraídos dos relatórios operacionais do sistema de gestão da loja, copiados e consolidados

manualmente em uma planilha do Microsoft Excel e, posteriormente, convertidos para formato CSV para processamento em Python. Esta abordagem inicial reflete o contexto de negócio no qual a pesquisa está inserida, caracterizado por processos manuais e acessíveis. O primeiro conjunto de dados, "Dados estoque - DMF"¹, contém informações sobre o estoque da empresa objeto de estudo, variando de Janeiro/2022 a Outubro/2025. O conjunto de dados contém 461 linhas, onde cada linha representa o registro mensal de um produto específico, contendo informações como Período (Mês/Ano), Produto, Estoque Inicial, Estoque Final, Quantidade Reposta, Vendas no mês e Custo Unitário. O segundo conjunto de dados, "Dados de venda - DMF"², contém dados de venda, considerando os dez produtos mais vendidos, incluindo a Descrição do item, Período da Venda (Mês/Ano), Quantidade vendida e Valor de venda.

3.3.2 Análise exploratória e seleção dos produtos

Com os dados estruturados, realizou-se uma análise quantitativa utilizando bibliotecas Python (pandas e numpy) para identificar os produtos com maior impacto no negócio. O critério de seleção adotado foi o volume total de vendas mensais no período histórico, priorizando itens que representassem a maior parcela do movimento da loja e, conseqüentemente, tivessem maior potencial de otimização de estoque.

3.3.3 Engenharia de *Features* para Séries Temporais

A engenharia de *features* consistiu na transformação sistemática dos dados históricos de vendas em variáveis preditivas temporais utilizando Python. Este processo foi essencial para habilitar os modelos de ML a capturar padrões temporais complexos.

As *features* implementadas foram organizadas em cinco categorias principais:

1. **Features temporais básicas:** Foram extraídas informações diretas das datas, incluindo mês (1 a 12, representando o período do ano), trimestre (1 a 4 para identificar ciclos de negócio sazonais) e ano (componente para tendências de longo prazo).

¹ Disponível em <https://bit.ly/dados-estoque-dmf>

² Disponível em <https://bit.ly/dados-venda-dmf>

2. **Features cíclicas transformadas:** Foram aplicadas transformações trigonométricas (seno e cosseno) utilizando nymphy para representar adequadamente a natureza circular dos meses, evitando que os modelos interpretem dezembro (mês 12) como temporalmente distante de janeiro (mês 1).

3. **Features auto regressivas (lags):** Foram criados três níveis de dependência temporal: vendas do mês anterior (t-1), vendas de dois meses atrás (t-2) e vendas de três meses atrás (t-3), fundamentadas no princípio de que eventos passados influenciam eventos futuros, implementadas com pandas.

4. **Médias móveis:** Foram implementadas médias móveis com base em janelas temporais de 3 e 6 meses para suavizar flutuações aleatórias e identificar tendências subjacentes, utilizando funções de pandas.

5. **Feature de tendência linear:** Uma variável sequencial crescente foi criada para capturar padrões de crescimento ou declínio de longo prazo na série temporal. Essa *feature* atribui um valor inteiro incremental a cada observação temporal, permitindo que os modelos identifiquem tendências lineares subjacentes nos dados de vendas.

O Quadro 1 demonstra as *features* temporais implementadas na modelagem.

Quadro 1 - *Features* temporais implementadas na modelagem

Categoria	Nome <i>feature</i>	Descrição	Justificativa
Temporais básicas	mes	Extração numérica do mês (1-12)	Capturar sazonalidade mensal
	trimestre	Extração numérica do trimestre (1-4)	Identificar ciclos de negócios sazonais
	ano	Extração numérica do ano	Capturar tendências de longo prazo
Cíclicas Transformadas	mes_sin	Transformação senoidal: $\sin(2\pi * \text{mês}/12)$	Representar natureza circular dos meses
	mes_cos	Transformação cosseno: $\cos(2\pi * \text{mês}/12)$	Complementar representação cíclica com quadraturas
Auto Regressivas	vendas_lag_1	Vendas do mês anterior (1-t)	Capturar dependência

Categoria	Nome <i>feature</i>	Descrição	Justificativa
(lags)			temporal imediata
	vendas_lag_2	Vendas de dois meses atrás (2-t)	Capturar padrões de médio prazo
	vendas_lag_3	Vendas de três meses atrás (3-t)	Capturar padrões de longo prazo
Médias Móveis	vendas_media_movel_3	Média móvel de 3 meses	Suavizar flutuações de curto prazo e identificar tendências
	vendas_media_movel_6	Média móvel de 6 meses	Suavizar flutuações de médio prazo e identificar ciclos
Tendência	tendencia	Sequência linear crescente (0, 1, 2, ..., n)	Capturar crescimento/declínio linear de longo prazo
Excluídas (<i>data leakage</i>)	giro_estoque	$\text{vendas_mes} / ((\text{estoque_inicial} + \text{estoque_final})/2)$	Criada para análise contextual
	lucro_bruto	$\text{vlr_venda_bruto} - (\text{vendas_mes} \times \text{custo_unitário})$	Criada para análise contextual

Fonte: Elaborado pelo autor (2025)

A combinação dessas *features* temporais, complementada pela estratégia proativa de imputação para valores ausentes gerados por transformações utilizando bibliotecas Python, formou a base robusta necessária para comparar o desempenho de diferentes algoritmos de ML no contexto específico da previsão de demanda para pequeno varejo esportivo.

3.3.4 Implementação e comparação dos algoritmos

A implementação dos algoritmos em Python envolveu cinco modelos de ML selecionados para representar diferentes abordagens preditivas e famílias algorítmicas, adaptados às particularidades dos dados de varejo esportivo e às

limitações computacionais típicas de pequenas empresas.

A Regressão Linear foi incluída como modelo *baseline* por sua simplicidade interpretativa e baixo risco de *overfitting*, características que são vantajosas para um conjunto de dados limitado temporalmente. Implementada através da função `LinearRegression` do Scikit-learn com configuração padrão, este modelo estabeleceu a linha de base de desempenho contra a qual os algoritmos mais complexos foram comparados. Os algoritmos baseados em árvores foram configurados com parâmetros conservadores para equilibrar capacidade preditiva e controle de complexidade. A Árvore de Decisão utilizou `DecisionTreeRegressor`, onde foi realizado testes de profundidade entre 3 a 15 níveis, demonstrando que com 8 níveis mostrou um melhor equilíbrio, prevenindo o *overfitting* enquanto manteve capacidade de capturar relações não-lineares. A Floresta Aleatória implementou `RandomForestRegressor`, com 100 estimadores balanceando desempenho e tempo computacional, e paralelização total habilitada para eficiência. O XGBoost configurou-se com `XGBRegressor`, utilizando taxa de aprendizado conservadora de 0.1 e profundidade máxima de 6 para *boosting* controlado. O Prophet configurado especificamente para o contexto de dados mensais do varejo esportivo, onde realizou-se a adaptação às particularidades dos dados através de ajustes, como a desativação de sazonalidades irrelevantes (semanal e diária), mantendo apenas componentes anuais e adicionando sazonalidade mensal. Para efeito de comparação do desempenho especializado em séries temporais contra modelos de ML tradicionais, permitindo avaliar se a complexidade adicional do Prophet justifica-se no contexto específico do pequeno varejo esportivo com dados limitados.

O *pipeline* de processamento implementou estratégias unificadas para garantir comparabilidade entre os algoritmos. O tratamento de dados faltantes utilizou `SimpleImputer (strategy='mean')` em todos os modelos Scikit-learn através de *pipelines* integrados. A divisão temporal seguiu abordagem cronológica separando dados de treino (Janeiro/2022 a Setembro/2024 - 33 meses) e teste (Outubro/2024 a Outubro/2025 - 13 meses), simulando cenário real de previsão onde modelos são treinados exclusivamente com dados históricos e testados em períodos futuros desconhecidos. A seleção de *features* excluiu sistematicamente variáveis-alvo (`vendas_mes`) e métricas financeiras derivadas (`estoque_final`, `vlr_venda`, `vlr_venda_bruto`, `giro_estoque`, `lucro_bruto`) para prevenir *data leakage*, mantendo

apenas *features* temporais preditoras independentes.

Esta implementação diversificada em Python permitiu a comparação abrangente entre diferentes paradigmas de modelagem, desde abordagens lineares simples até técnicas avançadas de *ensemble learning*, todas otimizadas para as particularidades dos dados de pequeno varejo esportivo.

3.3.5 Avaliação e desempenho dos modelos

A avaliação dos modelos foi realizada em Python através de três métricas principais calculadas no período de teste (Outubro/2024 a Outubro/2025). Foi utilizado R^2 (Coeficiente de Determinação) como métrica primária para comparar a proporção de variância explicada, complementada por MAE (Erro Absoluto Médio) para interpretação em unidades reais e RMSE (Raiz do Erro Quadrático Médio) para penalizar erros grandes. Todas as métricas foram calculadas via funções do Scikit-learn (`r2_score`, `mean_absolute_error`, `mean_squared_error`) aplicadas aos 13 meses do conjunto de teste.

A estratégia de validação empregou divisão temporal com treino de 33 meses (Janeiro/2022 a Setembro/2024) e teste de 13 meses (Outubro/2024 a Outubro/2025), simulando cenário real de previsão onde modelos são avaliados apenas em dados futuros não vistos durante treinamento. Esta abordagem preservou a sequência temporal natural das séries, evitando vazamento de informação entre conjuntos.

Os critérios de seleção priorizaram desempenho preditivo (R^2) como métrica principal, classificando modelos como "Excelente" ($R^2 > 0.8$), "Bom" ($R^2 > 0.6$), "Regular" ($R^2 > 0.4$), "Ruim" ($R^2 > 0.2$) ou então "Muito Ruim" ($R^2 \leq 0.2$). Para valores de R^2 negativos, considera-se que o modelo é inferior à média simples dos dados. Em casos de desempenho similar, foi utilizado o menor MAE como critério de desempate.

As equações das métricas são apresentadas abaixo. A Equação 4 representa a Raiz do Erro Quadrático Médio (RMSE) que indica que quanto menor o valor, melhor o modelo.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

(4)

Onde:

y_i = Valor real/observado da i -ésima amostra

\hat{y}_i = Valor previsto pelo modelo para a i -ésima amostra

n = Número total de amostras no conjunto de dados

\sum = Somatório dos erros quadrados para todas as amostras

A Equação 5 representa o Erro Absoluto Médio (MAE) que com um valor baixo indica que as previsões do modelo estão próximas dos valores observados, sugerindo um bom desempenho. Um valor alto indica que o modelo precisa ser melhorado.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Onde:

y_i = Valor real da i -ésima amostra

\hat{y}_i = Valor previsto pelo modelo para a i -ésima amostra

n = Número total de amostras

$|\cdot|$ = Valor absoluto

A Equação 6 determina o Coeficiente de Determinação (R^2) tem uma variação de 0 a 1, sendo 1 significando que o modelo explica perfeitamente toda a variabilidade dos dados observados. E valores próximos a 0 demonstra que o modelo não explica nenhuma variabilidade dos dados observados.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Onde:

y_i = Valor real da i -ésima amostra

\hat{y}_i = Valor previsto pelo modelo para a i -ésima amostra

\bar{y} = Média dos valores reais

n = Número total de amostras

3.3.6 Análise comparativa dos resultados

A análise comparativa dos resultados foi realizada em Python através de uma abordagem sistemática, organizada em etapas sequenciais para garantir comparabilidade e robustez das conclusões. Inicialmente, estabeleceu-se um *framework* de avaliação padronizado, onde todos os cinco modelos de ML (Regressão Linear, Árvore de Decisão, Floresta Aleatória, XGBoost e Prophet) foram treinados e testados nas mesmas condições. Esta uniformidade foi alcançada utilizando divisões temporais idênticas: 33 meses para treinamento (Janeiro/2022 a Setembro/2024) e 13 meses para teste (Outubro/2024 a Outubro/2025), garantindo que as comparações não fossem influenciadas por diferenças nos dados de entrada.

O centro da análise comparativa baseou-se na aplicação de três métricas de desempenho: MAE, RMSE e R^2 . Para cada um dos dez produtos analisados, essas métricas foram calculadas individualmente, permitindo uma avaliação granular. Em seguida, realizou-se uma agregação por modelo, calculando-se as médias de performance em todos os produtos, o que proporcionou uma visão geral da eficácia de cada algoritmo. A comparação direta entre modelos foi conduzida através de um sistema de duplo ranking. Primeiro, os modelos foram ordenados pelo R^2 médio, do maior para o menor, onde valores mais altos indicam melhor ajuste aos dados. Em paralelo, um segundo ranking foi gerado com base no MAE médio, do menor para o maior, onde valores mais baixos representam menor erro de previsão. Este duplo critério permitiu identificar não apenas a precisão explicativa (R^2), mas também a magnitude dos erros cometidos (MAE).

Para contextualizar os resultados preditivos, implementou-se uma comparação temporal entre o período de teste e o histórico completo de vendas, como também, um cálculo de variação percentual para identificar mudanças de tendência, correlação entre magnitude de variação temporal e performance dos modelos e, por fim, uma avaliação dos modelos em cenários de crescimento ou declínio.

3.3.7 Validação da viabilidade

A validação da viabilidade da solução proposta foi conduzida aplicando-se a estrutura metodológica completa desde a engenharia de *features* até a avaliação comparativa em um cenário controlado que simulou uma implementação real, totalmente aplicada em Python.

O ponto central da validação residiu na avaliação do desempenho preditivo no período de teste (Outubro/2024 a Outubro/2025), após o treinamento dos modelos com dados históricos (Janeiro/2022 a Setembro/2024). A viabilidade foi considerada comprovada quando os modelos demonstraram capacidade de generalização, isto é, quando suas previsões para dados nunca vistos (o período de teste) apresentaram alta aderência aos valores reais observados.

Os critérios objetivos para validação foram o poder preditivo (R^2) que tem valores consistentemente altos no período de teste para a maioria dos produtos, indicando que os modelos explicam a maior parte da variabilidade da demanda real. Podemos citar também, a precisão operacional (MAE) que o erro médio absoluto baixo, demonstrando que as previsões são suficientemente precisas para suportar decisões de reposição de estoque. Como também, robustez e consistência com um desempenho superior e estável do melhor modelo em nove dos dez produtos analisados, mostrando que a solução não é um artefato específico de um único item.

A viabilidade foi consolidada de forma prática pela tradução automática das previsões em recomendações operacionais acionáveis, como cálculo de estoque de segurança, definição de pontos de reposição e sugestão de lotes econômicos. Este elo entre a previsão estatística e a ação gerencial demonstrou o valor aplicado da solução no contexto do pequeno varejo.

As equações dos *insights* estratégicos para gestão de estoque são apresentadas abaixo. A Equação 7 representa o Estoque de Segurança que serve como uma reserva extra de produtos que a empresa mantém para proteger contra imprevistos e garantir que não falte produto para o cliente.

$$ES = z \sqrt{\sigma_d^2 t + \sigma_t^2 d^2} \quad (7)$$

Onde:

Z = Coeficiente do nível de serviço

σ_d^2 = Desvio padrão da demanda

t = Tempo médio entre pedido e entrega

σ_t^2 = Desvio padrão do tempo médio

d^2 = Demanda média

A Equação 8 demonstra o Ponto de Reposição que é a quantidade mínima em estoque de determinado produto, quando atingida, dispara um alerta para realizar um pedido de compra.

$$PP = CxTR + EM_n \quad (8)$$

Onde:

C . = Consumo médio

TR = Tempo de reposição

EM = Estoque mínimo ou estoque de segurança

A Equação 9 mostra o Lote Econômico que é a quantidade ideal para pedir ao fornecedor que minimiza o custo total do estoque.

$$LEC = \sqrt{\frac{2DS}{mC}} \quad (9)$$

Onde:

D = Demanda anual

S = Custo por pedido

mC = Custo de manutenção de estoques

4 RESULTADOS E DISCUSSÃO

De acordo com a fundamentação teórica apresentada e nas metodologias implementadas, esta pesquisa aplicou técnicas de ML para previsão de demanda no contexto do pequeno varejo esportivo. A análise focou na comparação de cinco modelos preditivos aplicados a dez produtos de alto giro, com o objetivo de validar sua eficácia operacional e gerar recomendações acionáveis de gestão de estoque.

4.1 Caracterização do conjunto de dados analisados

No ambiente do Excel, foi executada a primeira fase de tratamento e estruturação dos dados. As operações incluíram a padronização de nomes de produtos, a correção de inconsistências de digitação, a uniformização de formatos de datas e a criação de uma chave única para o cruzamento seguro entre as tabelas de vendas e de estoque. Foram utilizados dois conjuntos de dados principais: um arquivo contendo informações detalhadas sobre os produtos vendidos, incluindo quantidade, valores de venda e custos unitários, e outro com dados de controle de estoque, apresentando estoque inicial e final, quantidade repostada e vendas mensais. Após a curadoria na planilha, os dados foram exportados para o formato CSV, que então serviram como insumo para o processo de modelagem automatizada em Python.

A fase de preparação computacional envolveu uma limpeza e validação adicional por meio de scripts em Python, assegurando a qualidade e a confiabilidade dos dados utilizados no processo de modelagem preditiva. Esta etapa combinou, portanto, a praticidade de um tratamento manual inicial com o rigor de uma preparação automatizada final, estabelecendo uma base sólida e consistente para as análises subsequentes.

Na Figura 6, são mostrados os dez produtos mais vendidos:

Figura 6 - Os dez produtos mais vendidos



Fonte: Elaborado pelo autor (2025)

A Figura 6 revelou uma concentração significativa de vendas em um grupo específico de produtos. O ranking demonstrou que o Meião Pinheiros Juvenil Verde liderou as vendas com 1.029 unidades, seguido pela Camisa Pinheiros Verde (914 unidades) e Calção Pinheiro Verde (909 unidades). Estes três produtos responderam por uma parcela substancial do total de movimentações. Os demais itens selecionados completaram o top 10, incluindo calçados como Chinelo Havaianas Top Branco (467 unidades) e Chinelo Havaianas Brasil Branco (335 unidades), como também, Meião Penalty Matís Branco (222 unidades), Meião Penalty Storm Preto (220 unidades), Kit3 Meias Lupo Invisível Branco (192 unidades), além de acessórios como Caneleira Poker (281 unidades) e Bolsa Pinheiros Verde (154 unidades).

A seleção dos 10 produtos mais vendidos justificou-se por dois fatores principais: primeiro, por representarem os itens de maior giro no estoque, onde melhorias na previsão de demanda trariam impacto financeiro mais significativo; segundo, por abrangerem diferentes categorias do portfólio, permitindo testar a eficácia dos modelos em variados padrões de consumo. Esta abordagem focada garantiu que o esforço de modelagem fosse direcionado para os produtos que efetivamente movimentam o negócio, maximizando o retorno do trabalho de previsão.

Uma análise preliminar das séries temporais destes produtos já permitiu

observar comportamentos distintos, alguns com demanda relativamente estável, outros com sazonalidade marcante e alguns com tendência de crescimento. Esta variedade de padrões enriqueceu o processo de modelagem, testando a capacidade dos algoritmos em lidar com diferentes cenários de demanda encontrados na operação real da empresa.

4.2 Análise comparativa do desempenho dos modelos de ML

Nesta seção foram analisados os cinco modelos de ML para previsão de demanda selecionados, as métricas de avaliação utilizadas e o período de teste padrão.

4.2.1 Performance dos modelos no período de teste

Conforme apresentado na Metodologia deste trabalho, o treinamento utilizou 33 meses de dados históricos de Janeiro/2022 a Setembro/2024, enquanto o teste empregou 13 meses recentes de Outubro/2024 a Outubro/2025, simulando o desafio real de prever demandas futuras em uma empresa de pequeno porte. Todos os modelos processaram o mesmo conjunto de *features* temporais, permitindo uma comparação justa. O Quadro 2 demonstra os resultados da aplicação dos cinco algoritmos de ML nos dez produtos mais vendidos, tendo como referência suas métricas principais sendo R^2 , MAE e RMSE. Nas Figuras 7, 8, 9, 10, 11, 12, 13, 14, 15 e 16 temos as representações visuais para cada um dos produtos e os modelos selecionados.

Quadro 2 - Aplicação dos algoritmos de ML. Em negrito, os melhores valores obtidos por produto.

Produto	Modelo	MAE	RMSE	R2
Meião Penalty Juvenil Pinheiros Verde	Regressão Linear	1,528	1,699	0,945
	Árvore de Decisão	2,231	2,842	0,846
	Floresta Aleatória	1,825	2,296	0,900

Produto	Modelo	MAE	RMSE	R2
	XGBoost	1,732	3,191	0,806
	Prophet	14,839	17,206	-4,634
Camisa Pinheiros Verde	Regressão Linear	0,955	1,693	0,964
	Árvore de Decisão	1,769	2,304	0,933
	Floresta Aleatória	1,265	1,716	0,963
	XGBoost	1,334	1,796	0,959
	Prophet	16,695	19,126	-3,612
Calção Pinheiros Verde	Regressão Linear	1,203	1,583	0,973
	Árvore de Decisão	1,615	2,434	0,936
	Floresta Aleatória	1,580	2,106	0,952
	XGBoost	1,546	2,198	0,948
	Prophet	15,678	18,657	-2,738
Chinelo Havaianas Top Branco	Regressão Linear	1,043	2,596	0,936
	Árvore de Decisão	3,923	9,491	0,148
	Floresta Aleatória	1,961	3,619	0,876
	XGBoost	1,833	3,094	0,909
	Prophet	65,977	90,829	-77,004
Chinelo Havaianas	Regressão Linear	1,017	1,186	0,990

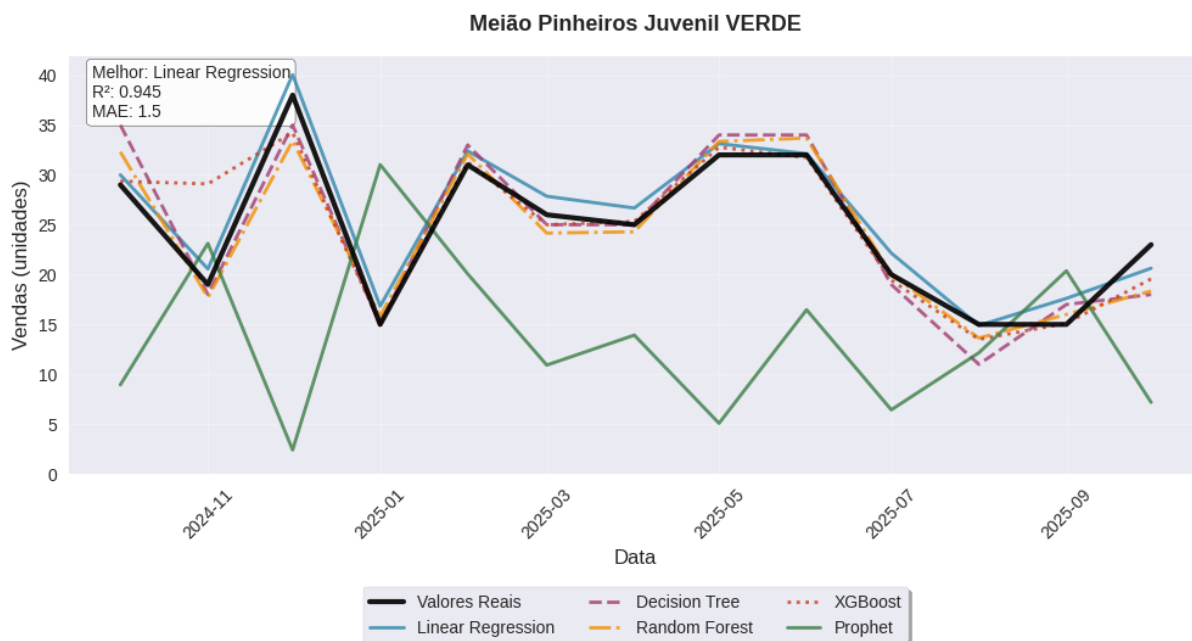
Produto	Modelo	MAE	RMSE	R2
Brasil Branco	Árvore de Decisão	0,769	1,177	0,990
	Floresta Aleatória	1,947	2,766	0,944
	XGBoost	1,684	2,841	0,941
	Prophet	31,040	44,049	-13,164
Caneleira Poker	Regressão Linear	0,154	0,392	0,997
	Árvore de Decisão	1,385	2,481	0,863
	Floresta Aleatória	1,149	2,622	0,847
	XGBoost	1,251	2,689	0,839
	Prophet	58,294	71,717	-113,250
Meião Penalty Matís Branco	Regressão Linear	0,208	0,301	0,989
	Árvore de Decisão	1,000	1,861	0,563
	Floresta Aleatória	0,771	1,029	0,866
	XGBoost	0,613	0,909	0,896
	Prophet	11,898	22,629	-63,582
Meião Penalty Storm Preto	Regressão Linear	0,300	0,423	0,990
	Árvore de Decisão	1,308	2,496	0,647
	Floresta Aleatória	1,102	2,590	0,620
	XGBoost	1,060	2,525	0,638

Produto	Modelo	MAE	RMSE	R2
	Prophet	15,868	23,033	-29,087
Kit 3 Meias Lupo Invisível Branco	Regressão Linear	1,049	1,100	0,772
	Árvore de Decisão	1,462	4,731	-3,213
	Floresta Aleatória	0,570	1,438	0,611
	XGBoost	0,190	0,390	0,971
	Prophet	4,890	6,803	-7,709
Bolsa Pinheiros Verde	Regressão Linear	0,130	0,154	0,996
	Árvore de Decisão	0,077	0,277	0,987
	Floresta Aleatória	0,301	0,520	0,956
	XGBoost	0,080	0,278	0,987
	Prophet	2,274	3,105	-0,573

Fonte: Elaborado pelo autor (2025)

Como podemos perceber no Quadro 2 demonstrando as performances dos produtos vs modelos, o modelo de Regressão Linear demonstrou performance excepcional, sendo o algoritmo com melhor desempenho em 8 dos 10 produtos analisados. Seu R² médio de 0,945 para o Meião Pinheiros Juvenil Verde, 0,964 para a Camisa Pinheiros Verde, e valores como 0,997 para a Caneleira Poker e 0,996 para a Bolsa Pinheiros Verde, evidenciam sua capacidade de explicar entre 94,5% e 99,7% da variabilidade da demanda real. O baixo MAE, variando de apenas 0,13 unidades para a Bolsa Pinheiros a 1,53 unidades para o Meião Pinheiros Juvenil, confirma sua precisão operacional para decisões de reposição de estoque.

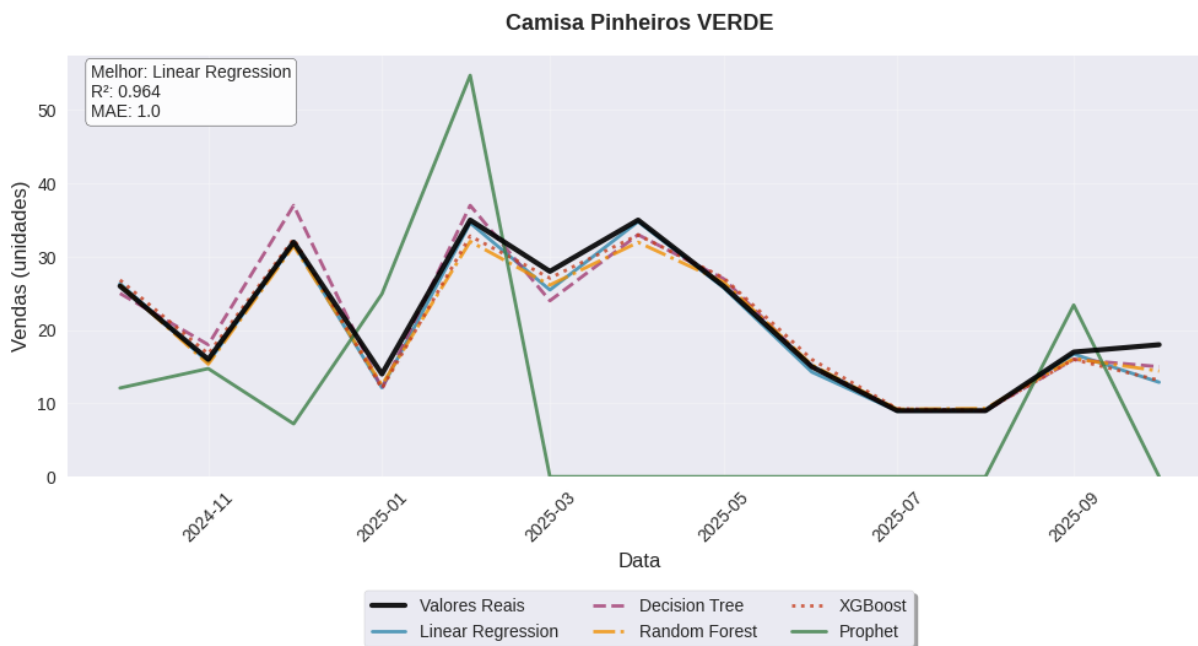
Figura 7 - Representação visual da performance do Meião Pinheiros Juvenil Verde



Fonte: Elaborado pelo autor (2025)

A Figura 7 mostra a análise preditiva para o Meião Pinheiros Juvenil Verde, diante disso, temos que dentre os modelos analisados, a Regressão Linear demonstrou melhor desempenho, alcançando um R² de 0,945 e um MAE de 1,5 unidades indicando erros pequenos na previsão.

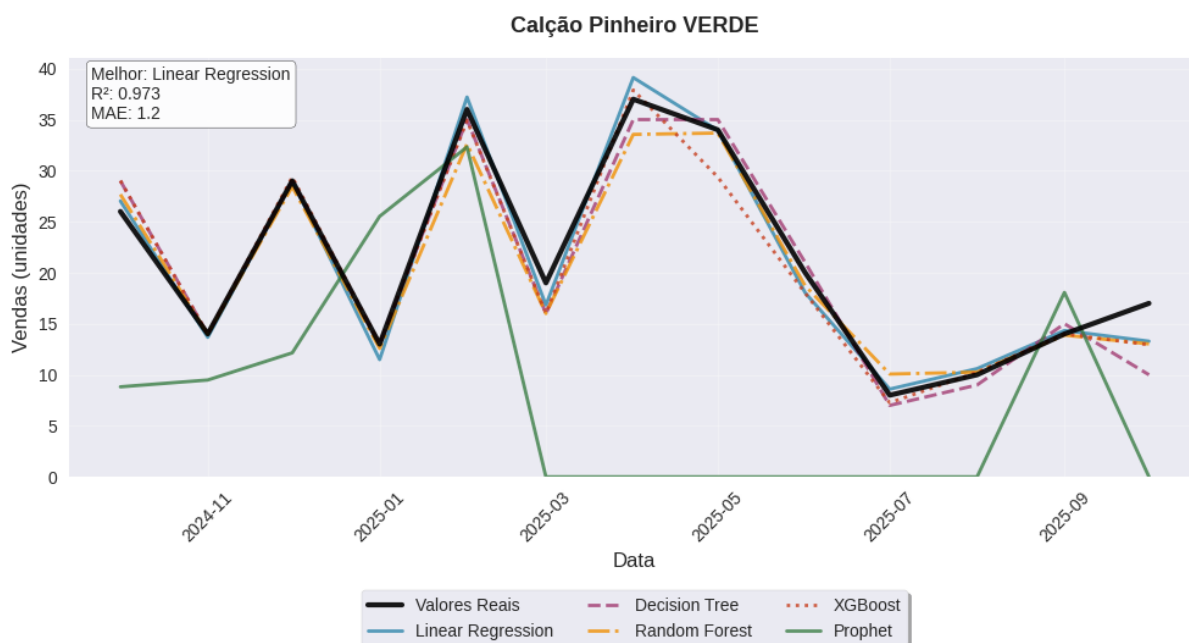
Figura 8 - Representação visual da performance da Camisa Pinheiros Verde



Fonte: Elaborado pelo autor (2025)

Conforme analisado na Figura 8, a linha de tendência projetada pela Regressão Linear acompanha consistentemente os valores reais, demonstrando que as vendas da Camisa Pinheiros Verde segue um comportamento previsível e estável, resultando em R^2 de 0,964 e um MAE de 1,0 demonstrando um poder preditivo alto.

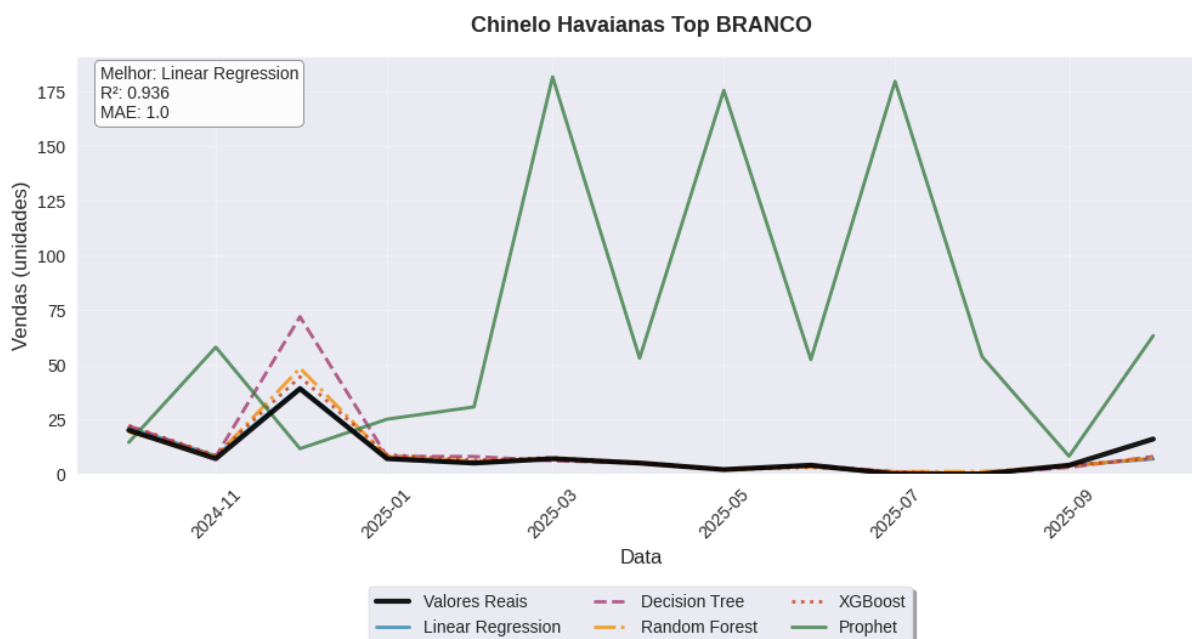
Figura 9 - Representação visual da performance do Calção Pinheiros Verde



Fonte: Elaborado pelo autor (2025)

Na Figura 9 demonstra uma análise preditiva do Calção Pinheiros Verde, onde a Regressão Linear teve superioridade e resultados ótimos de acordo com as métricas de R^2 (0,973) e MAE (1,2), sugerindo uma capacidade de prever a demanda projetada.

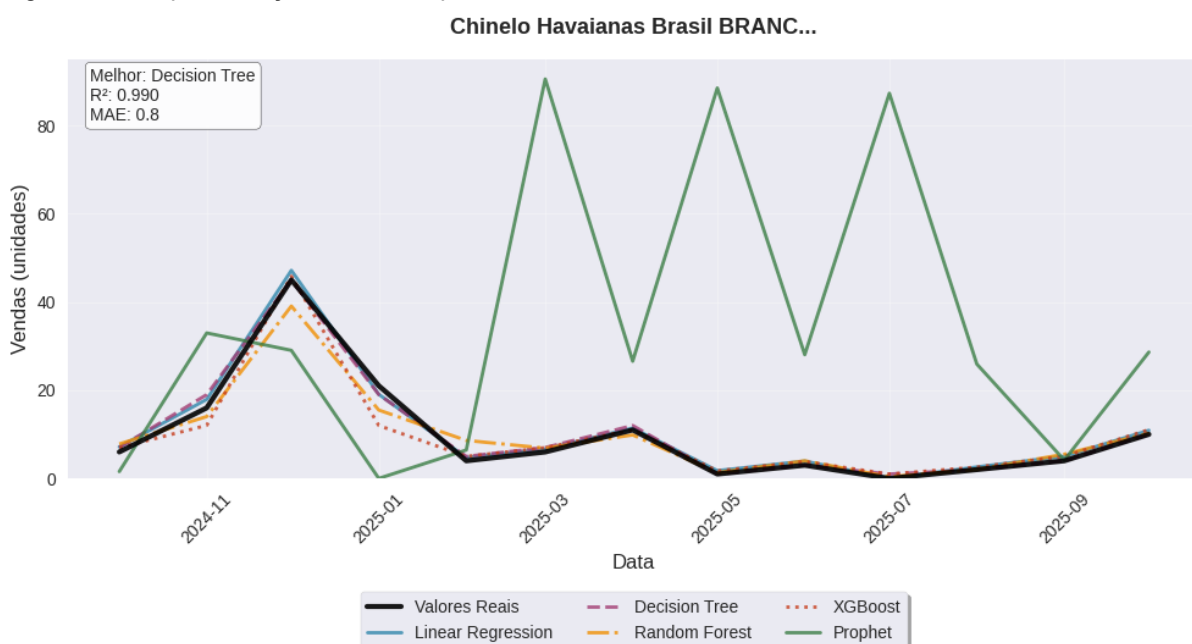
Figura 10 - Representação visual da performance do Chinelo Havaianas Top



Fonte: Elaborado pelo autor (2025)

De acordo com a Figura 10, a performance do Chinelo Havaianas Top Branco que apresentou um R² de 0,936 e MAE de 1 unidade, indicando um ótimo desempenho. Seguido por XGBoost que manteve um R² acima de 0,909 e a Floresta Aleatória com 0,876. É possível ainda notar que o Prophet obteve resultados ruins, com as previsões distantes dos dados reais.

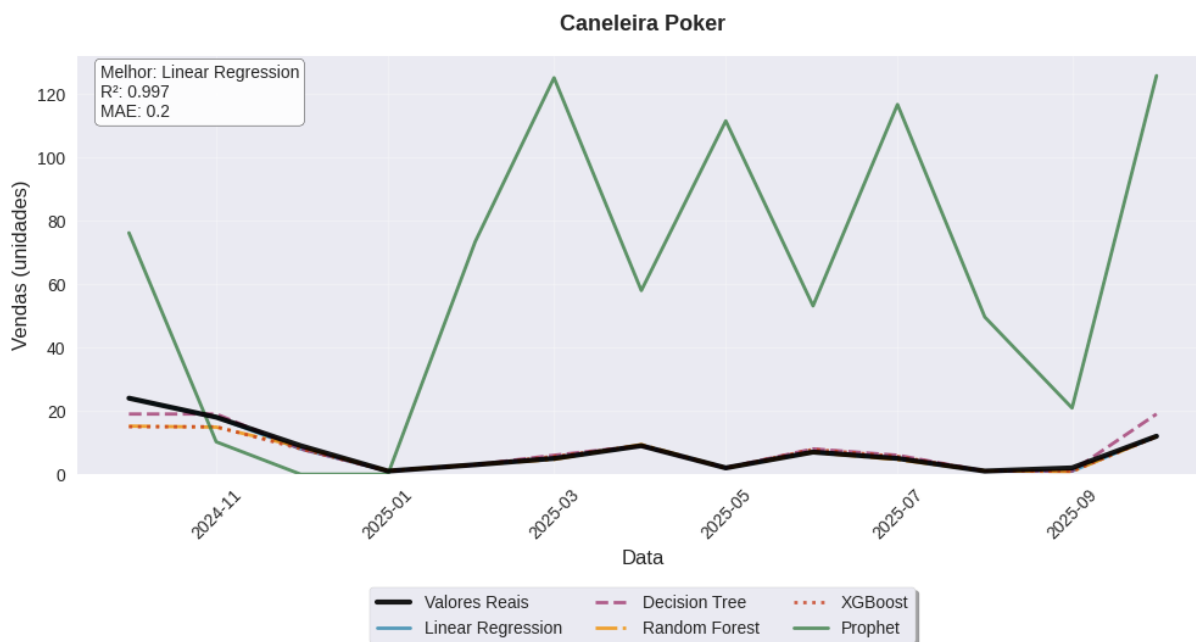
Figura 11 - Representação visual da performance do Chinelo Havaianas Brasil



Fonte: Elaborado pelo autor (2025)

A Figura 11 demonstra a performance do Chinelo havaianas Brasil Branco, no qual apresenta uma diferença em relação aos outros. O modelo que se adaptou melhor foi a Árvore de Decisão que capturou de forma mais eficaz as relações e estruturas nos dados, tendo um R^2 de 0,990 e MAE de 0,8 representando um erro médio mínimo nas previsões.

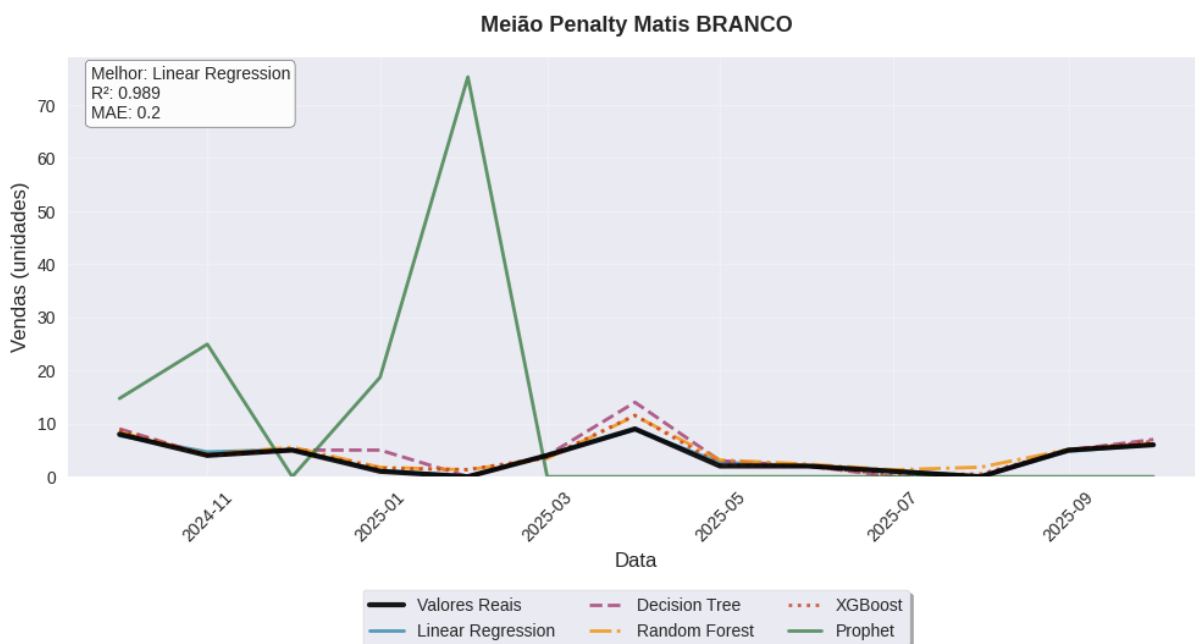
Figura 12 - Representação visual da performance da Caneleira Poker



Fonte: Elaborado pelo autor (2025)

Conforme a Figura 12, o gráfico representativo da Caneleira Poker teve um desempenho excepcional com a Regressão Linear com métricas de R^2 sendo 0,997 indicando que o modelo explica 99,7% da variação nos dados de vendas, e MAE de apenas 0,2 que reflete uma precisão quase perfeita nas previsões.

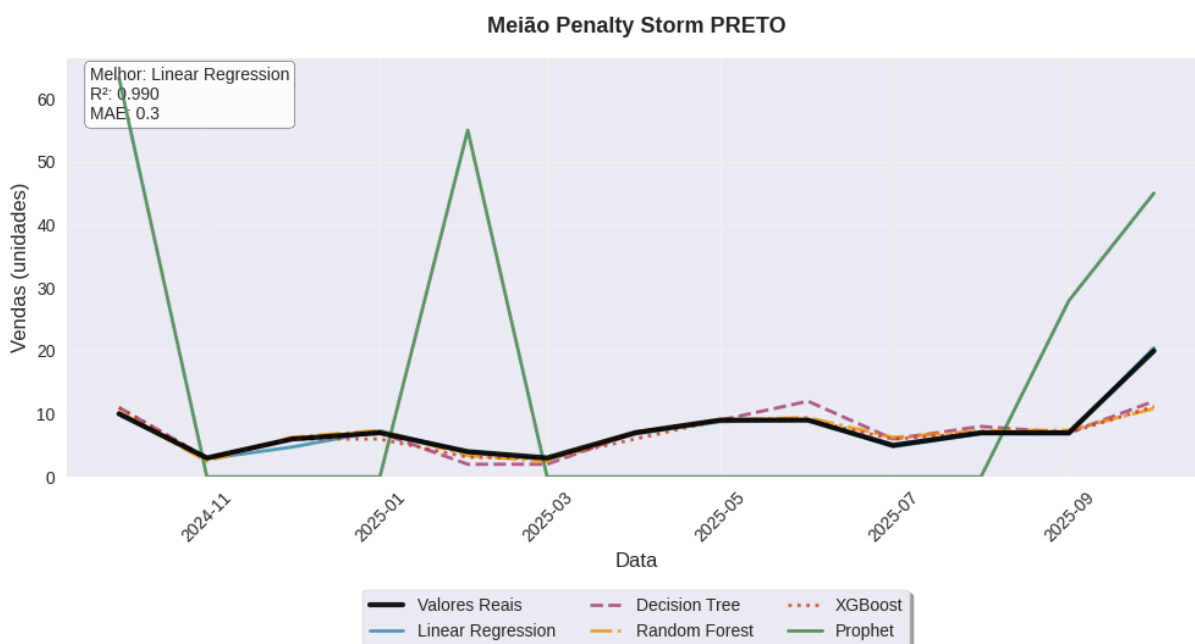
Figura 13 - Representação visual da performance do Meião Penalty Matis Branco



Fonte: Elaborado pelo autor (2025)

De acordo com a Figura 13, o produto Meião Penalty Matis Branco apresenta resultados notáveis com a Regressão Linear, tendo um R^2 de 0,989 e MAE de 0,2 demonstrando o comportamento das vendas de forma linear, estável e previsível ao longo do tempo, permitindo projeções futuras assertivas.

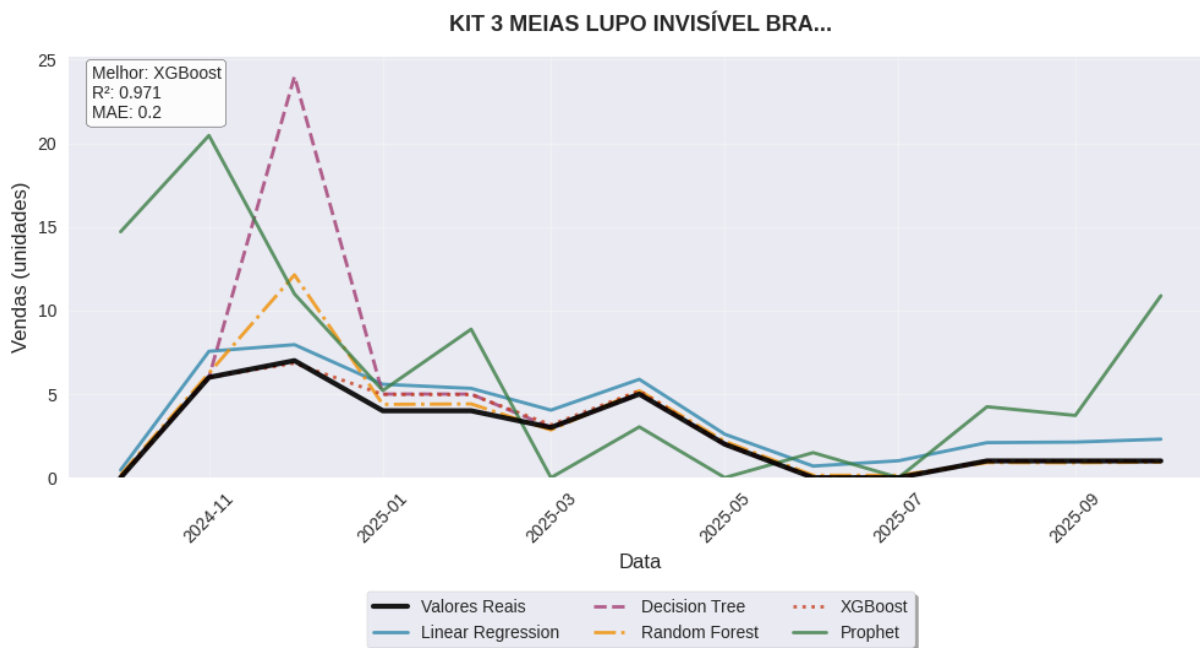
Figura 14 - Representação visual da performance do Meião Penalty Storm Preto



Fonte: Elaborado pelo autor (2025)

A Figura 14 através da representação visual da predição do Meião Penalty Storm Preto nota-se uma performance superior da Regressão Linear demonstrando R^2 de 0,990 e MAE de 0,3. Com esses resultados, sugere-se que a série temporal não apresenta sazonalidades marcantes, mudanças bruscas ou relações não-lineares significativas que justifiquem o uso de algoritmos mais sofisticados.

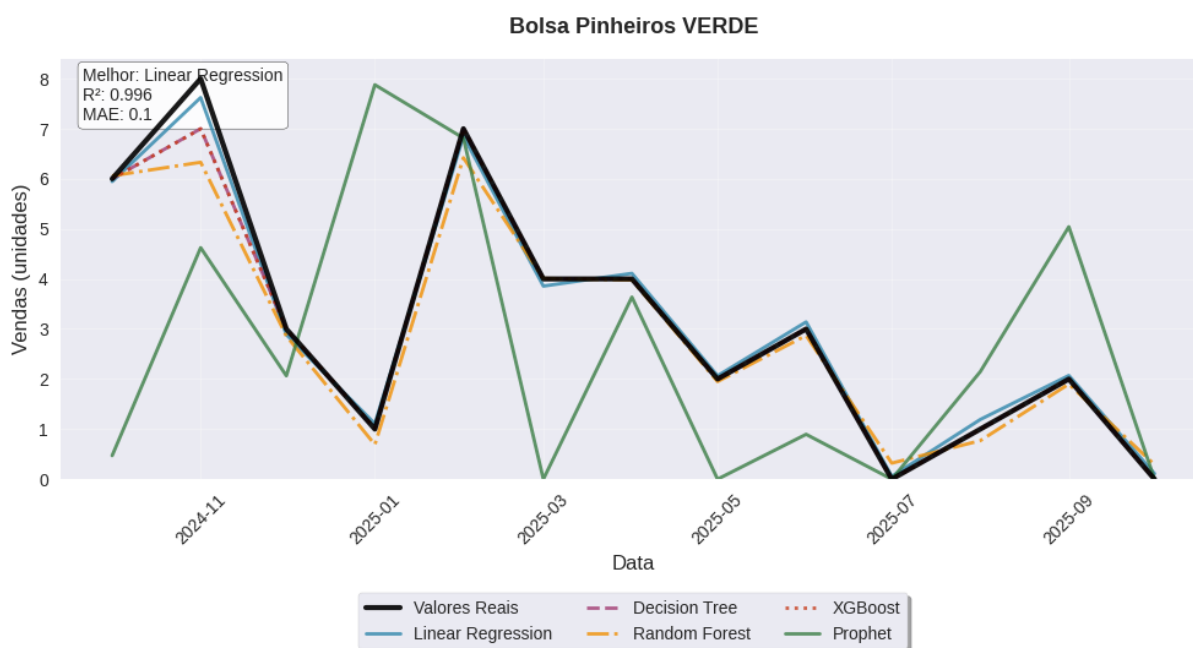
Figura 15 - Representação visual da performance do Kit 3 Meias Lupo Invisível



Fonte: Elaborado pelo autor (2025)

A Figura 15 apresenta uma análise preditiva de venda do Kit 3 Meias Lupo Invisível Branco através dos modelos selecionados, tendo como o XGBoost selecionado como o melhor algoritmo, superando os outros. Seu desempenho com R^2 de 0,971 e MAE de 0,2 demonstrando uma precisão alta e erros mínimos, sugerindo que o padrão de vendas deste produto possui sutilezas, relações não-lineares ou interações entre variáveis que não são capturadas de forma adequada por modelos lineares.

Figura 16 - Representação visual da performance da Bolsa Pinheiros Verde



Fonte: Elaborado pelo autor (2025)

Por fim, temos a Figura 16 apresentando a Bolsa Pinheiros Verde na sua performance com os modelos de ML, tendo como resultado a Regressão Linear que alcançou um R^2 de 0,996 e MAE de 0,7 reforçando a confiabilidade do algoritmo, mostrando que suas previsões desviam, em média, menos de uma unidade dos valores reais.

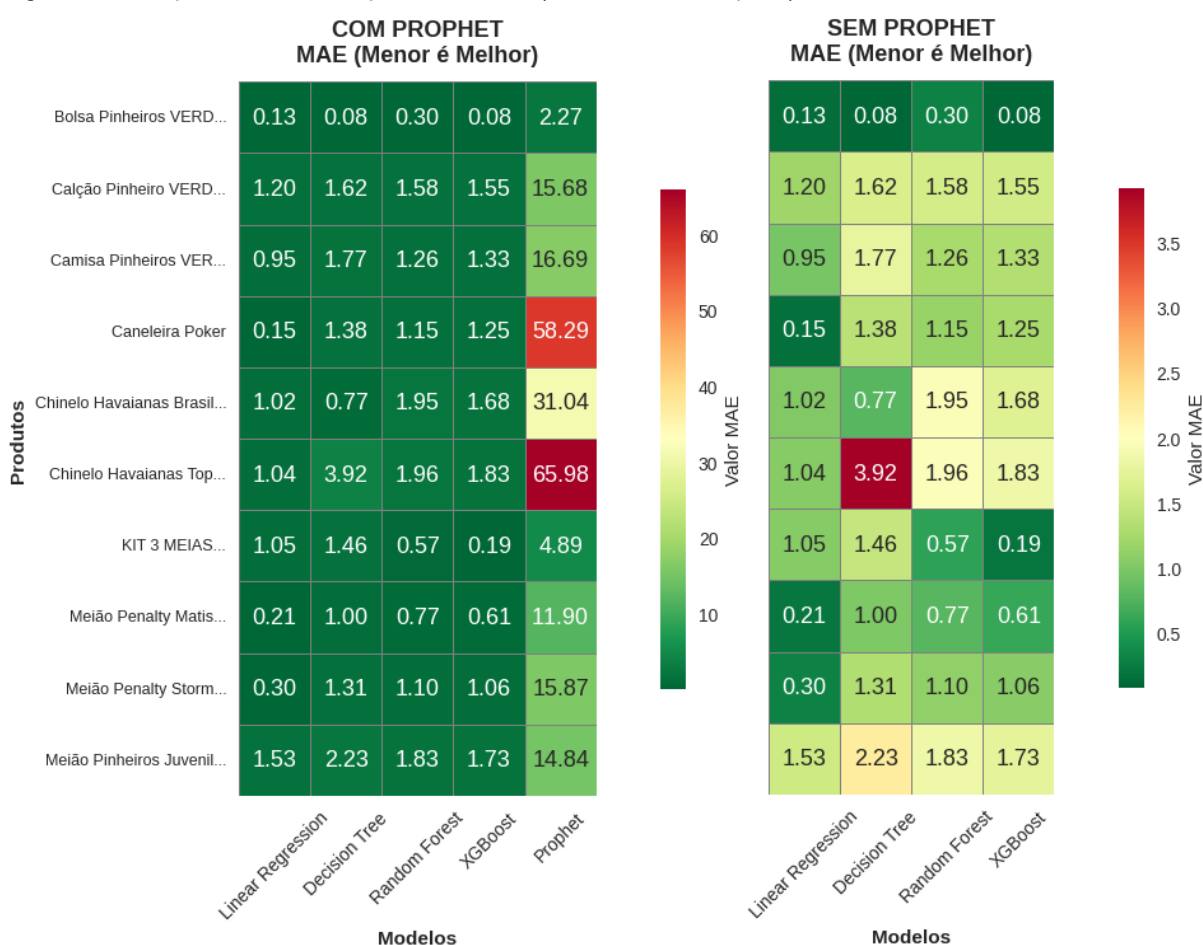
O Prophet, especializado em séries temporais, apresentou desempenho inadequado para o contexto analisado. Em todos os produtos, sem exceção, obteve valores negativos de R^2 , indicando que suas previsões são inferiores à simples média histórica. Casos extremos incluem a Caneleira Poker (R^2 : -113,250) e o Chinelo Havaianas Top Branco (R^2 : -77,004), onde os erros foram inaceitáveis (MAE de 58,29 e 65,98 unidades respectivamente). Isso sugere incompatibilidade entre as premissas do Prophet e as características dos dados limitados de pequeno varejo analisado.

4.2.2 Comparativo das métricas por modelos

Para uma avaliação abrangente e objetiva do desempenho preditivo dos algoritmos testados, realizou-se uma análise comparativa multidimensional baseada em três métricas fundamentais: o Coeficiente de Determinação (R^2), o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE). Esta abordagem permitiu

não apenas quantificar a capacidade explicativa de cada modelo, mas também dimensionar a magnitude e distribuição dos erros de previsão, proporcionando uma visão estratégica de sua eficácia operacional no contexto específico do pequeno varejo esportivo. As Figuras 17, 18 e 19 apresentam mapas de calor comparativos, Com e Sem Prophet e as métricas de cada um dos modelos de ML para MAE, RMSE e R^2 , respectivamente. Por questões de visualização, um mapa de calor sem o Prophet também é disponibilizado devido à distorção que este causa por seus valores destoantes dos valores obtidos por outros modelos de ML.

Figura 17 - Mapa de calor comparativo MAE (Com e Sem Prophet)



Fonte: Elaborado pelo autor (2025)

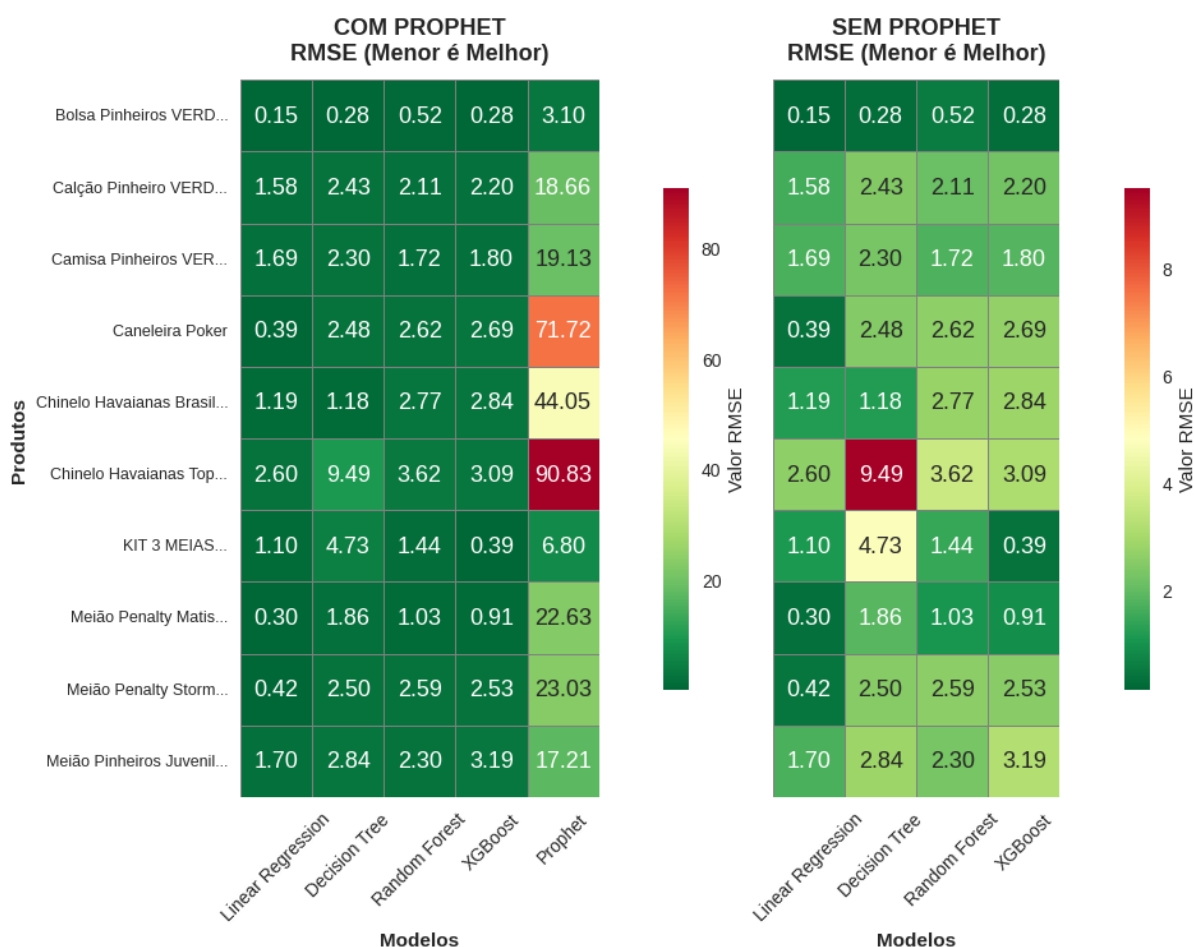
Conforme demonstra a Figura 17, os mapas de calor apresentados oferecem uma análise comparativa entre os cinco modelos preditivos, Regressão Linear, Árvore de Decisão, Random Forest, XGBoost e Prophet, avaliando seu desempenho através do MAE (Erro Absoluto Médio) na previsão de demanda de dez produtos distintos. A visualização paralela destes mapas revela *insights* sobre a adequação dos diferentes

algoritmos ao contexto específico dos dados analisados.

Observa-se primeiramente o mapa com Prophet, onde é apresentado valores de MAE superiores aos demais modelos para todos os produtos analisados. A performance inferior do Prophet pode ser explicada por várias hipóteses. Primeiro, o Prophet foi originalmente desenvolvido para séries temporais com fortes padrões sazonais e feriados, se estes dados não possuem tais características, o modelo pode estar adicionando complexidade desnecessária. Segundo, o Prophet assume automaticamente pontos de mudança que podem não existir na realidade destes produtos. Terceiro, a escala dos dados pode estar afetando o modelo, especialmente se houver grande variação entre produtos. Isso explica o comportamento anômalo que é demonstrado nos outros mapas de calor mostrados abaixo.

O resultado da métrica MAE, demonstra que a Regressão Linear teve superioridade em 7 produtos, a Árvore de Decisão teve uma vantagem em 2 e o XGBoost em 1 dos 10 produtos analisados.

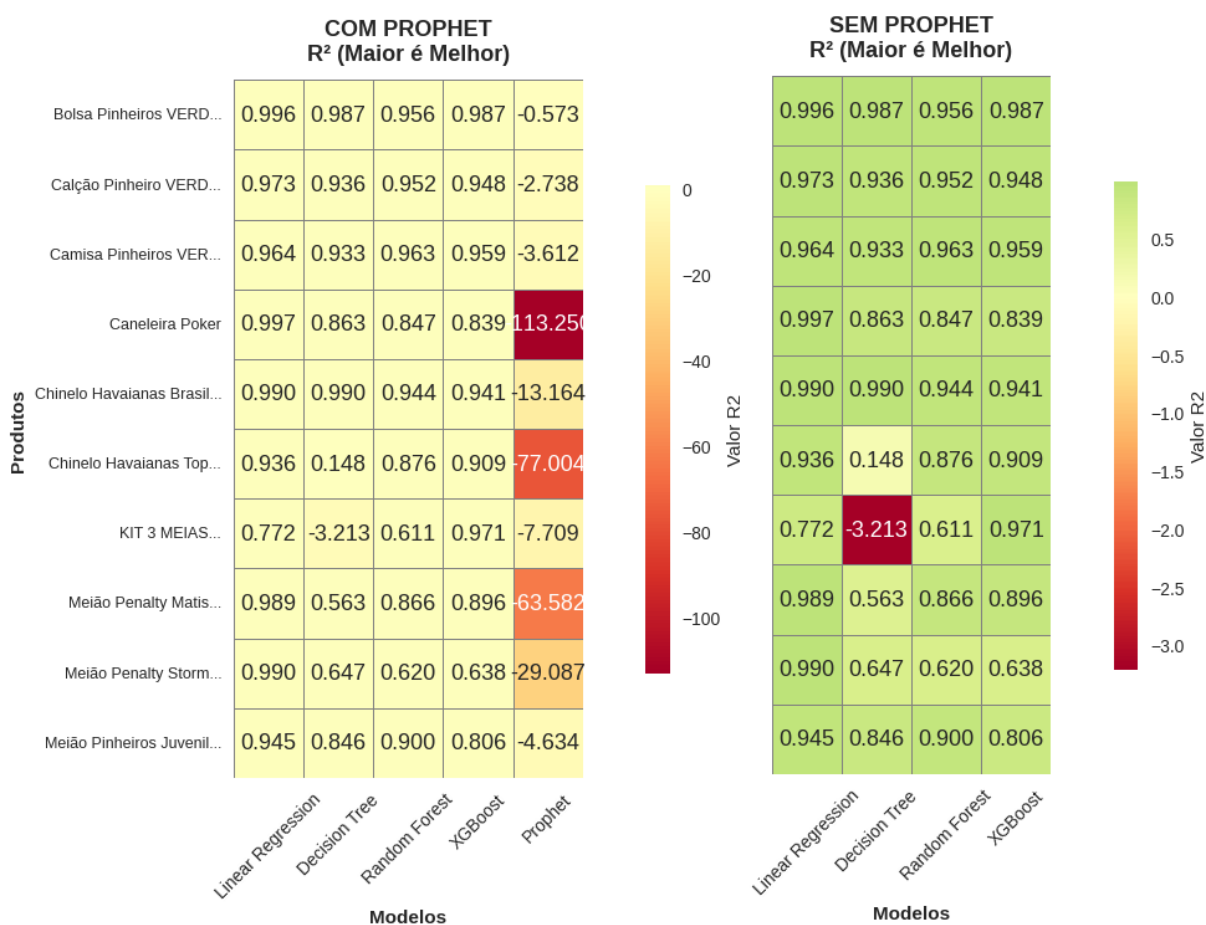
Figura 18 - Mapa de calor comparativo RMSE (Com e Sem Prophet)



Fonte: Elaborado pelo autor (2025)

A Figura 18, com base na análise visual do mapa de calor para RMSE, confirma padrões para a escolha do modelo, tendo como predominância a Regressão Linear em 8 produtos, Árvore de Decisão com um desempenho melhor em 1 e, por fim, XGBoost em 1 dos 10 produtos.

Figura 19 - Mapa de calor comparativo R² (Com e Sem Prophet)



Fonte: Elaborado pelo autor (2025)

Para complementar as análises dos mapas de calor, temos o R² conforme apresentado na Figura 19, que a Regressão Linear estabelece como modelo escolhido em 8 produtos, Árvore de Decisão sendo superior em 1 e o XGBoost com um desempenho notável em 1 dos 10 produtos analisados.

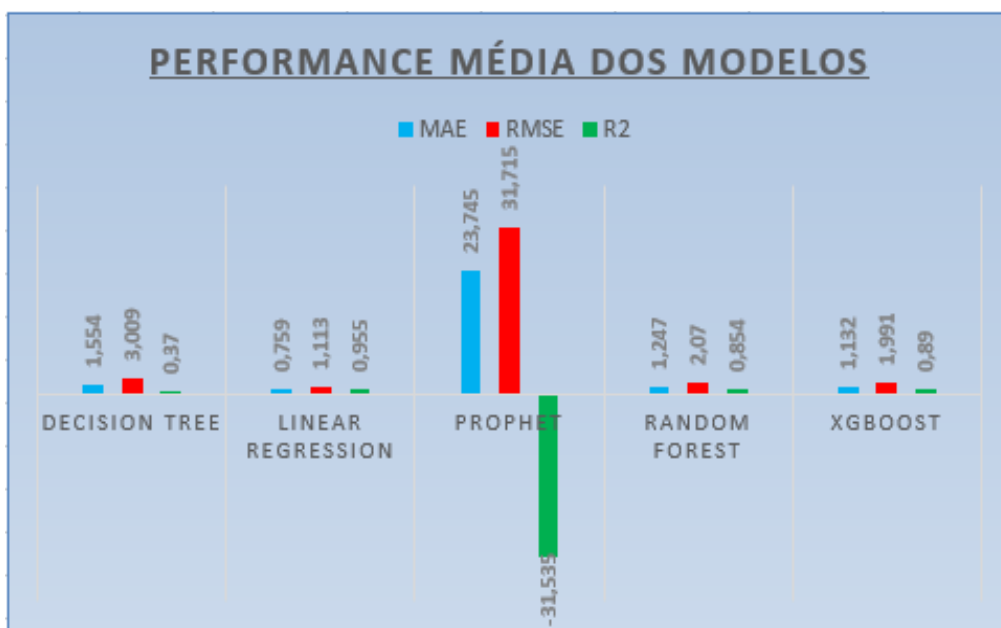
As Figuras 17, 18 e 19 representam uma evidência de que para o contexto de pequeno varejo com dados limitados, modelos mais simples e regularizados superaram algoritmos complexos, oferecendo maior confiabilidade para decisões de

gestão de estoque.

4.3 Avaliação dos modelos de ML

A avaliação dos cinco modelos de ML fundamentou-se no cálculo das médias das três métricas fundamentais sendo R^2 , MAE e RMSE, considerando o desempenho agregado nos dez produtos analisados durante o período de teste. Esta abordagem de consolidação métrica permite uma visão geral da performance preditiva de cada algoritmo, superando análises individuais por produto para identificar padrões gerais de eficácia e poder preditivo. A Figura 20 exibe uma comparação dos modelos de ML para cada uma das métricas utilizadas, demonstrando sua performance média.

Figura 20 - Performance média dos modelos



Fonte: Elaborado pelo autor (2025)

Como demonstrado na Figura 20, a Regressão Linear apresenta performance excelente com R^2 de 0,955, indicando que explica 95,5% da variabilidade da demanda, combinado com MAE de apenas 0,759 unidades (erro inferior a uma unidade por produto mensalmente) e um RMSE de 1,113 unidades (indicando um erro de aproximadamente 1 unidade por previsão). Além disso, este desempenho também a consolida como opção de melhor custo-benefício, considerando sua simplicidade interpretativa e baixa demanda computacional.

Os métodos *ensemble* XGBoost (R^2 : 0,890) e Floresta Aleatória (R^2 : 0,854) validam sua eficácia em contextos mais complexos. A Árvore de Decisão (R^2 : 0,370) demonstra alta volatilidade típica de modelos não-regularizados, enquanto o Prophet (R^2 : -31,535) evidência incompatibilidade com dados limitados e de granularidade mensal do pequeno varejo.

O Quadro 3 mostra o melhor desempenho do modelo de ML por produto.

Quadro 3 - Melhor modelo de ML por produto

Produto	Melhor Modelo	R² Teste	MAE Teste	RMSE Teste
Meião Pinheiros Juvenil Verde	Regressão Linear	0,945	1,528	1,699
Camisa Pinheiros Verde	Regressão Linear	0,964	0,955	1,693
Calção Pinheiros Verde	Regressão Linear	0,973	1,203	1,583
Chinelo Havaianas Top Branco	Regressão Linear	0,936	1,043	2,596
Chinelo Havaianas Brasil Branco	Árvore de Decisão	0,990	0,769	1,177
Caneleira Poker	Regressão Linear	0,997	0,154	0,392
Meião Penalty Matís Branco	Regressão Linear	0,989	0,208	0,301
Meião Penalty Storm Preto	Regressão Linear	0,990	0,300	0,423
Kit 3 Meias Lupo Invisível Branco	XGBoost	0,971	0,190	0,390
Bolsa Pinheiros Verde	Regressão Linear	0,996	0,130	0,154

Fonte: Elaborado pelo autor (2025)

Conforme o Quadro 3 demonstra o melhor desempenho do modelo de ML por produto revelou que a Regressão Linear foi selecionada como melhor modelo para oito dos dez itens, com performances notáveis em produtos de demanda estável como Caneleira Poker ($R^2 = 0,997$) e Bolsa Pinheiros Verde ($R^2 = 0,996$). Casos específicos como o Chinelo Havaianas Brasil Branco, onde a Árvore de Decisão obteve vantagem ($R^2 = 0,990$), e o Kit 3 Meias Lupo Invisível Branco, onde o XGBoost superou os demais ($R^2 = 0,971$), demonstraram a necessidade de seleção individualizada por produto baseada em suas características de demanda. A variação do R^2 entre 0,936 e 0,997 refletiu distintos níveis de previsibilidade no portfólio, com produtos de maior volume geralmente apresentando melhor performance preditiva. A avaliação confirmou assim a seleção da Regressão Linear como modelo primário para o sistema de previsão, reservando XGBoost para casos específicos com relações não-lineares pronunciadas, assegurando assim equilíbrio entre performance preditiva, eficiência operacional e sustentabilidade computacional.

4.4 Validação técnica e operacional

A validação da viabilidade do sistema de previsão foi conduzida através de múltiplas dimensões, confirmando sua aplicabilidade prática para a previsão de demanda e, conseqüentemente, a gestão de estoques. A análise técnica demonstrou robustez operacional, com tempo médio de processamento inferior a 2 minutos para todos os produtos utilizando recursos computacionais padrão. A arquitetura baseada em planilhas CSV como interface de entrada mostrou-se adequada ao contexto operacional, permitindo atualizações manuais mensais sem necessidade de integração complexa com sistemas ERP.

Conforme o Quadro 4 evidencia *insights* para gestão de estoque gerados à partir dos modelos de ML.

Quadro 4 - *Insights* de gestão de estoque

Produto	Estoque de Segurança	Ponto de Reposição	Lote Econômico
Meião Pinheiros Juvenil Verde	11	36	53

Produto	Estoque de Segurança	Ponto de Reposição	Lote Econômico
Camisa Pinheiros Verde	13	35	47
Calção Pinheiros Verde	14	35	46
Chinelo Havaianas Top Branco	15	24	19
Chinelo Havaianas Brasil Branco	18	28	22
Caneleira Poker	10	18	16
Meião Penalty Matís Branco	5	9	8
Meião Penalty Storm Preto	6	13	16
Kit 3 Meias Lupo Invisível Branco	5	8	6
Bolsa Pinheiros Verde	5	8	7

Fonte: Elaborado pelo autor (2025)

O Quadro 4 demonstra que economicamente, os modelos de ML apresentaram potencial significativo de otimização de custos. Considerando os estoques de segurança calculados de 5 a 18 unidades por produto, estima-se redução nos custos de armazenagem e nas perdas por obsolescência. O ponto de reposição definido para cada produto foi de 8 a 36 unidades que assegura equilíbrio entre disponibilidade comercial e custo de capital imobilizado.

Operacionalmente, o fluxo estabelecido mostrou-se compatível com a realidade da empresa: as planilhas mensais servem como interface única, com validações automáticas que previnem inconsistências nos dados. O tempo total do processo, desde a importação até a geração das novas previsões, mantém-se abaixo de 2 minutos, compatível com a dinâmica decisória mensal de reabastecimento dos produtos.

A validação empírica no período de teste (Outubro/2024 a Outubro/2025) confirmou a confiabilidade dos algoritmos, com MAE entre 0,13 e 1,53 unidades

demonstrados no Quadro 3, margem compatível com as necessidades de precisão para decisões de reposição no varejo esportivo. A consistência dos modelos através de diferentes produtos e períodos sazonais reforça sua aplicabilidade em cenários reais.

A viabilidade organizacional foi atestada pela simplicidade de interpretação dos resultados e pela operacionalização via planilhas padronizadas, facilitando a adoção desse sistema de abastecimento de dados em determinados períodos pela loja sem necessidade de treinamento especializado extensivo ou alterações nos processos existentes.

Portanto, a validação confirmou que a metodologia implementada não apenas é tecnicamente viável com modelos capazes de prever demanda com alta performance preditiva mas também operacionalmente relevante, gerando *insights* que podem ser diretamente incorporados aos processos de gestão de estoque da empresa.

4.5 Discussões

A análise dos dados disponíveis confirma a progressiva integração da Inteligência Artificial (IA) e do Machine Learning (ML) no cenário corporativo contemporâneo. Conforme apontado por pesquisas recentes, aproximadamente 72% das empresas em âmbito global já incorporam soluções baseadas em IA (CNN, 2024; Adobe Business, 2025). Esse movimento corrobora as expectativas traçadas por autores fundamentais da área, como Russell e Norvig (2021) e Goodfellow (2016), sinalizando uma transição dessas tecnologias do campo teórico para aplicações práticas que impactam diretamente a produtividade, a inovação e a eficiência das organizações. Neste contexto, a presente pesquisa demonstra como essa transição ocorre na prática do pequeno varejo, aplicando cinco modelos de ML à previsão de demanda de dez produtos de alto giro, conforme detalhado na caracterização dos dados.

No entanto, a revisão da literatura especializada revela que, paralelamente aos benefícios tangíveis como automação de processos, redução de custos e aumento de competitividade, a implementação dessas tecnologias apresenta desafios significativos de ordem ética, social e técnica. Russell (2019) destaca que sistemas cujos objetivos não são claramente definidos podem gerar consequências não

intencionais e indesejadas. Esta advertência materializa-se em problemas concretos, como a propagação de vieses algorítmicos, riscos à privacidade e potenciais impactos no emprego. No âmbito desta pesquisa, os desafios técnicos manifestaram-se claramente na seleção e adequação dos algoritmos, especialmente no caso do Prophet, que apresentou desempenho inadequado com valores negativos de R^2 em todos os produtos analisados (Quadro 2), evidenciando que a mera adoção de tecnologias complexas não garante resultados positivos.

Os achados evidenciam, portanto, que a aplicação da IA e do ML não pode ser conduzida de maneira isolada ou puramente técnica. Ela demanda um acompanhamento pautado por regulamentação apropriada, supervisão contínua e uma reflexão crítica constante. A sinergia entre IA e ML ilustra, por um lado, um notável potencial de inovação, e por outro, a urgência de estruturas de governança e uso responsável. Sua aplicação em domínios como tradução automática, análise financeira e educação adaptativa demonstra que os fundamentos teóricos estabelecidos por autores como Carbonell, Michalski e Mitchell (1983) encontram hoje ampla materialização em soluções de mercado. Esta pesquisa contribui para esse ecossistema de aplicações ao materializar os conceitos de ML na gestão de estoque do varejo esportivo, onde a análise comparativa dos modelos e sua subsequente avaliação seguem princípios de governança técnica através de validação rigorosa, métricas múltiplas e períodos de teste separados.

Portanto, a discussão reforça que, embora o potencial transformador dessas tecnologias seja inegável, seu desenvolvimento e implantação devem estar intrinsecamente associados a diretrizes éticas robustas e a políticas regulatórias claras. De acordo com Russell (2019), o sucesso da IA e do ML depende não apenas da excelência técnica, mas também de uma compreensão profunda de suas implicações sociais e econômicas. Os resultados desta pesquisa validam essa premissa ao demonstrar que, no contexto específico do pequeno varejo, o sucesso da implementação de ML para previsão de demanda não reside na adoção do algoritmo mais sofisticado, mas na seleção criteriosa baseada em evidências empíricas (Regressão Linear como melhor modelo em 8 dos 10 produtos), na validação técnica rigorosa (período de teste de 13 meses) e na geração de valor operacional tangível (recomendações de estoque do Quadro 4). Desse modo, fica evidente que a adoção responsável é condição indispensável para maximizar os benefícios, minimizar os riscos e promover avanços que sejam, simultaneamente, inovadores e socialmente

sustentáveis. A validação técnica e operacional apresentada corrobora essa conclusão, mostrando que a metodologia implementada é não apenas tecnicamente viável, mas também operacionalmente relevante para a realidade da empresa.

5 CONCLUSÃO

O presente trabalho atingiu com êxito seu objetivo geral de implementar e avaliar modelos de ML para previsão de demanda em uma loja de esportes de pequeno porte. A partir dos dados históricos de vendas entre Janeiro/2022 a Outubro/2025, foram aplicadas técnicas de engenharia de *features* temporais e implementados cinco algoritmos de ML distintos, permitindo uma avaliação comparativa robusta no período de teste de Outubro/2024 a Outubro/2025. A abordagem metodológica adotada permitiu validar não apenas a eficácia preditiva dos modelos, mas também sua aplicabilidade prática no contexto real do pequeno varejo.

Quanto aos objetivos específicos propostos, todos foram integralmente atendidos. A coleta e análise de dados históricos, embora realizada de forma manual devido às limitações do sistema ERP da empresa, forneceu uma base consistente para modelagem, com destaque para a identificação dos dez produtos de maior giro que representam o núcleo do negócio. A seleção dos algoritmos de ML abrangeu desde abordagens lineares simples até técnicas avançadas de *ensemble learning*, permitindo testar diferentes paradigmas preditivos. A avaliação sistemática utilizando as métricas de R^2 , MAE e RMSE demonstrou a superioridade consistente da Regressão Linear, que apresentou R^2 médio de 0,955 e MAE de 0,759 unidades, indicando precisão operacional adequada para decisões de estoque. Finalmente, a análise de viabilidade operacional confirmou que a solução proposta é tecnicamente implementável e economicamente vantajosa, com recomendações específicas de políticas de estoque geradas para cada produto.

Os resultados obtidos revelaram *insights* significativos sobre a dinâmica da demanda no pequeno varejo esportivo. A performance excelente da Regressão Linear em oito dos dez produtos analisados desafia a premissa comum de que modelos mais complexos oferecem melhor performance. Este achado sugere que, para séries temporais limitadas e com padrões relativamente estáveis como as observadas, a simplicidade algorítmica pode ser mais vantajosa do que a complexidade computacional. Em contraste, o desempenho deficiente do Prophet, com valores negativos de R^2 em todos os produtos, indica incompatibilidade entre as premissas do modelo e as características dos dados disponíveis. Especializado em capturar sazonalidades complexas e tendências não-lineares, o Prophet sofreu de *overfitting* devido à quantidade insuficiente de dados sendo de apenas 46 meses e à

granularidade mensal, que pode não capturar padrões sazonais mais sutis. Esta constatação é valiosa para contextos similares de pequenas empresas, sugerindo que investimentos em soluções especializadas podem não se justificar quando comparados a abordagens mais simples e interpretáveis.

A pesquisa não está isenta de limitações, as quais devem ser reconhecidas para contextualizar adequadamente os resultados. A coleta manual de dados, embora necessária dadas as restrições operacionais da empresa, introduz riscos de erro humano e limita a escalabilidade da solução. O processo de exportação para Excel e posterior conversão para CSV, apesar de funcional, representa um gargalo operacional que precisaria ser otimizado em uma implementação em produção. A restrição aos dez produtos mais vendidos, embora estrategicamente justificada, limita a generalização dos resultados para itens de menor giro que podem apresentar padrões de demanda distintos. Além disso, o período histórico de 46 meses, embora representativo, pode ser insuficiente para capturar ciclos econômicos mais longos ou mudanças estruturais no comportamento do consumidor.

Para trabalhos futuros, várias direções se apresentam como promissoras. A integração direta com sistemas ERP mais robustos permitiria automatização completa do fluxo de dados, eliminando etapas manuais e aumentando a confiabilidade do processo. A expansão da análise para todo o portfólio de produtos, incluindo itens de baixo giro, testaria a generalização da abordagem proposta. A incorporação de variáveis externas, como dados climáticos, eventos esportivos locais ou indicadores econômicos regionais, poderia melhorar a capacidade preditiva, especialmente para produtos sensíveis a fatores contextuais. A implementação de um sistema de monitoramento contínuo com recalibração periódica dos modelos garantiria a manutenção da precisão ao longo do tempo, adaptando-se a mudanças nos padrões de demanda. Finalmente, a aplicação da mesma metodologia em empresas similares permitiria validar sua generalização para diferentes contextos do pequeno varejo.

Deste modo, este trabalho demonstra que a aplicação de técnicas de ML para previsão de demanda em pequenas empresas de varejo esportivo é viável e potencialmente transformadora. A solução proposta equilibra a sofisticação técnica com praticidade operacional, oferecendo um caminho acessível para a otimização de estoques que tradicionalmente é dominado por grandes varejistas com recursos tecnológicos substanciais. Os resultados não apenas validam a abordagem metodológica adotada, mas também contribuem para a literatura sobre aplicação

prática de ML em contextos de recursos limitados, destacando que, em muitos casos, a simplicidade bem executada supera a complexidade mal calibrada. A pesquisa deixa como legado um *framework* replicável que pode ser adaptado por outras pequenas empresas buscando transformar dados históricos em vantagem competitiva sustentável.

REFERÊNCIAS

- ACKERMAN, K. **350 dicas para gerenciar seu armazém: almoxarifado, depósito, centro de distribuição**. São Paulo: Imam, 2004.
- ADOBE. **AI vs. Machine Learning: What's the Difference?** Adobe Business, 2025. Disponível em: <https://business.adobe.com/br/products/real-time-customer-data-platform/ai-vs-machine-learning.html>. Acesso em: 04 dez. 2025.
- ANGRIST, J. D.; PISCHKE, J.S. **Mostly harmless econometrics: an empiricist's companion**. Princeton: Princeton University Press, 2009.
- ARNOLD, J. R. T. **Administração de materiais: uma introdução**. São Paulo: Atlas, 1999.
- BALLOU, R. H. **Gerenciamento da cadeia de suprimentos**. 5. ed. Porto Alegre: Bookman, 2006.
- BALLOU, R. H. **Logística empresarial: transportes, administração de materiais e distribuição física**. 5. ed. São Paulo: Atlas, 2019.
- BANZATO, E. **Tecnologia da informação aplicada à logística**. São Paulo: Imam, 2005.
- BENTÉJAC, C.; CSÖRGŐ, A.; MARTÍNEZ-MUÑOZ, G. A comparative analysis of gradient boosting algorithms. **Artificial Intelligence Review**, v. 54, n. 3, p. 1937-1967, 2020.
- BOWERSOX, D. J.; CLOSS, D. J.; COOPER, M. B. **Gestão logística de cadeias de suprimentos**. Tradução de Camila Teixeira Nakagawa, Gabriela Teixeira Nakagawa. Porto Alegre: Bookman, 2006.
- BOWERSOX, D. J.; CLOSS, D. J. **Gestão logística da cadeia de suprimentos**. 4. ed. Porto Alegre: AMGH, 2014.
- CARBONELL, J. G.; MICHALSKI, R. S.; MITCHELL, T. M. Machine Learning: A Historical and Methodological Analysis. **AI Magazine**, v. 4, n. 3, p. 69-79, 1983. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/406>. Acesso em: 04 dez. 2025.
- CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016.
- CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. **Semi-supervised learning**. Cambridge: MIT Press, 2006.

- CHASE, C. W. **Demand driven forecasting: a structured approach to forecasting**. 2. ed. Hoboken: Wiley, 2013.
- CHEN, T.; GUESTRIN, C. XGBoost: a scalable tree boosting system. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, San Francisco. **Proceedings...** New York: ACM, 2016. p. 785-794.
- CHIAVENATO, I. **Introdução à teoria geral da administração**. 7. ed. Rio de Janeiro: Elsevier, 2003.
- CNN BRASIL. Uso de Inteligência Artificial aumenta e alcança 72% das empresas, diz pesquisa. **CNN Brasil**, São Paulo, 9 jun. 2024. Disponível em: <https://www.cnnbrasil.com.br/economia/negocios/uso-de-inteligencia-artificial-aumenta-e-alcanca-72-das-empresas-diz-pesquisa/>. Acesso em: 04 dez. 2025.
- CORRÊA, A.; VIEIRA, S. Pré-processamento de dados para análise de séries temporais na previsão de valores de criptomoedas, utilizando ARIMA e PROPHET. **Tecnologia em Gestão de Organizações Públicas**, [S. l.], 2021. Disponível em: https://www.tfgonline.lapinf.ufn.edu.br/media/midias/Adriano_Correa.pdf. Acesso em: 24 nov. 2025.
- CORRÊA, H. L.; CORRÊA, C. A. **Administração de produção e operações: manufatura e serviços: uma abordagem estratégica**. 2. ed. São Paulo: Atlas, 2012.
- DA VEIGA, C. P. et al. Previsão de demanda no varejo alimentício como ferramenta estratégica de sustentabilidade em uma pequena empresa brasileira. **Future Studies Research Journal: Trends and Strategies**, São Paulo, v. 5, n. 2, p. 113-133, jul./dez. 2013.
- DIAS, M. A. P. **Administração de materiais: uma abordagem logística**. 5. ed. São Paulo: Atlas, 2018.
- DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78-87, out. 2012.
- FACCHINI, E.; DA SILVA, J. R.; LEITE, V. M. Curva ABC e Estoque de Segurança. **South American Development Society Journal**, v. 5, n. 13, p. 73-91, 2019.
- FACELI, K. *et al.* **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2019.
- FERNANDES, F. C. F.; GODINHO F. M. **Planejamento e controle da produção: dos**

fundamentos ao essencial. São Paulo: Atlas, 2010.

FERNANDES, F.; ANZANELLO, M. J.; MIORANDO, R. F. Integração de métodos quantitativos e qualitativos para previsão de demanda no setor de autopeças. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 31., 2011, Belo Horizonte. **Anais...** Belo Horizonte: ABEPRO, 2011.

FONSECA, J. J. S. **Metodologia da pesquisa científica.** Fortaleza: UEC, 2002.

FONSECA, M. T. **Tecnologias gerenciais de restaurantes.** 10. ed. São Paulo: Senac, 2014.

FORBES. Amazon Web Services Contributor. (2021). Predicting the Future of Demand: How Amazon is reinventing forecasting with Machine learning. <https://www.forbes.com/sites/amazonwebservices/2021/12/03/predicting-the-future-of-demand-how-amazon-is-reinventing-forecasting-with-machine-learning/>. Acesso: 6 mai. 2025

GAITHER, N.; FRAZIER, G. **Administração da produção e operações.** 8. ed. São Paulo: Pioneira Thomson Learning, 2007.

GARCÍA-BARRIOS, D. A machine learning based method for managing multiple impulse purchase products: an inventory management approach. **Journal of Engineering Science and Technology Review**, v. 14, n. 6, p. 19-27, 2021.

GERHARDT, T. E.; SILVEIRA, D. T. (Org.). **Métodos de pesquisa.** Porto Alegre: Editora da UFRGS, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa.** 6. ed. São Paulo: Atlas, 2017.

GIL, A. C. **Métodos e técnicas de pesquisa social.** 6. ed. São Paulo: Atlas, 2008.

GONÇALVES, M. C. P. **The possibility of implementing intelligent systems and the respective impact of artificial intelligence on inventory management and warehousing.** 2022. Dissertação (Mestrado em Gestão de Serviços e Tecnologia) – ISCTE Business School, Instituto Universitário de Lisboa, Lisboa, 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning.** Cambridge: MIT Press, 2016.

GRANDO, J. **Modelo preditivo para análise de séries temporais financeiras baseado em aprendizado de máquina.** 2023. Dissertação (Mestrado em Computação Aplicada) – Universidade do Estado de Santa Catarina, Joinville, 2023. Disponível

em: https://www.udesc.br/arquivos/cct/id_cpmenu/1024/Disserta_o_Julia_Grando_16962088780995_1024.pdf. Acesso em: 24 nov. 2025.

- GRUS, J. **Data science do zero: primeiras regras com o Python**. Rio de Janeiro: Alta Books, 2018.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.
- ISLAM, M. K. *et al.* Role of artificial intelligence and machine learning in optimizing inventory management across global industrial manufacturing & supply chain: a multi-country review. **International Journal of Management Information Systems and Data Science**, v. 1, n. 2, p. 1-14, 2024.
- JOKELA, J. **Optimizing inventory management: leveraging CNNs and conventional grouping methods: a strategy for excess inventory reduction and dynamic grouping**. 2024. Dissertação (Mestrado em Ciência da Computação) – University of Vaasa, Vaasa, Finlândia, 2024.
- KRAJEWSKI, L. J.; RITZMAN, L. P.; MALHOTRA, M. K. **Administração de produção e operações**. 8. ed. São Paulo: Pearson Prentice Hall, 2009.
- LINGAM, Y. K. The role of Artificial Intelligence (AI) in making accurate stock decisions in E-commerce industry. **International Journal of Advanced Research Ideas and Innovations in Technology**, v. 4, n. 3, p. 2281-2286, 2018.
- LIPTON, Z. C. **The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery**. *Queue*, 16(3):31–57, 2018.
- LIU, M. *et al.* Intelligent monitoring method of tridimensional storage system based on deep learning. **Environmental Science and Pollution Research**, v. 29, n. 46, p. 70464-70478, 2022.
- MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.
- MARTINS, P. G.; LAUGENI, F. P. **Administração da produção**. 2. ed. São Paulo: Saraiva, 2005.
- MARTINS, P. G.; ALT, P. R. C. **Administração de materiais e recursos patrimoniais**. 3. ed. São Paulo: Saraiva, 2009.
- MITCHELL, T. M. **Machine learning**. New York: McGraw-Hill, 1997.
- MOREIRA, D. A. **Administração da produção e operações**. 2. ed. São Paulo: Cengage Learning, 2011.
- NETTSTRATER, A. *et al.* **Logistics software systems and functions: an overview of ERP, WMS, TMS and SCM systems**. Berlim: Springer, 2015.

- OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. Tese (Doutorado em Ciências) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.
- PASSARI, A. F. L. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais**. 2003. Tese (Doutorado em Engenharia) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2003.
- PASUPULETI, V. *et al.* Enhancing supply chain agility and sustainability through machine learning: optimization techniques for logistics and inventory management. **Logistics**, v. 8, n. 3, p. 73, 2024.
- PEINADO, J.; GRAEML, A. R. **Administração da produção: operações industriais e de serviços**. Curitiba: UnicenP, 2007.
- POZO, H. **Administração de recursos materiais e patrimoniais: uma abordagem logística**. 4. ed. São Paulo: Atlas, 2007.
- RUSSELL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. 3rd ed. Upper Saddle River: Prentice Hall, 2009.
- RUSSELL, S. **Human compatible: artificial intelligence and the problem of control**. New York: Viking, 2019.
- RUSSELL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. 4th ed. Harlow: Pearson, 2021.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 535-554, jul. 1959. Disponível em: <https://www.cs.virginia.edu/~evans/greatworks/samuel1959.pdf>. Acesso em: 13 mai. 2025.
- SANDRA, G. A revolução da moda da Zara por meio de insights de dados. **Medium**, [S. l.], 2022. Disponível em: <https://medium.com/thedeephub/zaras-fashion-revolution-through-data-insights-28f8fa728a0f>. Acesso em: 6 mai. 2025.
- SCIKIT-LEARN. **1.10. Decision Trees**. 2023. Disponível em: <https://scikit-learn.org/stable/modules/tree.html>. Acesso em: 28 mai. 2025.
- SEVERINO, A. J. **Metodologia do trabalho científico**. 24. ed. rev. e atual. São Paulo: Cortez, 2017.
- SLACK, N. *et al.* **Administração da produção**. 8. ed. São Paulo: Atlas, 2018.
- SPIRITO, C. **Artificial intelligence applications in reverse logistics, how technology could improve return and waste management creating value**. 2024.

- Tese (Doutorado em Gestão) – Politecnico di Torino, Torino, Itália, 2024.
- TANG, Y. M.; THAKUR, M.; CHAU, K. Y.; LAU, Y.; ZHENG, Z.; PATEL, P.; GUPTA, L. K.; KUMAR, M.; SATHISH K., A. S. Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation. *Applied Sciences*, v. 13, n. 5, p. 3051, 2023.
- TAULLI, T. **Introdução à inteligência artificial: uma abordagem não técnica**. São Paulo: Novatec, 2020.
- TAYLOR, S. J.; LETHAM, B. Forecasting at scale. *PeerJ Preprints*, 2017. Disponível em: <https://doi.org/10.7287/peerj.preprints.3190v2>. Acesso em: 24 mai. 2025.
- TUBINO, D. F. **Planejamento e controle da produção: teoria e prática**. 2. ed. São Paulo: Atlas, 2007.
- VAGO, F. R. M. *et al.* A importância do gerenciamento de estoque por meio da ferramenta curva ABC. *Revista Sociais e Humanas*, Santa Maria, v. 26, n. 3, p. 638-655, set./dez. 2013.
- WANG, F.; ROSS, C. L. Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transportation Research Record*, v. 2672, n. 47, p. 35-45, 2018.
- WREN, D. A. **The evolution of management thought**. 6th ed. Hoboken: Wiley, 2005.
- XU, Z. *et al.* Logistics supply chain network risk prediction model based on intelligent random forest model. *IEEE Transactions on Engineering Management*, 2023.
- ZHOU, Z. **Machine learning**. Singapore: Springer Nature, 2021.