

UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE COMPUTAÇÃO
ENGENHARIA DE COMPUTAÇÃO

Brainer Sueverti de Campos

**Análise Comparativa de Modelos de Previsão de
Demanda Aplicados a Múltiplas Bases e Níveis
de Granularidade no Contexto do Varejo**

São Carlos - SP

2025

Brainer Sueverti de Campos

**Análise Comparativa de Modelos de Previsão de Demanda
Aplicados a Múltiplas Bases e Níveis de Granularidade no
Contexto do Varejo**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Computação da Universidade Federal de São Carlos, como requisito para a obtenção do título de Bacharel em Engenharia de Computação.

Orientação Prof. Dr. Alan Demétrius Baria Valejo

São Carlos - SP

2025

Dedico este trabalho ao meu pai, à minha mãe, à minha noiva e aos meus queridos amigos

Agradecimentos

A todos que contribuíram para que eu chegasse até aqui, expresso minha mais profunda gratidão. Esta jornada foi marcada por desafios, aprendizados e muitas alegrias, e cada um de vocês desempenhou um papel especial em minha caminhada. Sem o apoio, a compreensão e a motivação de todos, este momento não seria possível.

Em especial, agradeço aos meus pais, Luis e Lucelena, meus maiores exemplos de dedicação. Vocês sempre acreditaram em mim e trabalharam muito para me proporcionar as melhores oportunidades. O amor e o incentivo de vocês me inspiraram a persistir e buscar sempre o melhor.

À minha companheira, Ana Carla, minha eterna gratidão. Seu apoio e amor incondicional foram essenciais para que eu mantivesse o foco e enfrentasse os desafios com coragem. Sua presença constante foi um verdadeiro alicerce em minha trajetória, e este momento não seria o mesmo sem você.

Aos meus amigos da universidade, agradeço pela parceria e pelas memórias inesquecíveis. Vocês tornaram os momentos mais difíceis mais suportáveis e os momentos felizes ainda mais especiais. Levarei vocês para toda a vida.

Aos meus amigos de infância, minha eterna gratidão pelo companheirismo e por sempre acreditarem em mim. O apoio de vocês foi essencial em cada etapa desta jornada, tornando os desafios mais leves e as conquistas ainda mais especiais.

Por fim, dedico meu agradecimento a Deus, cuja presença em minha vida foi constante e essencial. Sua orientação e força me sustentaram nos momentos de incerteza e me guiaram por caminhos que me trouxeram até aqui. A Ele dedico esta conquista, fruto das bênçãos e da fé que me sustentaram em todos os momentos.

*“Porque Deus tanto amou o mundo que deu o seu Filho Unigênito, para que todo o que nele crer não pereça, mas tenha a vida eterna.”
(João 3:16)*

Resumo

O setor varejista é altamente dinâmico e competitivo, exigindo estratégias eficazes para a gestão de estoques, precificação e logística. A previsão de demanda se torna uma ferramenta essencial para otimizar recursos, reduzir custos operacionais e garantir a disponibilidade dos produtos, minimizando perdas e melhorando a experiência do consumidor. Além disso, em um mercado onde as preferências dos clientes mudam rapidamente e fatores sazonais influenciam diretamente as vendas, prever a demanda com precisão permite que as empresas se antecipem às oscilações do mercado e ajustem suas operações de forma ágil e eficiente. Este trabalho apresenta uma análise comparativa de modelos de previsão de demanda aplicados a múltiplas bases de dados do varejo, considerando diferentes níveis de granularidade, como vendas totais, por loja e por produto. Foram utilizados quatro algoritmos quantitativos amplamente adotados: ARIMA, Prophet, LSTM e Redes Neurais Convolucionais (CNNs), avaliando seu desempenho preditivo em bases de dados reconhecidas, como Favorita, M5 e Olist. As métricas de avaliação consideradas incluem RMSE e MAPE, além do tempo de treinamento e previsão. Os resultados evidenciaram a superioridade do modelo LSTM, que obteve um RMSE médio 7% melhor que o segundo melhor modelo. Na base M5, o LSTM se destacou com um RMSE médio de 1013, aproximadamente 43% melhor que o ARIMA, que foi o segundo melhor modelo na maioria das métricas gerais. As CNNs mostraram melhor desempenho na base Olist, enquanto o Prophet apresentou limitações na captura de padrões em séries temporais mais complexas. O estudo conclui que técnicas avançadas, como redes neurais, são fundamentais para lidar com a complexidade dos dados no varejo, proporcionando maior precisão e eficiência na previsão de demanda, auxiliando na tomada de decisões estratégicas para o setor.

Palavras-chave: Previsão de Demanda; Aprendizado de Máquina; Redes Neurais; Varejo;

Abstract

The retail sector is highly dynamic and competitive, requiring effective strategies for inventory management, pricing, and logistics. Demand forecasting becomes an essential tool to optimize resources, reduce operational costs, and ensure product availability, minimizing losses and improving the consumer experience. Furthermore, in a market where customer preferences change rapidly and seasonal factors directly influence sales, accurately forecasting demand allows companies to anticipate market fluctuations and adjust their operations quickly and efficiently. This paper presents a comparative analysis of demand forecasting models applied to multiple retail databases, considering different levels of granularity, such as total sales, by store, and by product. Four widely adopted quantitative algorithms were used: ARIMA, Prophet, LSTM, and Convolutional Neural Networks (CNNs), evaluating their predictive performance on recognized databases, such as Favorita, M5, and Olist. Evaluation metrics considered include RMSE and MAPE, in addition to training and prediction time. The results highlighted the superiority of the LSTM model, which obtained an average RMSE 7% better than the second best model. In the M5 database, LSTM stood out with an average RMSE of 1013, approximately 43% better than ARIMA, which was the second best model in most general metrics. CNNs showed better performance in the Olist database, while Prophet showed limitations in capturing patterns in more complex time series. The study concludes that advanced techniques, such as neural networks, are essential to deal with the complexity of data in retail, providing greater accuracy and efficiency in demand forecasting, aiding in strategic decision-making for the sector.

Keywords: Demand Forecasting; Machine Learning; Neural Networks; Retail;

Lista de ilustrações

Figura 1 – Exemplificação da Operação de <i>Max Pooling</i>	33
Figura 2 – Esquema da Base de Dados da Olist.	39
Figura 3 – Vendas Diárias Totais - Favorita.	40
Figura 4 – Vendas Diárias Totais por Loja - Favorita.	41
Figura 5 – Vendas Diárias Totais por Loja e Produto - Favorita.	42
Figura 6 – Vendas Diárias Totais - M5.	42
Figura 7 – Vendas Diárias Totais por Loja - M5.	43
Figura 8 – Vendas Diárias Totais por Loja e Produto - M5.	44
Figura 9 – Vendas Diárias Totais - Olist.	44
Figura 10 – Vendas Diárias Totais por Loja - Olist.	45
Figura 11 – Vendas Diárias Totais por Loja e Produto - Olist.	46
Figura 12 – Vendas Semanais Totais - Favorita.	46
Figura 13 – Previsões do Conjunto de Teste - Favorita.	52
Figura 14 – Previsões do Conjunto de Teste - M5.	52
Figura 15 – Previsões do Conjunto de Teste - Olist.	53
Figura 16 – Previsões do Conjunto de Teste por Loja - Favorita.	54
Figura 17 – Previsões do Conjunto de Teste por Loja - M5.	55
Figura 18 – Previsões do Conjunto de Teste por Loja - Olist.	56
Figura 19 – Previsões do Conjunto de Teste dos Produtos por Loja - Favorita.	58
Figura 20 – Previsões do Conjunto de Teste dos Produtos por Loja - M5.	59
Figura 21 – Previsões do Conjunto de Teste dos Produtos por Loja - Olist.	60
Figura 22 – <i>Treemap</i> de Tempo Total de Treino por Modelo.	61
Figura 23 – <i>Treemap</i> de Tempo Total de Previsão por Modelo.	61

Lista de tabelas

Tabela 1 – Tabela de Resultados Gerais - Favorita.	51
Tabela 2 – Tabela de Resultados Gerais - M5.	52
Tabela 3 – Tabela de Resultados Gerais - Olist.	53
Tabela 4 – Tabela de Resultado Médio das Lojas - Favorita.	53
Tabela 5 – Tabela de Resultado Médio das Lojas - M5.	55
Tabela 6 – Tabela de Resultado Médio das Lojas - Olist.	56
Tabela 7 – Tabela de Resultado Médio dos Produtos por Loja - Favorita.	57
Tabela 8 – Tabela de Resultado Médio dos Produtos por Loja - M5.	57
Tabela 9 – Tabela de Resultado Médio dos Produtos por Loja - Olist.	58
Tabela 10 – Contagem de Modelos nas 100 Melhores Previsões.	62
Tabela 11 – MAPE Médio das 100 Melhores Previsões por Modelo.	62
Tabela 12 – RMSE Médio por Modelo.	62
Tabela 13 – RMSE Médio por Base e Modelo.	63

Lista de abreviaturas e siglas

ARIMA	AutoRegressive Integrated Moving Average
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ReLU	Rectified Linear Unit
SARIMA	Seasonal AutoRegressive Integrated Moving Average
TanH	Tangent Hyperbolic

Sumário

1	INTRODUÇÃO	21
1.1	Objetivo	22
1.2	Objetivo Específico	22
1.3	Justificativa	22
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Previsão de Demanda	25
2.2	Séries Temporais	26
2.3	Modelos de Previsão	27
2.3.1	Arima	27
2.3.1.1	Sarima	28
2.3.2	Prophet	29
2.3.3	LSTM	30
2.3.4	CNN	31
2.4	Métricas de Avaliação	33
2.4.1	RMSE	33
2.4.2	MAPE	33
3	REVISÃO DE LITERATURA	35
4	METODOLOGIA	37
4.1	Descrição Geral	37
4.2	Bases	37
4.2.1	Favorita	38
4.2.2	M5	38
4.2.3	Olist	39
4.3	Análise Exploratória de Dados	40
4.4	Pré-processamento das Bases	43
4.5	Implementação dos Modelos	46
4.5.1	ARIMA	46
4.5.2	Prophet	47
4.5.3	LSTM	47
4.5.4	CNN	48
4.6	Métricas Coletadas	49
5	ANÁLISE E DISCUSSÃO DOS RESULTADOS	51

5.1	Resultado de Todas as Lojas e Itens	51
5.2	Resultado por Loja	53
5.3	Resultado por Loja e Produto	56
5.4	Resultado Temporal	59
5.5	Análise Geral de Resultado	61
6	CONCLUSÃO	65
6.1	Limitações	66
6.2	Trabalhos Futuros	66
	REFERÊNCIAS	69

1 Introdução

O varejo, setor dinâmico e fundamental da economia global, representa o ponto final da cadeia de suprimentos, conectando-se diretamente ao consumidor. Desde pequenos estabelecimentos até grandes redes internacionais, esse segmento é moldado por fatores econômicos, culturais e tecnológicos, além de ser altamente sensível às mudanças no comportamento do consumidor. Para se destacar nesse cenário competitivo, os varejistas precisam conciliar eficiência operacional com a experiência do cliente, garantindo que os produtos estejam disponíveis no momento e local adequados. A disponibilidade do bem, aliás, é um fator que influencia significativamente a demanda, como ressaltam Lustosa et al. (2008).

Uma gestão eficiente no varejo depende, em grande parte, de decisões fundamentadas por dados, e é nesse cenário que a previsão de demanda desempenha um papel estratégico. A capacidade de prever com robustez e precisão as demandas da empresa traz ganhos mais competitivos, de acordo com Novaes (2021). Por exemplo, estoques em excesso podem gerar custos adicionais de armazenamento e perdas por obsolescência, enquanto estoques insuficientes podem levar a rupturas e à insatisfação dos clientes. Além disso, prever a demanda auxilia no planejamento logístico, na definição de estratégias de marketing e no ajuste de preços, contribuindo para a maximização dos lucros.

A previsão de demanda também tem se tornado cada vez mais desafiadora devido à complexidade dos mercados modernos. O aumento da diversidade de produtos, as flutuações sazonais e as tendências emergentes tornam o comportamento do consumidor mais difícil de antecipar. Segundo Novaes (2021) aspectos políticos, econômicos e sociais, nacionais e internacionais, podem dificultar e colocar a previsão sujeita ao erro. Para lidar com esses desafios, varejistas estão recorrendo a tecnologias avançadas, como aprendizado de máquina e análise preditiva, que oferecem maior precisão e flexibilidade. Além de atender às necessidades operacionais, essas previsões podem ser utilizadas para identificar oportunidades de crescimento e adaptar rapidamente as estratégias às mudanças de mercado. Assim, de acordo com Ballou (2006) a aplicação de técnicas de previsão não é apenas uma vantagem competitiva, mas também uma necessidade para o negócio, já que afeta todas as áreas.

Nos últimos anos, inúmeros trabalhos demonstraram vantagens e limitações dos principais modelos de previsão em diferentes cenários. Entretanto, no contexto do varejo, poucos trabalhos fizeram essa análise considerando modelos tradicionais e baseados em redes neurais. Dessa forma, ainda há exploração científica, tais como análise de vantagens e limitações dos modelos no contexto em questão.

1.1 Objetivo

O objetivo deste trabalho é realizar uma análise comparativa utilizando diferentes algoritmos para previsão de demanda, avaliando seu desempenho em distintos segmentos do varejo e em diferentes níveis de granularidade: por produto, por loja e todas as lojas e produtos. Além disso, outra finalidade é identificar os algoritmos que oferecem maior precisão de previsão e eficiência computacional, considerando o contexto das bases. Ademais, este trabalho tem como objetivo contribuir para o entendimento de como diferentes algoritmos podem ser aplicados, de forma prática, em diferentes contextos do varejo, visando otimizar processos e promover melhores decisões para um gerenciamento assertivo da demanda.

1.2 Objetivo Específico

Este trabalho tem como objetivos específicos aprofundar a análise do desempenho de abordagens de previsão de demanda em bases de dados com diferentes características, analisando as previsões realizadas para as bases em três níveis principais: vendas totais, vendas por loja e vendas por item de uma loja específica. Além disso, outro propósito específico é mensurar e comparar o tempo necessário para treinamento e geração de previsões em cada abordagem, analisando a eficiência computacional. Ademais, outro ponto crucial são métricas que avaliem a precisão das previsões, identificando e mensurando os erros, de modo a compreender melhor o desempenho de cada abordagem. Por fim, espera-se destacar as estratégias mais eficazes para diferentes situações e proporcionar subsídios para futuras implementações no setor varejista.

1.3 Justificativa

A previsão de demanda possui um papel muito importante no varejo, já que está diretamente relacionada à capacidade das empresas de gerenciar estoques, otimizar operações logísticas e atender às expectativas dos consumidores. Além disso, em um mercado competitivo, onde atrasos ou excessos de estoque podem representar perdas financeiras, prever a demanda com precisão é uma necessidade estratégica. No entanto, a escolha da abordagem mais adequada depende de vários fatores, como a qualidade dos dados disponíveis, o nível de detalhamento e o poder computacional necessário para cada método. Dessa forma, este estudo se faz relevante pois oferece uma visão abrangente e comparativa de diferentes estratégias de previsão de demanda em contextos variados. Ao explorar cenários diversificados do varejo e analisar diferentes granularidades de dados, espera-se compreender os limites e vantagens de cada abordagem.

Portanto, este trabalho contribui para a literatura acadêmica e para a prática

empresarial, ao oferecer análises que podem guiar decisões mais eficazes no setor varejista. Ademais, a pesquisa se alinha às necessidades reais do mercado, auxiliando empresas nos desafios de um ambiente em constante transformação e com consumidores cada vez mais exigentes.

2 Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica que sustenta as análises realizadas neste trabalho. Dessa forma, o conceito de previsão de demanda é explorado, destacando sua importância e desafios para a eficiência operacional e competitividade no varejo.

Além disso, este capítulo introduz os conceitos teóricos dos modelos de previsão aplicados, detalhando suas abordagens, fundamentos matemáticos e características principais. Por fim, são apresentadas as métricas utilizadas para avaliar a precisão e a eficiência computacional dos modelos, auxiliando a análise crítica e interpretação dos resultados nos capítulos seguintes.

2.1 Previsão de Demanda

De acordo com Lustosa et al. (2008), a previsão de demanda baseia-se, em grande parte, na análise de dados históricos de vendas, utilizando-os como referência para projetar o comportamento futuro do consumo. Dessa forma, a previsão de demanda pode ser entendida como o processo de estimar, com base em dados históricos e variáveis externas, o volume de produtos ou serviços que serão requisitados em determinado período futuro. Essa prática combina análise estatística, técnicas computacionais e conhecimento de mercado para identificar padrões e tendências de consumo, permitindo antecipar o comportamento do cliente.

Conforme Novaes (2021), a previsão de demanda é essencial para o planejamento estratégico no varejo e em cadeias produtivas, pois permite antecipar a evolução futura do consumo. Essa prática fundamenta decisões como pedidos a fornecedores, programação de matérias-primas e recursos de produção, além de influenciar a alocação de mão de obra, infraestrutura, depósitos e capital, todos elementos que requerem tempo e organização para serem executados de forma eficiente. No varejo, a importância dessa prática não pode ser subestimada. A capacidade de prever a demanda com precisão impacta diretamente a eficiência operacional e a experiência do consumidor. Uma boa previsão ajuda a otimizar estoques, reduzir desperdícios e melhorar o planejamento logístico. Além disso, auxilia na alocação de recursos e na definição de estratégias de precificação, campanhas de marketing e lançamento de produtos.

De acordo com Lustosa et al. (2008), há dois tipos de métodos de previsão de demanda: qualitativa e quantitativa. Os métodos qualitativos utilizam julgamentos de especialistas, opiniões de consumidores e dados qualitativos, de modo a tentar estimar a demanda futura. Por outro lado, os métodos quantitativos utilizam dados quantitativos e

técnicas estatísticas para realizar a previsão. Neste trabalho será dado total enfoque em métodos quantitativos.

No entanto, a previsão de demanda também enfrenta diversos desafios. Segundo Novaes (2021) as previsões sempre estarão sujeitas a erros. A complexidade dos dados, marcada por sazonalidades, mudanças de comportamento e influências externas, torna a tarefa de prever o futuro algo intrinsecamente incerto. Outro desafio é a necessidade de equilibrar a precisão das previsões com a eficiência computacional, especialmente quando se lida com grandes volumes de dados e múltiplos níveis de granularidade. Superar essas dificuldades requer um entendimento profundo dos métodos preditivos, aliado a um conhecimento estratégico sobre o mercado.

2.2 Séries Temporais

Podemos definir, de acordo com Morettin e Toloí (2018), que séries temporais são observações ordenadas pelo tempo. Essas séries são compostas por observações registradas em intervalos de tempo regulares, como minutos, horas, dias, meses ou anos. A análise de séries temporais é essencial para compreender o comportamento dinâmico dos sistemas e apoiar a tomada de decisões baseadas em previsões futuras.

Conforme Morettin e Toloí (2018), podemos exemplificar algumas séries temporais:

- valores diários de poluição na cidade de São Paulo;
- valores mensais de temperatura na cidade de Cananéia-SP;
- índices diários da Bolsa de Valores de São Paulo;
- precipitação atmosférica anual na cidade de Fortaleza;
- número médio anual de manchas solares;
- registro de marés ao porto de Santos

As séries temporais apresentam características únicas que as diferenciam de outras formas de dados. Uma das mais importantes é a dependência temporal, ou autocorrelação, que reflete o impacto das observações passadas nas observações futuras. Além disso, séries temporais podem conter componentes distintos, como tendência, sazonalidade e ruído. A tendência representa o comportamento de longo prazo da série, como um aumento ou diminuição consistente ao longo do tempo. A sazonalidade refere-se a padrões repetitivos em intervalos regulares, como picos de vendas durante feriados. O ruído compreende flutuações aleatórias e imprevisíveis nos dados. Podemos definir a relação dessas componentes, segundo Morettin e Toloí (2018), com a fórmula:

$$Z_t = T_t + S_t + a_t, \quad (2.1)$$

onde Z_t é a série temporal com número de observações t , T_t é a tendência, S_t é a sazonalidade e a_t é o ruído.

A análise de séries temporais desempenha um papel fundamental em diversos setores, especialmente em previsão de demanda. Compreender as variáveis envolvidas e identificar padrões nos dados são etapas essenciais para construir modelos preditivos precisos, que possam auxiliar empresas e organizações a otimizar processos, reduzir custos e aumentar a eficiência operacional.

2.3 Modelos de Previsão

Conforme mencionado anteriormente, este trabalho foca exclusivamente em modelos quantitativos de previsão de demanda, utilizando as seguintes abordagens: ARIMA, *Prophet*, *Long Short-Term Memory* (LSTM) e Redes Neurais Convolucionais (Convolutional Neural Network - CNN). Esses modelos foram selecionados por sua relevância e ampla aplicação em cenários de previsão baseados em dados históricos.

2.3.1 Arima

O modelo ARIMA, cujo nome é derivado do inglês *Autoregressive Integrated Moving Average*, foi desenvolvido por Box e Jenkins (1970). Esse modelo surgiu como parte de uma metodologia mais ampla proposta pelos autores, conhecida como Metodologia Box-Jenkins, que revolucionou a análise de séries temporais ao oferecer uma abordagem sistemática para identificação, estimativa e validação de modelos. Amplamente reconhecido como uma das abordagens mais consolidadas para a análise e previsão de séries temporais, o ARIMA é utilizado para identificar e modelar padrões em dados históricos, permitindo a projeção de eventos futuros. Sua popularidade no campo da estatística e da ciência de dados deve-se à sua robustez e simplicidade, especialmente em cenários onde a série temporal apresenta dependências lineares significativas.

De acordo com Morettin e Toloi (2018), o ARIMA é configurado como ARIMA(p, d, q), onde os parâmetros p, d e q, controlam os componentes autorregressivo, de diferenciação e de média móvel, respectivamente.

A parte autorregressiva, representada por AR, descreve a relação entre o valor atual da série temporal e seus valores passados. Este componente é baseado na ideia de que a série temporal apresenta dependência linear com os valores anteriores. A equação para o modelo AR de ordem p , tem a estrutura apresentada por Morettin e Toloi (2018):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t, \quad (2.2)$$

onde Y_t é o valor observado no tempo t , $\phi_1, \phi_2, \dots, \phi_p$ são os coeficientes autorregressivos e ϵ_t é o erro aleatório no tempo t .

A parte de média móvel, representada por MA, modela a relação entre o valor atual da série e os erros passados. Esse componente é responsável por capturar o comportamento de média dos resíduos dos valores anteriores. Conforme Adhikari e Agrawal (2013), definimos o modelo de média móvel de ordem q pela equação:

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad (2.3)$$

onde μ é a média da série, ϵ_t é o erro no tempo t , e $\theta_1, \theta_2, \dots, \theta_q$ são os coeficientes do modelo de média móvel.

O componente de integração, representado por d , trata da diferenciação da série temporal para torná-la estacionária, ou seja, para garantir que a série não tenha tendências de longo prazo que possam afetar a previsão. A diferenciação é aplicada repetidamente até que a série se torne estacionária, e é calculada pela diferença entre um valor e o valor anterior:

$$\Delta Y_t = Y_t - Y_{t-1}. \quad (2.4)$$

Sendo assim, o modelo ARIMA combina essas três componentes em uma equação geral que pode ser escrita, de acordo com Betti (2012), como:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t, \quad (2.5)$$

onde $\phi(B)$ e $\theta(B)$ são os polinômios de autorregressão e de média móvel, respectivamente, e B é o operador de defasagem.

2.3.1.1 Sarima

O modelo ARIMA é amplamente reconhecido pela sua eficácia na modelagem e previsão de séries temporais. Sua flexibilidade permite capturar padrões lineares nos dados, como tendências e correlações entre observações passadas. No entanto, em cenários onde as séries temporais apresentam sazonalidade, o modelo SARIMA (Seasonal Autoregressive Integrated Moving Average) surge como uma extensão natural e mais apropriada, ao incorporar componentes sazonais para capturar padrões que se repetem em intervalos regulares de tempo.

O modelo SARIMA é definido pela adição de parâmetros sazonais P, D, Q, m aos componentes tradicionais p, d, q do ARIMA. De acordo com Adhikari e Agrawal (2013), sua fórmula geral pode ser expressa como:

$$\Phi_p(B^s)\varphi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t, \quad (2.6)$$

onde $\Phi_p(B^s)$ e $\Theta_Q(B^s)$ são os polinômios autorregressivo e de média móvel sazonais, respectivamente, e $\varphi_p(B)$ e $\theta_q(B)$ são os polinômios autorregressivo e de média móvel não sazonais, respectivamente. Além disso, s é a periodicidade sazonal e $(1-B)^d$ e $(1-B^s)^D$ representam a diferenciação sazonais e não sazonais, respectivamente.

2.3.2 Prophet

O *Prophet* é um modelo de previsão de séries temporais desenvolvido pela equipe do Meta (*Facebook*) em 2017, projetado para lidar com dados que apresentam características como sazonalidade, tendências não lineares e efeitos de feriados. Seu principal objetivo é fornecer um método flexível e robusto que seja de fácil implementação, mesmo para usuários que não possuem amplo conhecimento em estatística ou aprendizado de máquina (FACEBOOK, 2023).

O modelo assume que a série temporal pode ser decomposta em três componentes principais: tendência, sazonalidade e eventos especiais. De acordo com Taylor e Letham (2017), a fórmula básica do *Prophet* pode ser expressa como:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (2.7)$$

onde $g(t)$ representa a tendência, $s(t)$ a sazonalidade, $h(t)$ os efeitos dos feriados e ϵ_t o erro.

A tendência pode ser calculada em diferentes formas, um modelo de crescimento saturado e um modelo linear por partes. Conforme Taylor e Letham (2017), podemos definir as fórmulas da seguinte forma:

$$a_j(t) = \begin{cases} 1, & \text{se } t \geq s_j \\ 0, & \text{outros} \end{cases} \quad (2.8)$$

$$\gamma_j = (s_j - m - \sum_{l < j} \gamma_l) \left(1 - \frac{k - \sum_{l < j} \delta_l}{k - \sum_{l \leq j} \delta_l}\right) \quad (2.9)$$

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))} \quad (2.10)$$

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma), \quad (2.11)$$

onde j , um inteiro, varia de 1 até S , sendo S a quantidade de pontos de mudança, δ um vetor de ajustes de taxas, em que δ_j é a mudança na taxa que ocorre no momento s_j .

A componente de sazonalidade, de acordo com Taylor e Letham (2017), pode ser descrita:

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (2.12)$$

sendo P o período (365,25 para dados anuais e 7 para dados semanais, por exemplo), a_n e b_n parâmetros de ajuste da sazonalidade e N o número do período.

Por fim, temos a componente dos feriados, a qual definimos segundo Taylor e Letham (2017), como:

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)] \quad (2.13)$$

$$h(t) = Z(t)\kappa \quad (2.14)$$

$$\kappa \sim Normal(0, v^2), \quad (2.15)$$

onde v é o nível de impacto dos feriados.

2.3.3 LSTM

As redes LSTM (*Long Short-Term Memory*), introduzidas por Hochreiter e Schmidhuber (1997), são uma variação das Redes Neurais Recorrentes (RNNs) projetadas para modelar sequências temporais e capturar dependências e padrões de longo prazo nos dados. Por isso, as LSTMs são amplamente utilizadas em problemas que envolvem séries temporais, como previsão de demanda, por exemplo.

De acordo com Amorim et al. (2023), a arquitetura das LSTMs inclui células de memória e três portões principais: o portão de entrada, o portão de esquecimento e o portão de saída. Essa estrutura permite que a rede retenha informações relevantes por períodos mais longos, dependendo da natureza dos dados.

O portão de esquecimento controla quais informações novas serão adicionadas à célula. De acordo com Olah (2015), podemos definir esse portão matematicamente, como:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (2.16)$$

onde f_t é o vetor da porta de esquecimento no tempo t , σ é a função *sigmoid*, W_f são os pesos associados, h_{t-1} é o estado anterior, x_t é a entrada no tempo t e b_f é o vetor de *bias*.

Ademais, o portão de esquecimento, que determina quais informações antigas devem ser descartadas. Conforme Olah (2015), primeiro, é decidido quais valores serão atualizados:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.17)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C). \quad (2.18)$$

Em seguida, a célula é atualizada:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (2.19)$$

onde C_{t-1} é o antigo estado.

Por fim, temos o portão de saída, que decide quais dados serão usados para a próxima etapa, que pode ser definido, conforme Olah (2015), como:

$$o(t) = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.20)$$

$$h_t = o_t * \tanh(C_t). \quad (2.21)$$

Portanto, essas equações permitem que as redes LSTM aprendam a manter informações relevantes e descartar irrelevantes em cada passo temporal, podendo ser altamente eficazes em prever demanda com base em dados históricos, onde os padrões de longo prazo e sazonalidade desempenham um papel crucial.

2.3.4 CNN

As Redes Neurais Convolucionais, do inglês *Convolutional Neural Networks* (CNNs), introduzidas por Fukushima (1980) e originalmente concebidas para tarefas de visão computacional, podem ser utilizadas para séries temporais, visualizando os dados com apenas uma dimensão, de acordo com Goodfellow, Bengio e Courville (2016). Além disso, podemos utilizar a seguinte definição: “As redes convolucionais são simplesmente redes neurais que usam a convolução no lugar da multiplicação geral de matrizes em pelo menos uma de suas camadas” (GOODFELLOW; BENGIO; COURVILLE, 2016).

As CNNs são formadas por algumas etapas, sendo elas: convolução, função de ativação, *pooling* e camada totalmente conectada.

Na camada de convolução é realizada a operação de convolução entre a matriz de entrada e algum filtro convolucional, denominado *kernel*. A fórmula de convolução pode ser definida como:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau) \cdot h(t - \tau) \cdot d\tau. \quad (2.22)$$

Em seguida, geralmente, são aplicadas funções de ativação. As funções de ativação têm como objetivo introduzir não-linearidade. As funções de ativação mais comuns são: *Rectified Linear Unit* (ReLU), *Sigmoid* e Tangente Hiperbólica (TanH). Podemos definir as fórmulas, de acordo com Alzubaidi et al. (2021), como:

$$ReLU(x) = \max(0, x) \quad (2.23)$$

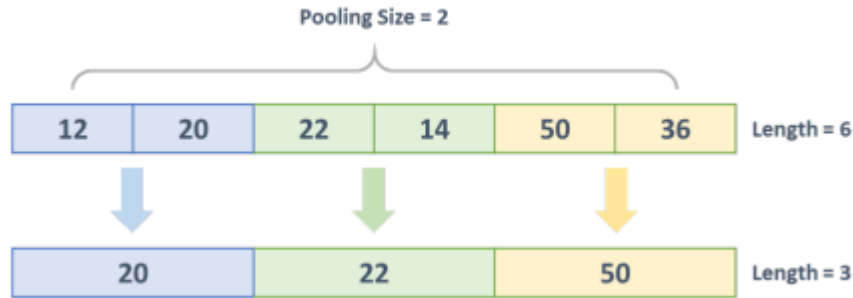
$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (2.24)$$

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (2.25)$$

De acordo com Goodfellow, Bengio e Courville (2016), a camada de *pooling*, ou agrupamento, aplica operações de estatística descritiva que auxiliam no processamento dos dados. Além disso, essa etapa contribui para tornar a rede neural mais invariável a pequenas variações nos dados de entrada e melhora a eficiência computacional.

As funções de *pooling* mais utilizadas em redes neurais convolucionais são o *max pooling* e o *average pooling*. De acordo com Zhao e Zhang (2024), essas funções reduzem as dimensões espaciais das entradas. No caso do *max pooling*, é selecionado o maior valor dentro de uma região específica, enquanto no *average pooling* é calculada a média dos valores dessa mesma região. A Figura 1 exemplifica a operação de *max pooling* para um vetor de 6 posições com tamanho 2 de *pooling*.

Finalmente, temos a camada totalmente conectada, que segundo Ragab et al. (2020) é uma etapa especial e essencial das CNNs. Essa camada desempenha o papel crucial de consolidar as características extraídas pelas camadas anteriores, transformando-as em representações que permitem a realização da previsão final. Em sua estrutura, cada neurônio está conectado a todos os neurônios da camada precedente, possibilitando uma combinação abrangente e eficiente das informações aprendidas ao longo do modelo.

Figura 1 – Exemplificação da Operação de *Max Pooling*.

Fonte: (RAGAB et al., 2020)

2.4 Métricas de Avaliação

Para avaliar a eficácia das previsões de cada modelo, serão utilizadas duas métricas: RMSE e MAPE. Essas métricas nos ajudam a medir o tamanho do erro entre o que o modelo previu e o que realmente aconteceu.

2.4.1 RMSE

O *Root Mean Squared Error* (RMSE), conhecido em português como Raiz do Erro Quadrático Médio, é amplamente utilizado por sua capacidade de penalizar maiores diferenças entre os valores previstos e os valores reais. Ele faz isso elevando os erros ao quadrado antes de calcular a média e, posteriormente, extraindo sua raiz quadrada. Essa característica torna o RMSE particularmente útil em situações onde erros significativos são mais prejudiciais.

A fórmula do RMSE, conforme Matta et al. (2021), é dada por:

$$e_t = y_t - \hat{y}_t \quad (2.26)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}, \quad (2.27)$$

sendo y_t o valor real e \hat{y}_t o valor previsto.

2.4.2 MAPE

O *Mean Absolute Percentage Error* (MAPE), conhecido em português como Erro Percentual Absoluto Médio, mede o erro em termos percentuais, fornecendo uma visão relativa da precisão do modelo em relação ao tamanho da demanda prevista. Sua principal vantagem está na facilidade de interpretação, pois é expresso como uma porcentagem

média dos erros absolutos em relação aos valores reais. Isso permite compreender, em média, o quanto as previsões diferem dos valores reais, independentemente da escala dos dados.

A fórmula do MAPE, conforme Matta et al. (2021), é dada por:

$$MAPE = \frac{1}{n} \left(\sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \right) \cdot 100. \quad (2.28)$$

Uma das principais limitações do MAPE é o problema da explosão em valores próximos de zero. Como o cálculo do MAPE envolve dividir o erro absoluto pelo valor real, quando os valores reais são muito pequenos ou se aproximam de zero, o erro percentual se torna excessivamente grande, inflacionando os resultados. Em cenários onde há uma alta ocorrência de valores baixos ou próximos de zero na série temporal, o MAPE pode se tornar uma métrica inadequada para avaliar a qualidade das previsões, destacando a necessidade de usar métricas complementares, como o RMSE, que não sofrem desse tipo de limitação.

3 Revisão de Literatura

A previsão eficaz de demanda é essencial para a gestão de estoques no varejo, como destacado por Shak et al. (2024) em seu artigo “Optimizing Retail Demand Forecasting: A Performance Evaluation of Machine Learning Models Including LSTM and Gradient Boosting”. O artigo avalia cinco modelos de aprendizado de máquina – Regressão Linear, Regressor de Árvore de Decisão, Regressor de Floresta Aleatória, *Gradient Boosting* e Redes Neurais do tipo LSTM – aplicados a um conjunto de dados que inclui vendas transacionais, promoções, eventos de calendário e fatores externos, como clima e indicadores econômicos. Os resultados indicaram que o LSTM superou os demais, evidenciando sua capacidade de capturar dependências temporais e padrões complexos de demanda. Embora *Gradient Boosting* e Floresta Aleatória também tenham apresentado bons resultados, a Regressão Linear e o Regressor de Árvore de Decisão demonstraram limitações. Este estudo reforça a eficácia de modelos avançados, especialmente o LSTM, para aprimorar a precisão das previsões de demanda, oferecendo *insights* valiosos para a otimização de estoques e operações no varejo.

O artigo “Enhancing Time Series Product Demand Forecasting With Hybrid Attention-Based Deep Learning Model”, proposto por Zhang et al. (2024), apresenta uma abordagem inovadora para previsão de séries temporais, com foco em demandas de produtos, utilizando técnicas avançadas de aprendizado profundo. A proposta central do estudo é o modelo Hybrid Attention-based Long Short-Term Memory (HA-LSTM), que combina mecanismos de atenção multi-head com camadas LSTM tradicionais. Essa integração permite capturar tanto dependências de longo prazo quanto padrões temporais locais em dados complexos de séries temporais. O modelo foi avaliado utilizando um conjunto de dados em larga escala da competição “Predict Future Sales”, demonstrando eficácia na manipulação de séries temporais reais com sazonalidades e tendências múltiplas. Os resultados experimentais mostram que o HA-LSTM supera métodos consagrados, como ARIMA, *Prophet* e LSTM padrão. O HA-LSTM não apenas melhora a precisão das previsões, mas também oferece pesos de atenção interpretáveis, fornecendo ideias sobre a importância relativa de diferentes instantes temporais nas previsões. Este trabalho contribui significativamente para o campo de análise de séries temporais com aprendizado profundo e apresenta implicações práticas para o aprimoramento da previsão de demanda no varejo e na gestão da cadeia de suprimentos.

O artigo “Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning”, proposto por Nasser et al. (2023), investiga a aplicação de aprendizado de máquina para prever a demanda no varejo, utilizando dados históricos de mais de seis anos, incluindo variáveis

externas como condições meteorológicas e impactos da COVID-19. Avaliando três categorias de produtos perecíveis, o estudo mostrou que o modelo *Extra Tree Regressor* (ETR) superou redes LSTM, com maior vantagem para carnes frescas e diferenças menores para frutas. O ETR também apresentou resultados consistentes com outros modelos baseados em árvores, como *XGBoost*, Florestas Aleatórias e *Gradient Boosting*, consolidando a eficácia desses métodos em comparação com LSTM. Os resultados achados oferecem ideias para otimizar a previsão de demanda em diferentes horizontes temporais, contribuindo para a eficiência da gestão da cadeia de suprimentos no varejo.

O artigo “Retail Demand Forecasting Using Spatial-Temporal Gradient Boosting Methods”, apresentado por Wang Woon Kian Chong e Hedenstierna (2024), aborda a necessidade crescente de algoritmos eficientes para previsão de demanda no varejo, especialmente diante da explosão no número de unidades de manutenção de estoque (SKUs) impulsionada pelo crescimento do comércio eletrônico. Os autores apresentam um novo algoritmo de aprendizado de máquina, denominado *Spatial-Temporal Gradient Boosting Tree* (ST-GBT), que integra informações transversais e temporais à metodologia tradicional de árvores de decisão baseadas em gradiente. Este algoritmo é capaz de prever demandas de inúmeros SKUs simultaneamente, demonstrando alta precisão e capacidade de automação. Além disso, o estudo identifica fatores relacionados ao varejo e propõe novos indicadores, como estatísticas de ordem superior e taxas de juros livres de risco, para aprimorar a tarefa de previsão. Experimentos numéricos realizados com registros históricos de transações de uma grande empresa de *e-commerce* reforçam as vantagens comparativas do ST-GBT, destacando sua superioridade em termos de exatidão e eficiência.

O artigo “Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail”, proposto por Oliveira e Ramos (2024) explora o desempenho de modelos baseados em *Transformers* na previsão de demanda no varejo. Foram avaliados modelos como *Transformer*, *Informer*, *Autoformer*, *PatchTST* e *Temporal Fusion Transformer* (TFT) em comparação com métodos tradicionais, como *AutoARIMA* e *AutoETS*, utilizando o conjunto de dados da competição M5, o mesmo usado neste trabalho. Os resultados evidenciam que os modelos baseados em *Transformers* superaram significativamente os métodos tradicionais. *Transformer*, *Informer* e TFT lideraram as métricas de desempenho, enquanto *Autoformer* e *PatchTST*, embora superiores aos métodos tradicionais, apresentaram resultados ligeiramente inferiores, sugerindo espaço para ajustes. O estudo também aponta um equilíbrio entre a complexidade dos modelos e a eficiência computacional, com os *Transformers* oferecendo maior precisão, mas exigindo maior capacidade computacional. Esses achados destacam o potencial dos *Transformers* para melhorar a previsão de demanda no varejo, desde que os desafios computacionais sejam gerenciados adequadamente.

4 Metodologia

Este capítulo descreve como os métodos e técnicas foram utilizados para conduzir a análise. A metodologia foi estruturada para garantir a consistência no tratamento dos dados, a aplicação adequada dos modelos de previsão e a coleta de métricas de análise comparativa.

Ademais, vale ressaltar que todas as implementações foram feitas utilizando a linguagem de programação *Python*, que é amplamente usada para aplicações de ciência de dados, inclusive previsão de demanda. O código-fonte da aplicação encontra-se disponível no repositório do *GitHub*, Campos (2025), acessível em: <<https://github.com/midas-ufscar/2024-2-TCC-2-Brainer-Sueverti-de-Campos>>.

4.1 Descrição Geral

Inicialmente, os dados históricos de demanda das bases foram coletados. Em seguida, buscou-se compreender as características das séries temporais, identificando padrões, sazonalidades e possíveis inconsistências nos dados. Ademais, o tratamento das bases incluiu limpeza, normalização e organização dos dados, garantindo sua adequação aos modelos escolhidos.

Sendo assim, os modelos foram aplicados considerando diferentes granularidades e bases, explorando as variações de comportamento entre os conjuntos de dados. O desempenho foi avaliado utilizando métricas de previsão, como RMSE e MAPE, além de métricas temporais, como o tempo de treinamento e previsão.

4.2 Bases

O *Kaggle* é uma plataforma online amplamente reconhecida na comunidade de ciência de dados e aprendizado de máquina. Fundada em 2010 e adquirida pelo *Google* em 2017, ela se tornou um repositório confiável para datasets, além de oferecer competições de aprendizado de máquina, tutoriais, notebooks interativos e uma comunidade ativa de profissionais e entusiastas.

Uma das principais vantagens do *Kaggle* é a qualidade e a diversidade dos datasets disponíveis, abrangendo desde dados simples e bem estruturados até conjuntos mais complexos, frequentemente utilizados em competições de alto nível. Os dados podem ser fornecidos por empresas, organizações ou indivíduos que buscam soluções para problemas reais ou querem disponibilizar informações para análises acadêmicas e comerciais.

Quanto à confiabilidade, o *Kaggle* permite que os usuários avaliem e comentem os datasets, contribuindo para a validação da qualidade das informações. Muitos datasets no *Kaggle* são bem documentados, com descrições claras e informações sobre a origem, o que garante maior transparência e confiança. Entretanto, como qualquer plataforma aberta, é importante verificar a procedência do *dataset*.

Sendo assim, devido à relevância da plataforma, todas as bases selecionadas foram extraídas do *Kaggle* e são apresentadas nas próximas subseções.

4.2.1 Favorita

A base de dados utilizada na competição “Corporación Favorita Grocery Sales Forecasting”, disponível em Favorita et al. (2017), é muito utilizada para estudos e aplicações relacionadas à previsão. Esta base foi originalmente disponibilizada para uma competição de aprendizado de máquina cujo objetivo era prever as vendas unitárias diárias de diversos produtos em múltiplas lojas de supermercados da rede *Corporación Favorita*, uma grande varejista equatoriana.

A base é composta por diversos arquivos que contêm informações detalhadas sobre as vendas históricas, características dos produtos, dados das lojas e fatores externos que podem influenciar o comportamento das vendas, como feriados e promoções. Além disso, estão incluídos dados macroeconômicos, como índices de preço, que podem ser úteis na modelagem das séries temporais.

A confiabilidade dessa base é reforçada pela curadoria feita pelo *Kaggle* e pela grande adesão de participantes da competição, que contribuem com análises e validações. Assim, ela se tornou uma referência amplamente utilizada em estudos de previsão de demanda e séries temporais.

4.2.2 M5

A base de dados da competição “M5 Forecasting - Accuracy”, disponível em Howard et al. (2020), também é muito utilizada para estudos de previsão de vendas. Essa competição foi organizada pela *Walmart*, uma das maiores redes de varejo do mundo, fundada em 1962 nos Estados Unidos e atualmente presente em dezenas de países.

A base contém um histórico detalhado de cinco anos de vendas diárias, abrangendo 3.049 produtos distribuídos em 10 lojas localizadas em três estados dos Estados Unidos. Além dos dados de vendas, o conjunto inclui informações contextuais, como categorias e subcategorias dos produtos, variações de preço, impacto de promoções e feriados, além de outros fatores externos que podem influenciar a demanda.

O desafio proposto pela competição é prever as vendas futuras em múltiplos níveis de granularidade, desde o desempenho individual de um produto em uma loja específica até

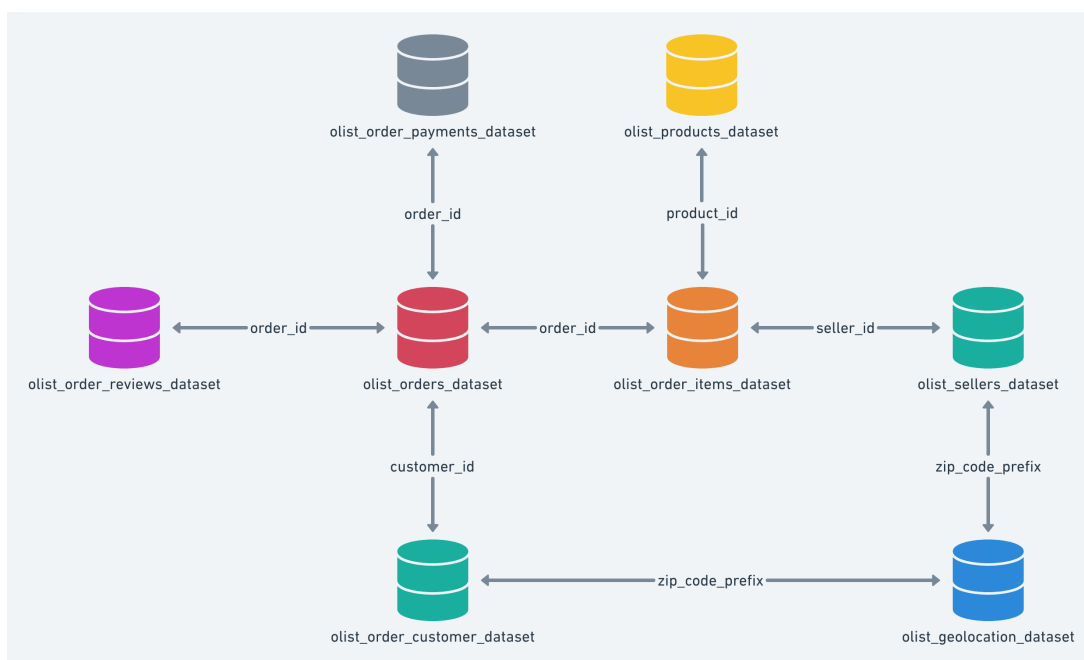
projeções mais agregadas por região ou categoria. Essa característica exige modelos robustos que sejam capazes de capturar padrões temporais complexos e interações contextuais em diferentes escalas.

4.2.3 Olist

A base de dados da *Olist*, disponível em Olist e Sionek (2018), oferece um panorama abrangente sobre o comportamento do comércio eletrônico no Brasil. A *Olist* é uma plataforma de *e-commerce* que conecta pequenos e médios vendedores a grandes *marketplaces*, possibilitando que lojistas tenham alcance nacional e maior visibilidade de seus produtos. A empresa, fundada em 2015, tem como foco democratizar o acesso ao mercado online para empreendedores de diversas categorias, oferecendo soluções integradas de gestão, logística e marketing.

O conjunto de dados inclui informações detalhadas sobre transações realizadas na plataforma, abrangendo um período significativo de operações. Ele contém dados relacionados a pedidos, avaliações de clientes, informações demográficas, descrição dos produtos, categorias e prazos de entrega. Essa riqueza de informações torna a base altamente relevante para estudos e análises, como previsão de demanda, análise de comportamento de compra e estratégias de marketing. A Figura 2 ilustra como é o esquema da base de dados e como se relacionam.

Figura 2 – Esquema da Base de Dados da Olist.



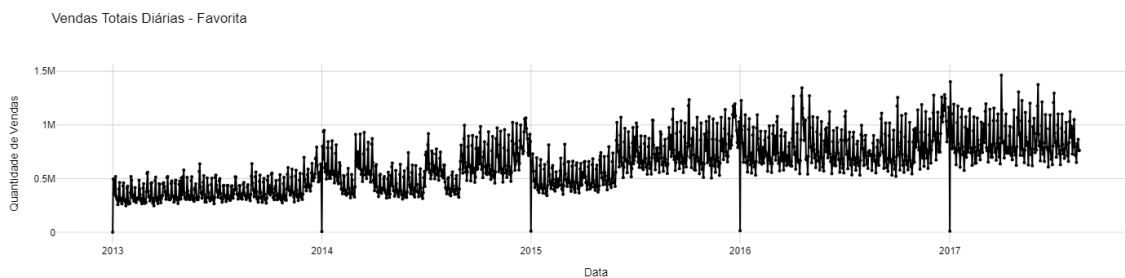
Fonte: (OLIST; SIONEK, 2018)

4.3 Análise Exploratória de Dados

A etapa de Análise Exploratória de Dados foi conduzida com o objetivo de compreender as características e a qualidade das bases de dados utilizadas no estudo. Durante essa etapa, verificou-se que nenhuma das bases apresentou problemas estruturais, como valores ausentes, inconsistências ou duplicações significativas. Essa constatação reflete a alta qualidade das bases selecionadas, que foram bem estruturadas e organizadas, o que simplificou o processo de preparação dos dados para as etapas seguintes da metodologia.

A base Favorita não apresentou problemas estruturais nos dados. No entanto, é evidente a presença de um elevado nível de ruído na curva, o que pode dificultar uma análise mais precisa. A Figura 3 exibe a série temporal das vendas diárias da base, onde se pode observar uma forte sazonalidade, caracterizada por padrões recorrentes ao longo do tempo. Além disso, há uma clara tendência de crescimento nas vendas, indicando um aumento gradual na atividade comercial.

Figura 3 – Vendas Diárias Totais - Favorita.

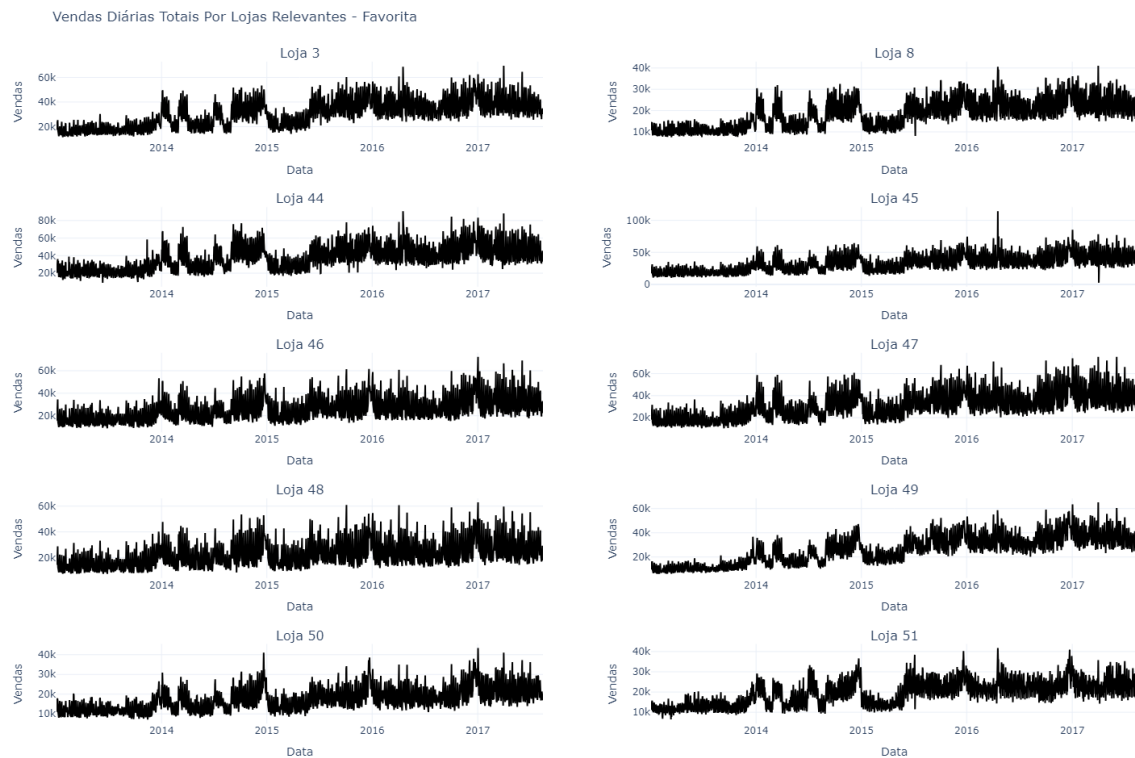


Fonte: Autor

Além disso, por meio da Figura 4, é possível observar graficamente o comportamento de algumas lojas da base da Favorita. A análise gráfica deixa visível que há sazonalidade presente em todas as séries temporais. Ademais, é possível notar que algumas lojas demonstram uma sazonalidade mais acentuada do que outras. Observa-se, também, uma leve tendência de alta em algumas séries, o que pode indicar um crescimento gradual nas vendas ao longo do tempo, possivelmente influenciado por estratégias de marketing, expansão do mercado ou aumento na fidelização dos clientes.

A Figura 5 ilustra graficamente o comportamento das séries temporais de produtos em algumas lojas da base da Favorita. A análise visual permite identificar uma presença bastante significativa de ruídos nos gráficos, o que pode dificultar a identificação de padrões consistentes ao longo do tempo. Além disso, observa-se que alguns produtos apresentam períodos consideráveis sem registros de vendas em determinadas lojas, indicando possíveis lacunas de dados.

A base da M5 também não apresentou dificuldades estruturais. Sua principal

Figura 4 – Vendas Diárias Totais por Loja - Favorita.

Fonte: Autor

diferença das outras bases é não possuir datas reais, mas apenas números sequenciais de tempo, de forma a possuir mais de 1900 dias de vendas. A Figura 6 ilustra a série temporal semanal das vendas totais da M5. Assim, é possível observar uma tendência de crescimento e uma sazonalidade muito forte.

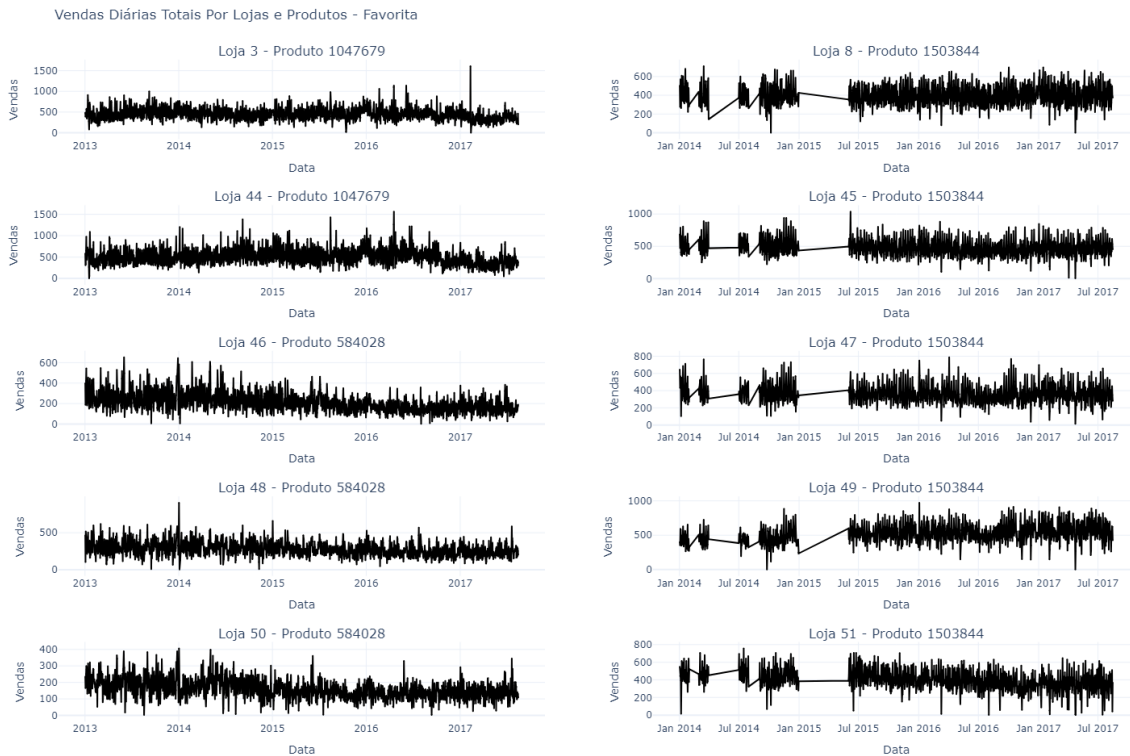
Além disso, de acordo com a Figura 7, é possível observar um comportamento muito parecido entre todas as 10 lojas da base M5, com sazonalidades fortes, ruídos e uma leve tendência de crescimento.

Os produtos das lojas da base M5 apresentam ruídos nas séries temporais e, alguns, apresentam um espaço de tempo sem vendas diárias. Além disso, as séries apresentam uma forte sazonalidade, contudo nem todas apresentam tendência de alta nas vendas.

A base da *Olist* não apresentou nenhum problema em sua estrutura. Ela possui aproximadamente dois anos de dados, com aproximadamente 3095 lojas e 32951 produtos. A Figura 9 mostra a série temporal das vendas totais diárias, ou seja, o acumulado diário de todas as vendas de todas as lojas e produtos. Ademais, é possível observar que a série temporal possui um comportamento sazonal, possui um grande pico em 24 de novembro de 2017, e baixas no início e final da curva.

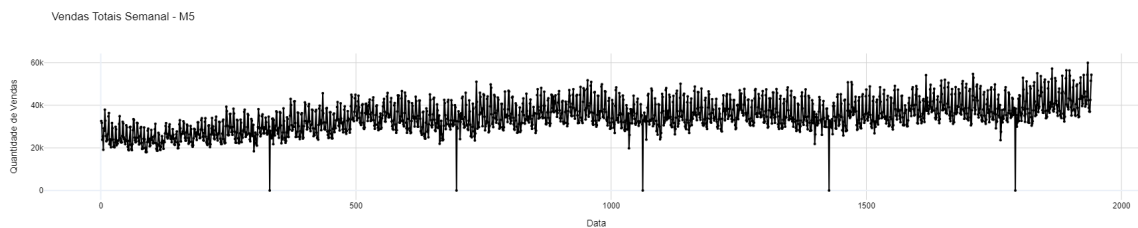
A Figura 10 ilustra graficamente as séries de vendas de algumas lojas da base da *Olist*, de modo que é possível observar uma diferença considerável no comportamento entre

Figura 5 – Vendas Diárias Totais por Loja e Produto - Favorita.



Fonte: Autor

Figura 6 – Vendas Diárias Totais - M5.

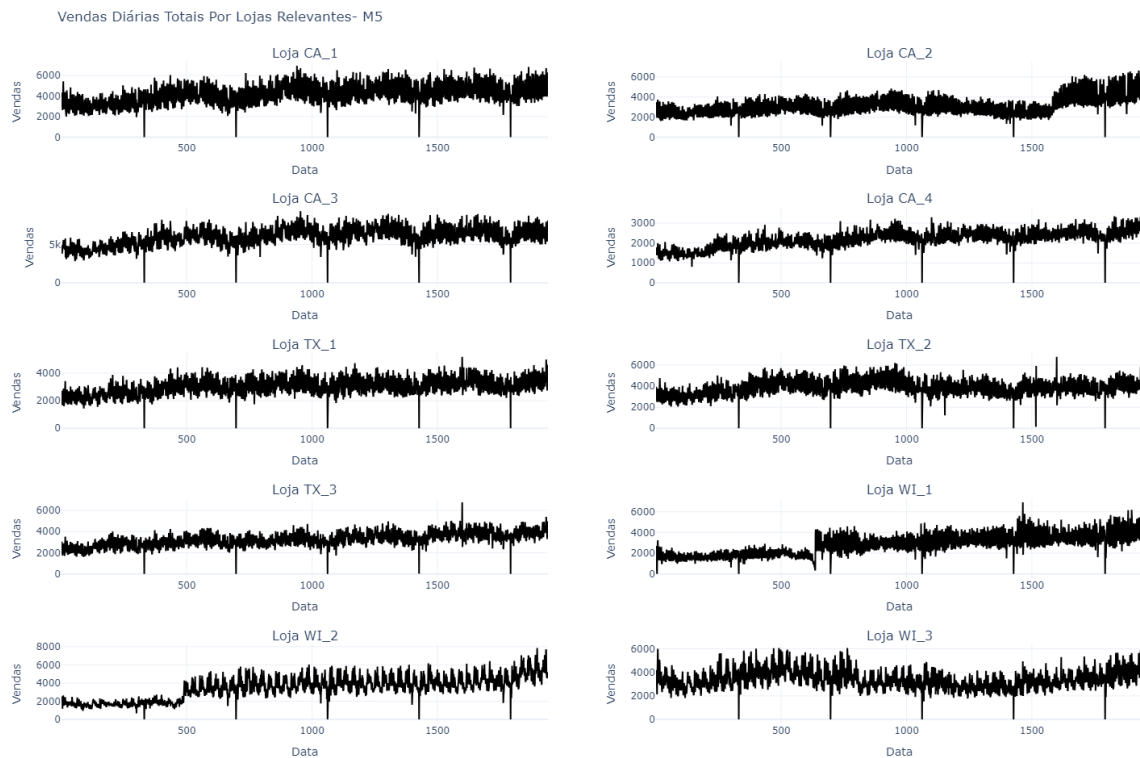


Fonte: Autor

algumas curvas. Além disso, nota-se grandes picos de vendas característicos, que destoam dos outros pontos.

Conforme a Figura 11, os produtos de algumas lojas apresentam comportamentos completamente distintos entre si, com dinâmicas de vendas que variam amplamente de um produto para outro. Outro aspecto relevante é a presença de longos períodos sem registros de vendas para diversos produtos. Isso pode ser explicado pelo fato de essas lojas serem menores e, em muitos casos, possuírem um portfólio mais diversificado. Lojas com essa configuração tendem a ter uma demanda menos estável, com oscilações mais acentuadas.

De forma geral, a análise exploratória reforçou a confiabilidade das bases de dados, assegurando um ponto de partida sólido para o desenvolvimento das etapas de modelagem,

Figura 7 – Vendas Diárias Totais por Loja - M5.

Fonte: Autor

validação e coleta de métricas. A qualidade dos dados contribuiu significativamente para a precisão e robustez das análises realizadas ao longo deste trabalho.

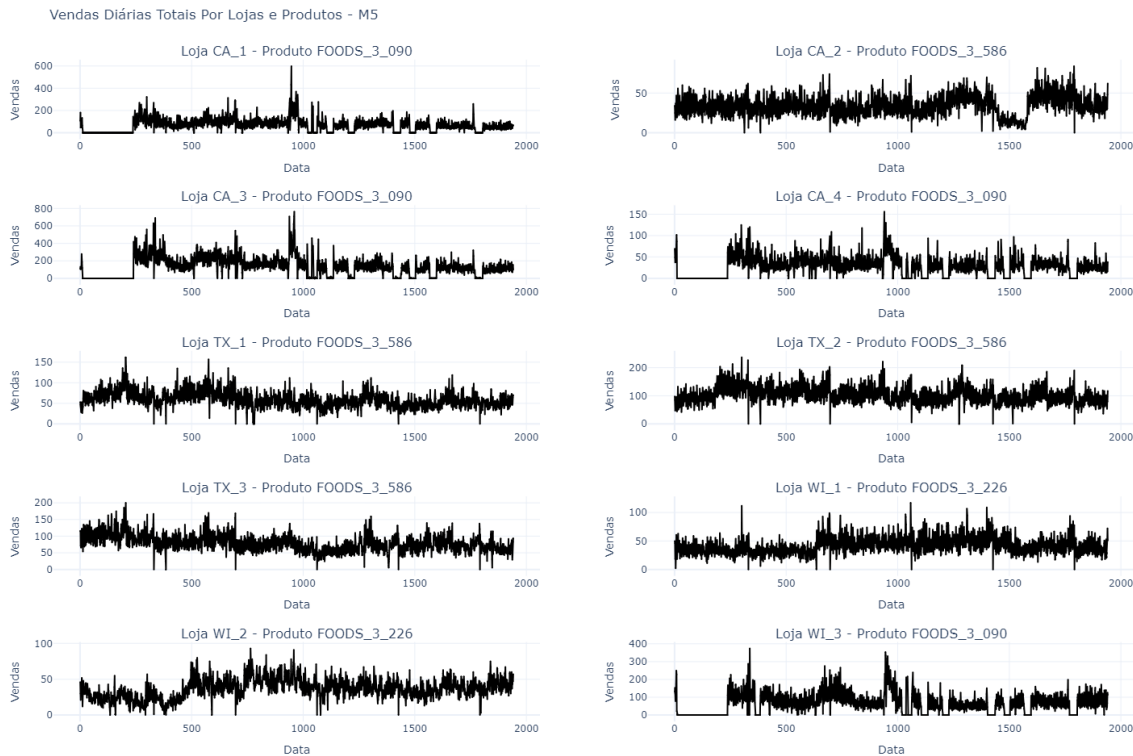
4.4 Pré-processamento das Bases

A etapa de pré-processamento das bases de dados desempenha um papel fundamental na preparação dos dados para garantir a qualidade e a consistência necessárias para a aplicação dos modelos preditivos. O objetivo principal dessa etapa é realizar ajustes, transformações e limpezas que tornem as bases adequadas para o processo de modelagem.

No caso das bases utilizadas neste trabalho, não foram necessárias mudanças complexas, uma vez que apresentaram boa qualidade estrutural desde o início. As ações realizadas focaram em tarefas pontuais, como a conversão de formatos de colunas para garantir a compatibilidade com os modelos, especialmente em relação a dados temporais e numéricos. Essas conversões incluíram ajustes de tipos de dados, como a transformação de datas para formatos padrão e a normalização de valores numéricos, quando necessário.

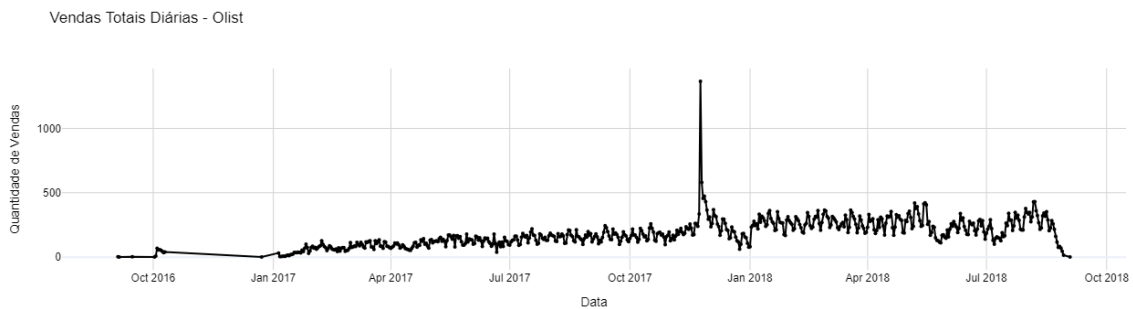
Além disso, foi realizada a exclusão das últimas semanas de registros devido à incompletude dos dados nesse período. Essas semanas apresentaram uma redução abrupta no volume de vendas, o que poderia atrapalhar os modelos. A decisão de remover esses

Figura 8 – Vendas Diárias Totais por Loja e Produto - M5.



Fonte: Autor

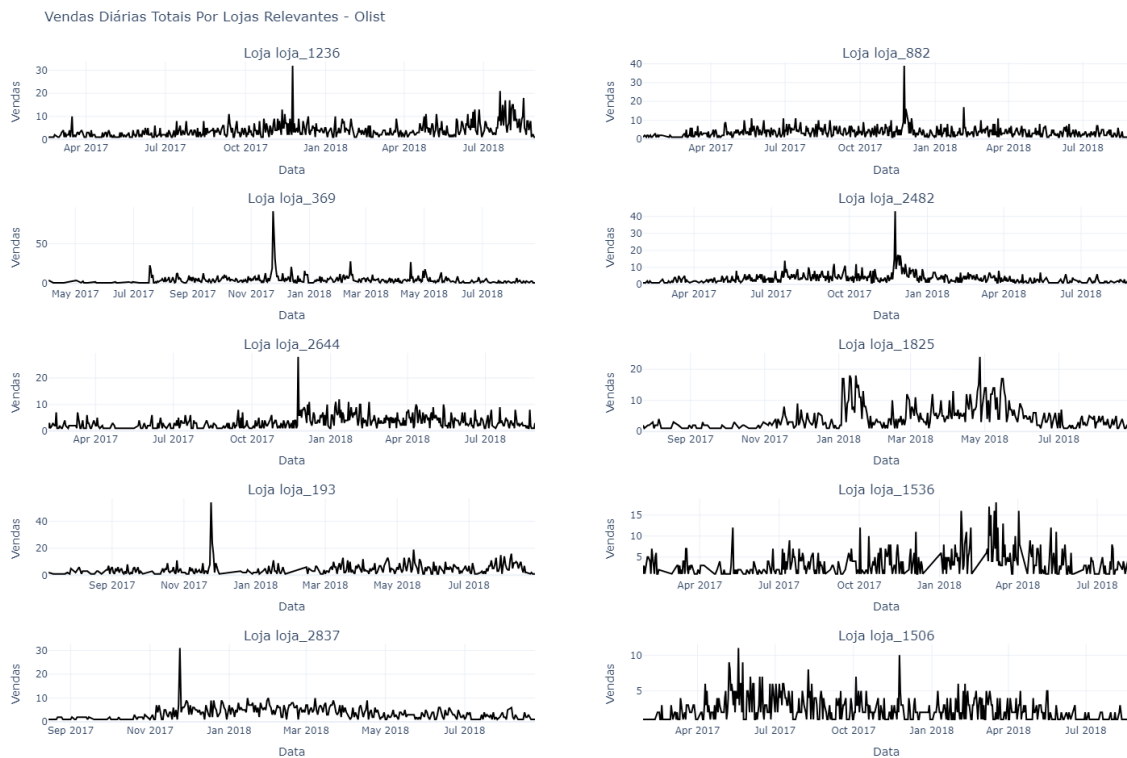
Figura 9 – Vendas Diárias Totais - Olist.



Fonte: Autor

dados foi tomada para preservar a integridade das análises e evitar que a presença de informações inconsistentes prejudicasse a performance dos algoritmos ou comprometesse a confiabilidade das métricas obtidas.

Todas as bases foram agregadas em nível semanal com o objetivo de reduzir os ruídos presentes nos dados e facilitar a generalização dos modelos preditivos. Essa agregação semanal permitiu simplificar as análises e tornar os padrões mais claros, beneficiando a capacidade dos modelos de capturar tendências e sazonalidades relevantes. A Figura 12 exemplifica, através da base da favorita, a série temporal de vendas totais semanais.

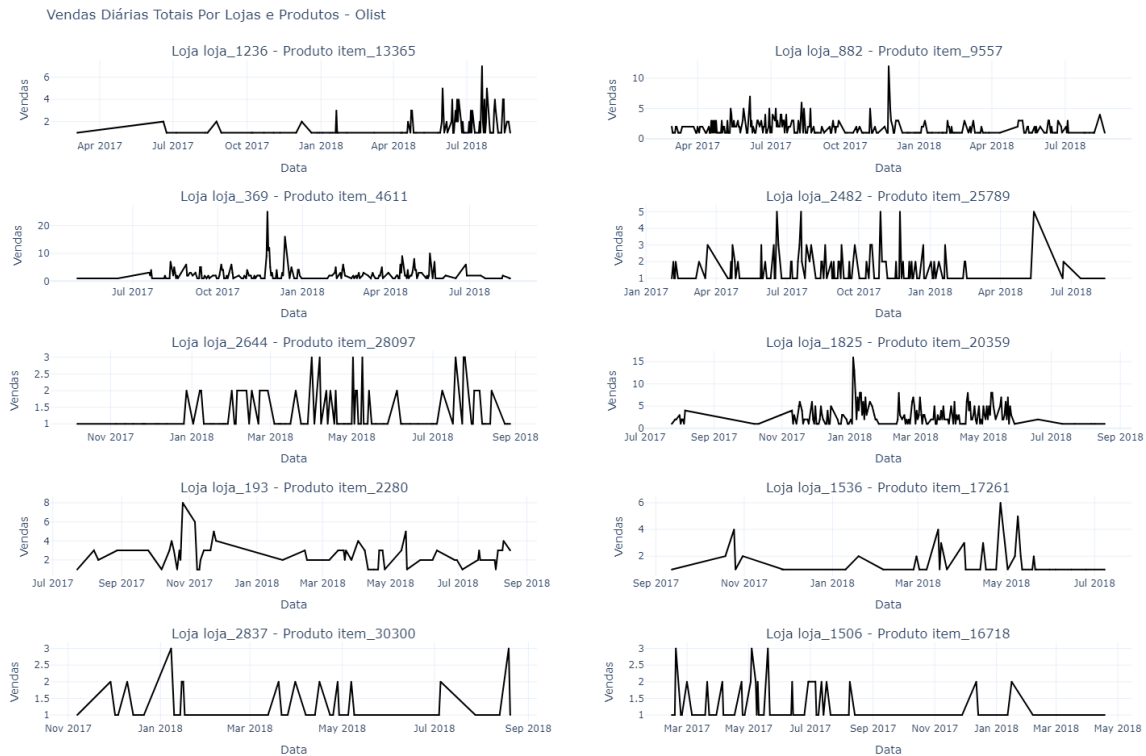
Figura 10 – Vendas Diárias Totais por Loja - Olist.

Fonte: Autor

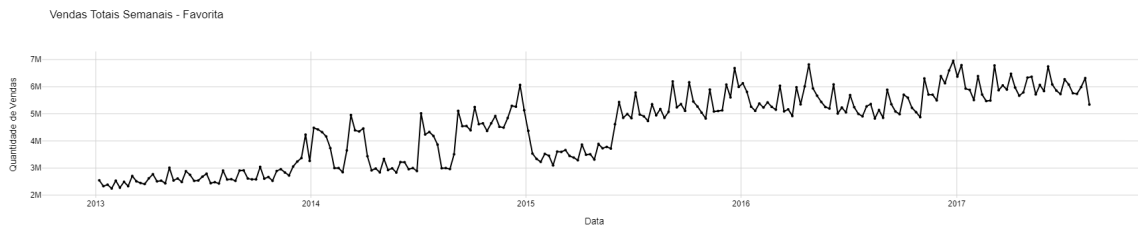
As bases possuem informações, como datas (agora consolidadas em semanas), lojas, produtos e respectivas quantidades de vendas. Essa estrutura possibilita uma abordagem flexível para realizar previsões em diferentes níveis de granularidade. Por exemplo, para prever as vendas gerais de uma loja, basta somar as vendas semanais de todos os itens associados àquela loja. Essa estratégia não apenas simplifica o processamento dos dados, mas também mantém a integridade das informações necessárias para gerar ideias valiosas em múltiplos contextos de previsão.

Ademais, para garantir uma avaliação robusta do desempenho dos modelos, foi realizada a separação das bases de dados em conjuntos de treino e teste utilizando uma divisão padrão de 70% para treino e 30% para teste. Essa abordagem é amplamente utilizada em projetos de modelagem preditiva, pois permite que os modelos sejam treinados com uma quantidade significativa de dados, enquanto o conjunto de teste, mantido separado, é usado para avaliar a capacidade de generalização do modelo em dados não vistos. Essa divisão foi aplicada uniformemente a todas as bases utilizadas no estudo, assegurando consistência nos experimentos e possibilitando uma comparação direta e justa entre os diferentes modelos implementados.

Em resumo, o pré-processamento foi uma etapa essencial para ajustar e refinar os dados antes da aplicação dos modelos, garantindo que as bases estivessem consistentes, padronizadas e prontas para fornecer resultados precisos e robustos. O cuidado dedicado a

Figura 11 – Vendas Diárias Totais por Loja e Produto - Olist.

Fonte: Autor

Figura 12 – Vendas Semanais Totais - Favorita.

Fonte: Autor

essa etapa contribui para o sucesso das análises subsequentes e reforça a importância de um pré-processamento bem conduzido no ciclo de vida de projetos envolvendo aprendizado de máquina e análise de dados.

4.5 Implementação dos Modelos

4.5.1 ARIMA

A implementação do modelo ARIMA foi conduzida utilizando uma abordagem automatizada para a seleção dos parâmetros ideais, aproveitando as capacidades do pacote “pmdarima”. Esse modelo é amplamente utilizado na análise de séries temporais devido à

sua capacidade de modelar padrões de tendência, sazonalidade e autocorrelação, sendo adequado para os dados analisados.

Para a configuração do ARIMA, foi adotado um modelo sazonal com periodicidade anual, considerando a natureza semanal dos dados. A seleção dos parâmetros foi realizada de maneira automática, o que garantiu a identificação de combinações ótimas para capturar as características específicas da série, como ciclos sazonais e variações ao longo do tempo.

Em resumo, o ARIMA foi implementado de forma estratégica, utilizando técnicas automatizadas e adaptadas à natureza sazonal e à granularidade dos dados, garantindo previsões eficientes e de alta qualidade.

4.5.2 Prophet

A implementação do modelo *Prophet* foi conduzida com o objetivo de explorar sua eficácia na previsão de séries temporais em dados de vendas no varejo. Este modelo é amplamente reconhecido por sua facilidade de uso, flexibilidade e desempenho robusto ao lidar com padrões sazonais e tendências.

A configuração inicial do *Prophet* foi projetada para capturar sazonalidades anuais de maneira multiplicativa, refletindo o impacto proporcional das variações sazonais nos dados de vendas. A sazonalidade semanal e diária foi desativada, já que os dados estavam agregados em granularidade semanal, tornando essas periodicidades irrelevantes para a análise. Essa escolha simplificou o modelo sem comprometer sua capacidade de representar padrões sazonais significativos.

Além disso, o parâmetro de sensibilidade a mudanças foi ajustado para moderar a detecção de pontos de inflexão na tendência, permitindo capturar alterações estruturais nos dados sem superajustar o modelo. A largura do intervalo de confiança foi configurada para 95%, oferecendo estimativas preditivas mais abrangentes e úteis para tomadas de decisão estratégicas no contexto de planejamento e gerenciamento de estoques.

Assim, a implementação do *Prophet* demonstrou-se simples e com apenas alguns ajustes de hiperparâmetros.

4.5.3 LSTM

A implementação do modelo LSTM foi conduzida utilizando a biblioteca *TensorFlow/Keras*, amplamente reconhecida por sua eficiência e versatilidade no desenvolvimento de redes neurais profundas.

O modelo foi construído de forma sequencial, começando com uma camada LSTM inicial configurada para retornar sequências, permitindo a propagação de informações temporais para a camada subsequente. Em seguida, uma segunda camada LSTM foi

adicionada sem retorno de sequências, sintetizando os dados temporais em uma única representação densa. Finalmente, uma camada densa foi utilizada para produzir a previsão final, representando o valor da variável-alvo para o próximo período.

Para a otimização do modelo, utilizou-se o algoritmo Adam, um método amplamente adotado devido à sua eficácia em acelerar a convergência durante o treinamento. A função de perda escolhida foi o erro quadrático médio (`mean_squared_error`), que é apropriada para tarefas de regressão e penaliza desvios maiores, incentivando previsões mais precisas.

Para evitar o sobreajuste (*overfitting*) e aprimorar a generalização do modelo, foi implementada a técnica de *early stopping*, monitorando a função de perda durante o treinamento. Caso a perda não apresentasse melhoria após 50 épocas consecutivas, o treinamento era interrompido, e os pesos da melhor iteração eram restaurados. Essa abordagem garante um treinamento eficiente e também evita desperdício de recursos computacionais.

Assim, a implementação do LSTM, utilizando a biblioteca *TensorFlow/Keras*, destaca-se pela capacidade de modelar dados temporais de forma robusta e eficiente.

4.5.4 CNN

A implementação do modelo CNN também foi realizada utilizando a biblioteca *TensorFlow/Keras*. Apesar de serem mais comuns em aplicações de visão computacional, as CNNs também têm demonstrado grande eficácia em tarefas envolvendo séries temporais, pois conseguem extrair características locais relevantes e padrões complexos nos dados.

O modelo desenvolvido possui uma arquitetura sequencial composta por duas camadas convolucionais seguidas de normalização, *dropout* e *pooling*, além de camadas densas para processamento final. A primeira camada convolucional foi configurada com 32 filtros, um *kernel* de tamanho 3 e função de ativação ReLU, permitindo a extração inicial de padrões. Para evitar sobreajuste e melhorar a generalização, aplicou-se regularização L2 e uma camada de normalização *batch*, além de uma camada de *dropout* com taxa de 30%. A operação de *max pooling* foi utilizada para reduzir a dimensionalidade e manter as características mais relevantes. A segunda camada convolucional segue a mesma estrutura, mas utiliza 64 filtros para capturar padrões mais profundos e complexos nos dados.

Após as operações convolucionais, o modelo utiliza uma camada *flatten* para transformar os dados em um vetor unidimensional, preparando-os para as camadas densas. A camada totalmente conectada foi configurada com 64 neurônios e função de ativação ReLU, sendo seguida por normalização *batch* e *dropout*, reforçando a robustez do modelo. Por fim, a camada de saída, composta por um único neurônio, gera as previsões.

O modelo integra estratégias avançadas de regularização e prevenção de sobreajuste, como a regularização L2, que penaliza pesos excessivos, e *dropout*, que inativa neurônios

aleatórios durante o treinamento, promovendo maior generalização. A normalização *batch* também foi aplicada para estabilizar o processo de treinamento, normalizando as ativações em cada camada.

Essa arquitetura foi projetada para identificar padrões locais e temporais nos dados de maneira eficiente, explorando a capacidade das CNNs de lidar com interações complexas. Além disso, o design modular permite que o modelo seja escalado para grandes volumes de dados, mantendo a eficiência computacional. Os resultados esperados incluem previsões mais precisas e maior robustez em cenários de dados ruidosos, consolidando a abordagem CNN como uma alternativa poderosa em comparação com outros modelos analisados.

4.6 Métricas Coletadas

As métricas de desempenho RMSE e MAPE foram calculadas exclusivamente sobre o conjunto de teste, que corresponde a 30% dos dados em todas as bases utilizadas. Essas métricas permitem uma análise detalhada tanto da precisão preditiva quanto das características dos modelos em cenários reais. A avaliação foi realizada em todas as séries temporais relevantes, abrangendo tanto produtos quanto lojas, para garantir uma visão ampla e representativa do desempenho.

Além disso, foram coletados o tempo de treinamento e de previsão de todas as séries temporais e modelos. Essas métricas fornecem uma base sólida para a comparação entre os diferentes algoritmos, de modo a analisar a eficiência computacional em contextos práticos.

5 Análise e discussão dos resultados

O objetivo deste capítulo é realizar uma análise detalhada dos resultados obtidos pelas previsões dos modelos aplicados aos conjuntos de teste, avaliando sua eficácia e desempenho.

5.1 Resultado de Todas as Lojas e Itens

O resultado geral consiste na agregação das vendas realizadas em todas as lojas e produtos, organizadas por data, com o objetivo de identificar padrões e prever o comportamento global das vendas nas bases analisadas. Essa abordagem permite compreender tendências gerais, independentemente das especificidades de cada loja ou produto individual, fornecendo uma visão consolidada do desempenho das vendas ao longo do tempo.

A previsão realizada no conjunto de teste da base da Favorita destacou a superioridade do modelo LSTM, de acordo com a Tabela 1, que apresentou o desempenho mais robusto entre os métodos avaliados, alcançando as melhores métricas preditivas.

Tabela 1 – Tabela de Resultados Gerais - Favorita.

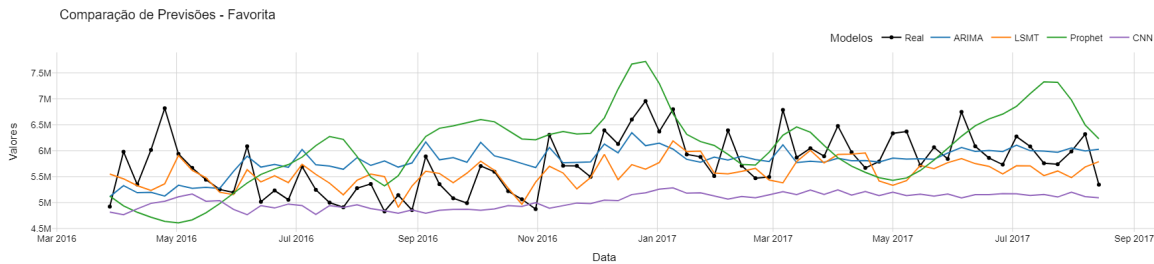
Modelo	RMSE (#)	MAPE (%)	Base
ARIMA	501644	7,17	Favorita
CNN	850827	11,85	Favorita
LSTM	497659	6,19	Favorita
Prophet	826610	12,21	Favorita

Fonte: Autor

Ademais, essa superioridade pode ser observada de forma clara na Figura 13. Nela, é possível perceber que os modelos LSTM e ARIMA foram os que mais se aproximaram da série temporal original, capturando com maior precisão suas variações e tendências ao longo do tempo.

No conjunto de teste da base M5, de acordo com a Tabela 2, o modelo LSTM apresentou um desempenho superior, com uma melhoria aproximada de 35% em relação ao ARIMA, que foi o segundo modelo mais bem-sucedido.

Além disso, na Figura 14 é possível analisar graficamente, nota-se que o LSTM consegue acompanhar muito bem o comportamento sazonal da série temporal original. Em contrapartida, os outros modelos não demonstraram o mesmo desempenho, apresentando maior dificuldade em reproduzir os padrões sazonais da série.

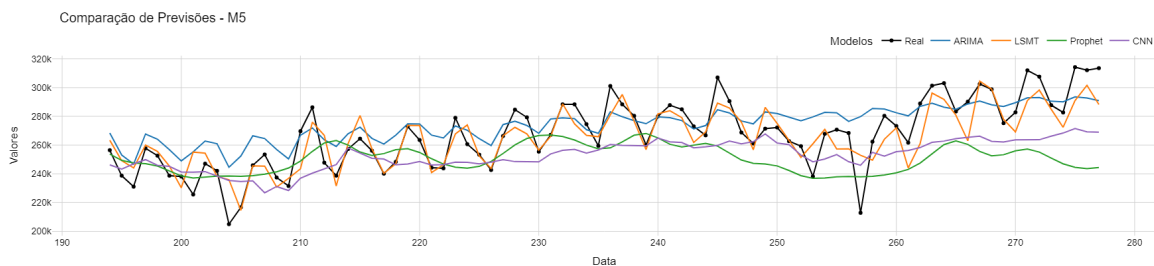
Figura 13 – Previsões do Conjunto de Teste - Favorita.

Fonte: Autor

Tabela 2 – Tabela de Resultados Gerais - M5.

Modelo	RMSE (#)	MAPE (%)	Base
ARIMA	16910	5,39	M5
CNN	22826	6,64	M5
LSTM	12593	3,51	M5
Prophet	27434	7,96	M5

Fonte: Autor

Figura 14 – Previsões do Conjunto de Teste - M5.

Fonte: Autor

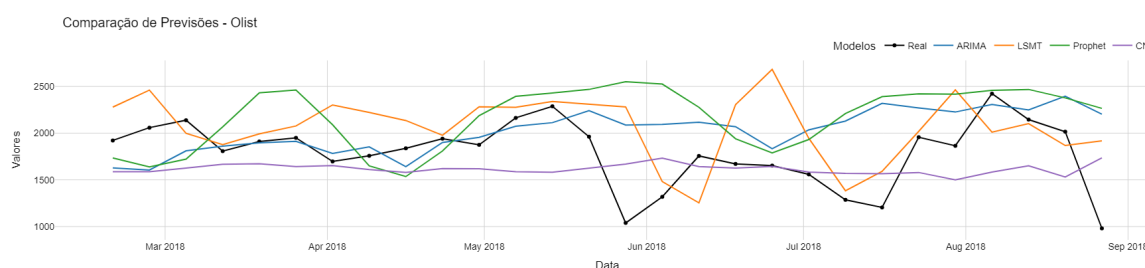
A base da *Olist* não seguiu o mesmo comportamento das outras bases em relação ao desempenho dos modelos. De acordo com a Tabela de Resultados 3, o modelo que desempenhou melhor foi a CNN. No entanto, é notável que a escala da métrica MAPE aumentou bastante, já que o melhor modelo apresentou 20%. Dessa forma, nenhum dos modelos teve um desempenho excepcionalmente bom.

Ademais, ao observar a Figura 15, nota-se que nenhum dos modelos conseguiu acompanhar adequadamente a série temporal original. Embora a CNN tenha apresentado o menor valor de MAPE, ela não conseguiu generalizar para capturar o comportamento da série de teste da *Olist*. O modelo LSTM, por sua vez, demonstrou certa capacidade de acompanhar o padrão da curva, mas apresentou previsões com valores sistematicamente mais altos do que os observados.

Tabela 3 – Tabela de Resultados Gerais - Olist.

Modelo	RMSE (#)	MAPE (%)	Base
ARIMA	488	25,17	Olist
CNN	414	20,84	Olist
LSTM	475	23,90	Olist
Prophet	615	33,36	Olist

Fonte: Autor

Figura 15 – Previsões do Conjunto de Teste - Olist.

Fonte: Autor

5.2 Resultado por Loja

Nesta seção, os testes foram realizados com as vendas agrupadas por lojas, restringindo a análise às 10 lojas mais relevantes de cada base. Esse agrupamento teve como objetivo concentrar os testes nas lojas que apresentaram maior representatividade em termos de volume de vendas, garantindo uma análise focada nos casos mais significativos de cada conjunto de dados.

As lojas da base da Favorita apresentaram padrões relativamente homogêneos, o que contribuiu para um desempenho consistente dos modelos. Entre eles, o LSTM destacou-se ao alcançar a melhor média nas métricas de avaliação, registrando um RMSE de 21.441 e um MAPE de 7,16. Embora os demais modelos não tenham apresentado resultados significativamente inferiores, nenhum deles conseguiu se aproximar do desempenho superior alcançado pelo LSTM.

Tabela 4 – Tabela de Resultado Médio das Lojas - Favorita.

Modelo	RMSE (#)	MAPE (%)	Base
ARIMA	24405	9,17	Favorita
CNN	30360	10,10	Favorita
LSTM	21441	7,16	Favorita
Prophet	36344	13,74	Favorita

Fonte: Autor

Além disso, a Figura 16 ilustra as séries temporais geradas por cada modelo para cada loja. A partir dessa visualização, torna-se evidente a superioridade do modelo LSTM, que demonstrou uma maior capacidade de acompanhar o comportamento da série temporal original de forma consistente em relação aos demais modelos.

Figura 16 – Previsões do Conjunto de Teste por Loja - Favorita.



Fonte: Autor

No caso da base M5, todas as lojas foram incluídas no teste, considerando que a base contém apenas 10 lojas. Apesar do número reduzido, os comportamentos das lojas apresentam variações significativas, com algumas exibindo um crescimento repentino acentuado, enquanto outras apresentam uma leve tendência de crescimento ao longo do tempo. Nesse cenário, o modelo LSTM destacou-se novamente, alcançando o maior sucesso entre os modelos avaliados. Conforme ilustrado na Tabela 5, o LSTM obteve métricas significativamente superiores, sendo aproximadamente duas vezes melhor que o segundo melhor modelo, as CNNs.

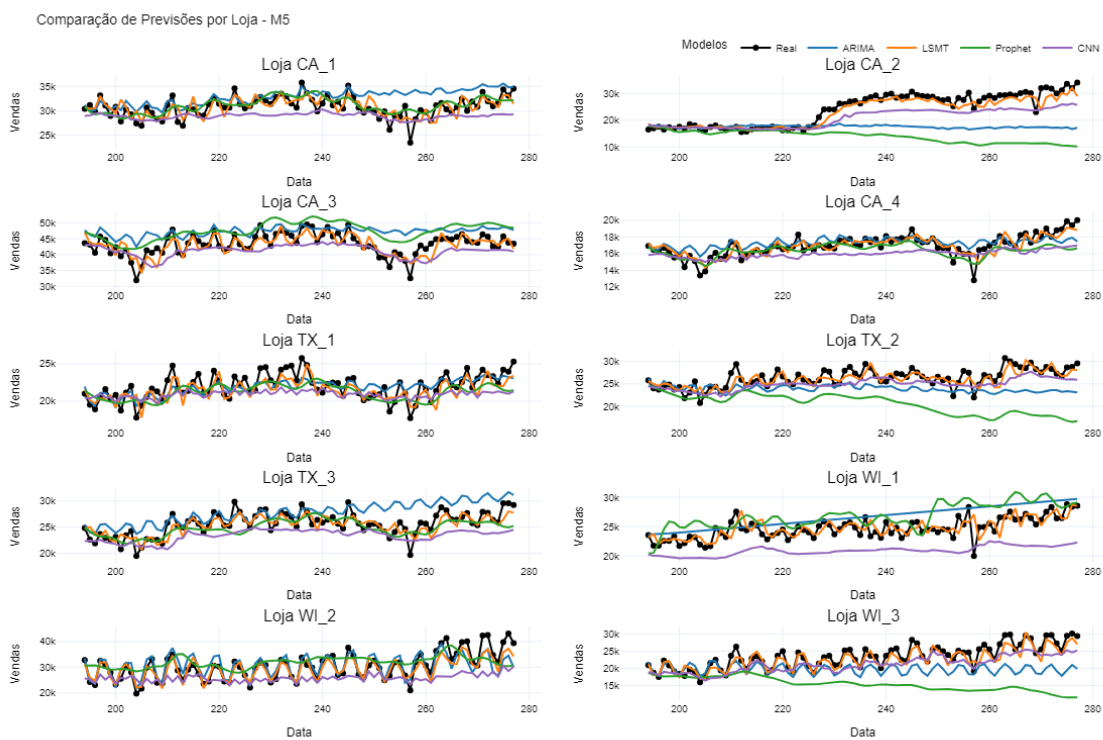
Ademais, ao analisar a Figura 17, fica evidente a superioridade do modelo LSTM em acompanhar a série temporal de teste original. Destaca-se, em particular, o caso da loja “CA_”, que apresenta um crescimento repentino nos dados, mas ainda assim foi bem representado pelo LSTM.

De acordo com a Tabela 6, todos os modelos apresentaram desempenho insatisfatório para as lojas na base da *Olist*, especialmente ao analisar pela métrica MAPE. Esse resultado foi influenciado pela presença de *outliers* gerados pela explosão da métrica MAPE, que

Tabela 5 – Tabela de Resultado Médio das Lojas - M5.

Modelo	RMSE (#)	MAPE (%)	Base
ARIMA	3525	10,31	M5
CNN	3218	9,44	M5
LSTM	1696	4,86	M5
Prophet	4774	14,43	M5

Fonte: Autor

Figura 17 – Previsões do Conjunto de Teste por Loja - M5.

Fonte: Autor

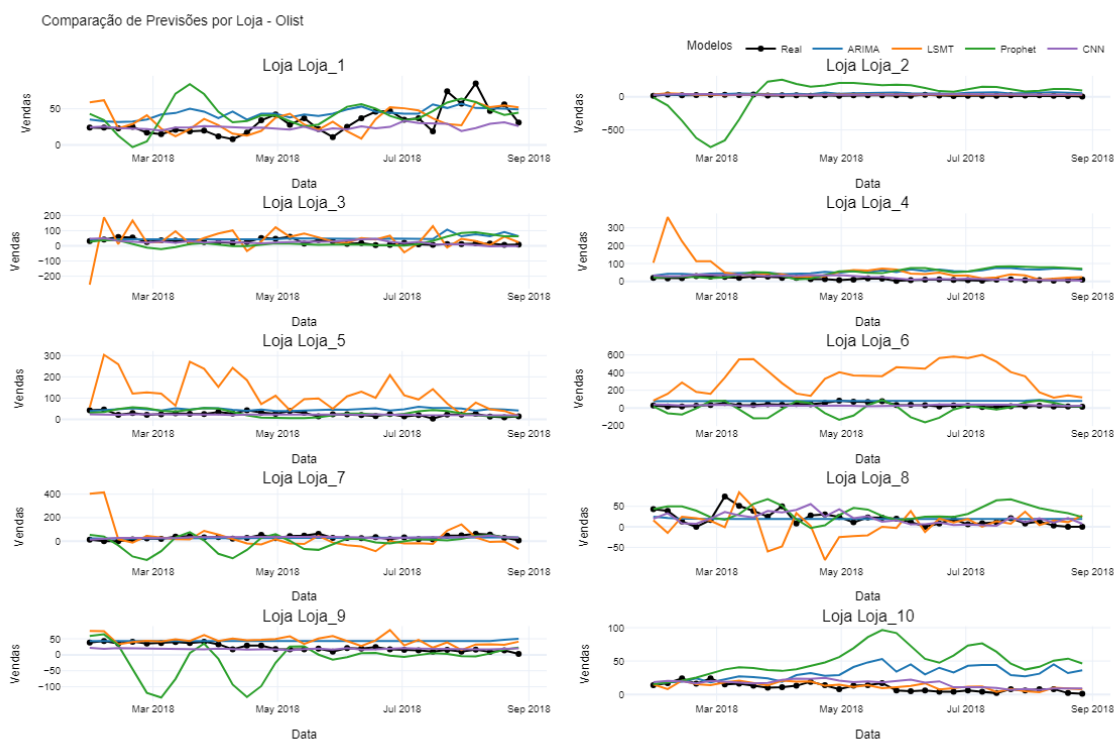
distorceu a média devido aos valores extremamente elevados. No entanto, ao considerar a métrica RMSE, as CNNs demonstraram desempenho superior aos demais modelos, com resultados aproximadamente duas vezes melhores que os do segundo melhor modelo, o ARIMA. Por outro lado, o LSTM apresentou o pior desempenho, evidenciando dificuldades em generalizar para uma base com volume de dados reduzido.

Na Figura 18, é possível observar com maior clareza o motivo da distorção na métrica MAPE. Os modelos LSTM e Prophet geraram séries temporais que divergem significativamente das séries temporais originais.

Tabela 6 – Tabela de Resultado Médio das Lojas - Olist.

Modelo	RMSE (#)	MAPE (%)	Base
ARIMA	28.3	$1,7 \times 10^{17}$	Olist
CNN	13.9	$1,3 \times 10^{17}$	Olist
LSMT	84.1	7.3×10^{17}	Olist
Prophet	69.7	$2,9 \times 10^{17}$	Olist

Fonte: Autor

Figura 18 – Previsões do Conjunto de Teste por Loja - Olist.

Fonte: Autor

5.3 Resultado por Loja e Produto

Nesta seção, os testes foram conduzidos considerando as vendas agrupadas pelos itens de maior relevância em cada loja. A análise foi restrita às 10 lojas mais representativas de cada base, assim como aos dois itens mais vendidos em cada uma delas. Esse agrupamento teve como objetivo priorizar os itens de maior representatividade em volume de vendas, assegurando uma abordagem focada nos casos mais significativos de cada conjunto de dados. Além disso, é importante destacar que, por questões de simplificação visual, as visualizações gráficas apresentam apenas o produto mais relevante de cada loja. No entanto, para uma análise mais completa dos resultados médios, foram considerados os dois produtos mais relevantes de cada loja.

Os produtos da base da Favorita exibiram comportamentos distintos entre si, evidenciando características únicas em suas séries temporais. Nesse cenário, o modelo LSTM destacou-se novamente como o de melhor desempenho, demonstrando sua capacidade de lidar com a variabilidade dos dados. No entanto, o modelo baseado em CNNs também apresentou resultados competitivos, ficando a apenas cerca de 1% de diferença nas métricas em relação ao LSTM.

Tabela 7 – Tabela de Resultado Médio dos Produtos por Loja - Favorita.

Modelo	RMSE(#)	MAPE(%)	Base
ARIMA	407	18.20	Favorita
CNN	355	13.21	Favorita
LSTM	324	12.26	Favorita
Prophet	479	20.96	Favorita

Fonte: Autor

Além disso, como ilustrado na Figura 19, os modelos LSTM e CNN demonstraram excelente capacidade em acompanhar as séries temporais originais, capturando de maneira consistente os padrões presentes nos dados. Por outro lado, os modelos *Prophet* e ARIMA apresentaram desempenhos significativamente inferiores, com previsões que se distanciaram bastante do comportamento real das séries. Em particular, o ARIMA, em alguns casos, gerou previsões em formato de retas, indicando uma clara limitação na generalização do modelo frente à complexidade dos dados analisados.

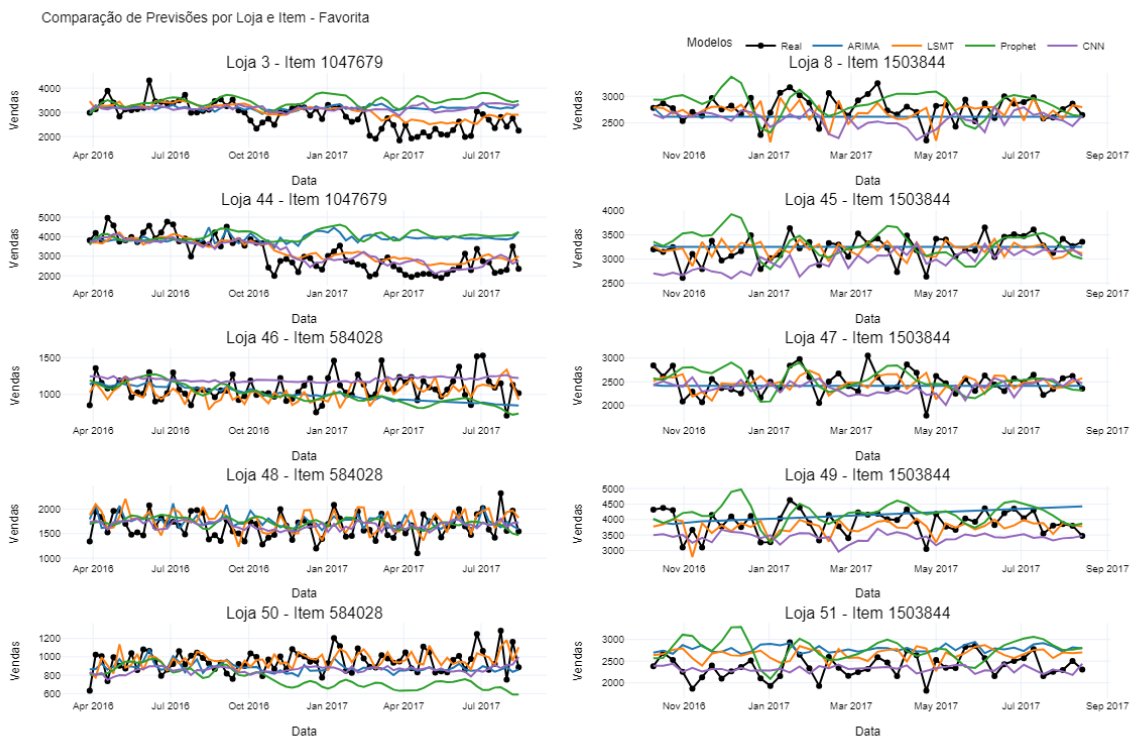
Alguns produtos da base M5 apresentaram padrões bastante semelhantes, enquanto outros exibiram comportamentos significativamente distintos, evidenciando a diversidade na dinâmica de vendas. Além disso, ocorreu novamente o problema de explosão nos valores da métrica MAPE, tornando essencial uma avaliação mais robusta do desempenho utilizando a métrica RMSE. Nesse contexto, o modelo LSTM destacou-se com o melhor desempenho médio, conforme ilustrado na Tabela 8, enquanto as CNNs mantiveram sua posição como o segundo melhor modelo.

Tabela 8 – Tabela de Resultado Médio dos Produtos por Loja - M5.

Modelo	RMSE(#)	MAPE(%)	Base
ARIMA	135.75	9.9×10^{18}	M5
CNN	108.90	7.4×10^{18}	M5
LSTM	92.35	4.3×10^{18}	M5
Prophet	132.85	7.8×10^{18}	M5

Fonte: Autor

Além disso, ao analisar graficamente a Figura 20, observa-se que o modelo LSTM reproduz de forma eficiente o comportamento da curva original, refletindo com precisão

Figura 19 – Previsões do Conjunto de Teste dos Produtos por Loja - Favorita.

Fonte: Autor

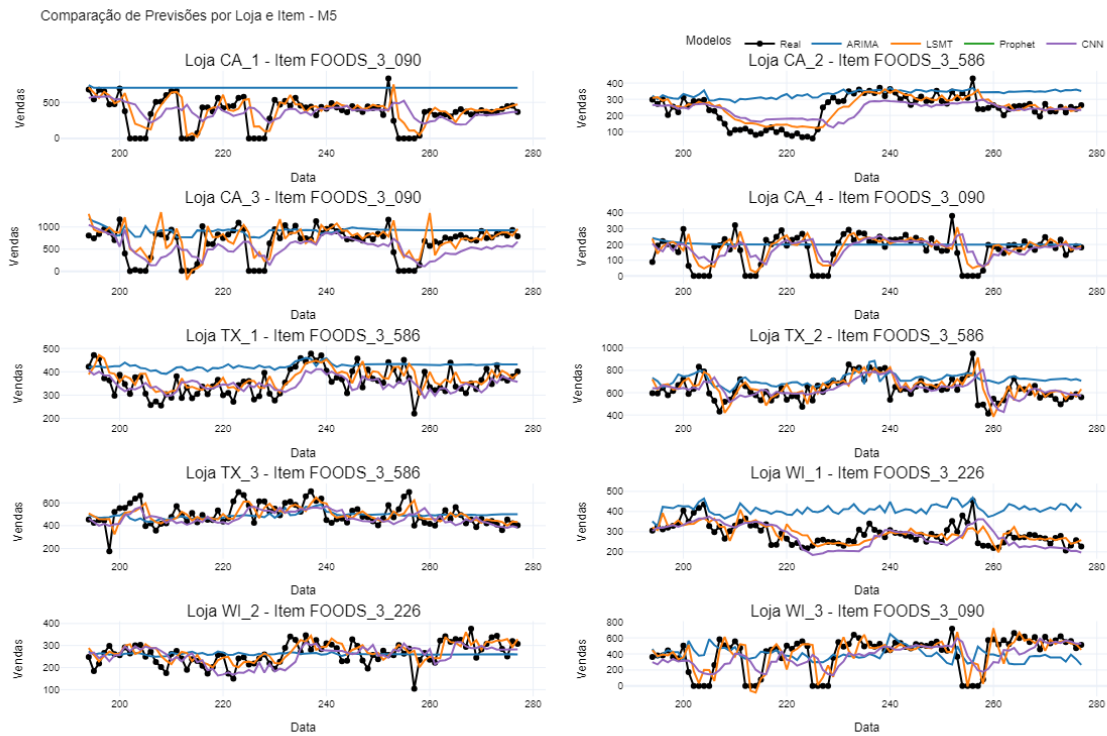
quedas e aumentos substanciais nas vendas. Por outro lado, o modelo ARIMA demonstra limitações em sua capacidade de generalização, frequentemente gerando retas ou padrões que pouco correspondem ao comportamento real da série temporal.

A base de dados da Olist, com poucos registros e baixo volume de vendas, gera distorções na métrica MAPE devido à frequência de valores próximos de zero. Nesse cenário, utilizando a métrica RMSE, as CNNs se destacaram com o melhor desempenho, superando em quase duas vezes o ARIMA, que foi o segundo melhor modelo. Em contrapartida, o LSTM e o *Prophet* apresentaram desempenhos significativamente inferiores, sendo aproximadamente 8 e 32 vezes piores que o modelo mais eficiente, respectivamente. Isso evidencia que esses métodos não são adequados para bases de dados com poucas informações disponíveis.

Tabela 9 – Tabela de Resultado Médio dos Produtos por Loja - Olist.

Modelo	RMSE(#)	MAPE(%)	Base
ARIMA	4.75	7.1×10^{17}	Olist
CNN	2.55	3.3×10^{17}	Olist
LSTM	19.65	1.2×10^{18}	Olist
Prophet	80.30	1.4×10^{19}	Olist

Fonte: Autor

Figura 20 – Previsões do Conjunto de Teste dos Produtos por Loja - M5.

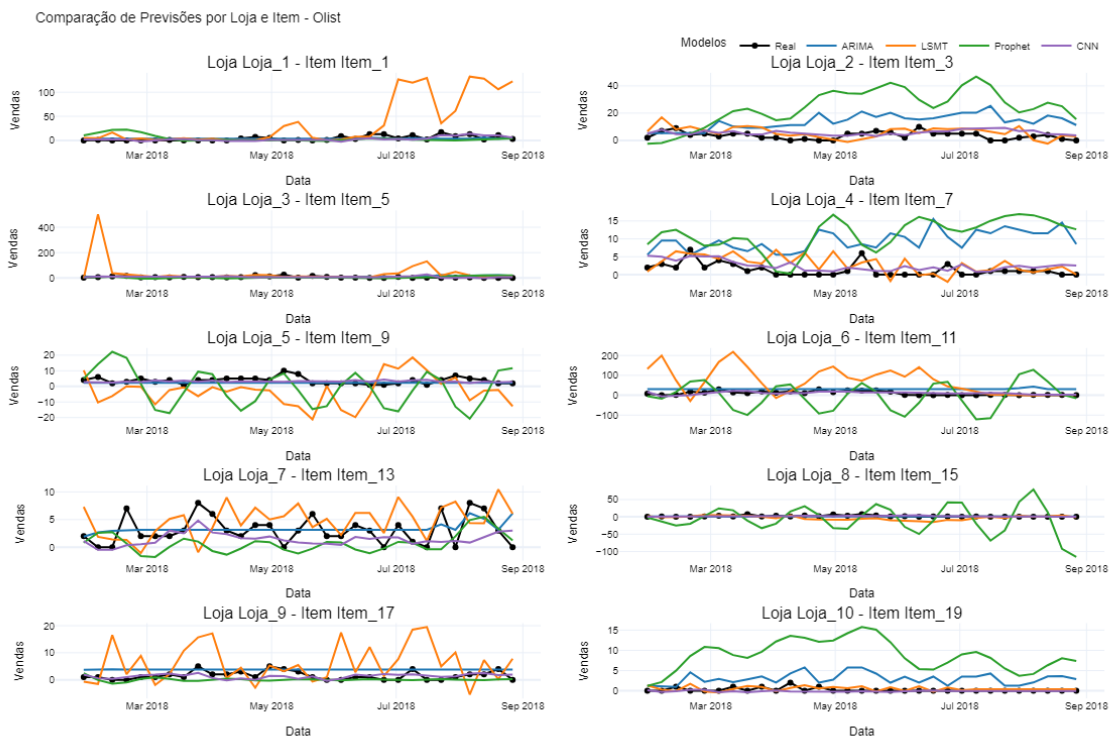
Fonte: Autor

Além disso, a Figura 21 ilustra que, para séries temporais com poucos dados, os modelos LSTM e *Prophet* apresentam um desempenho insatisfatório, divergindo significativamente da amplitude da série temporal original. Em contrapartida, o modelo CNN demonstra um bom ajuste, acompanhando de forma consistente as flutuações, mesmo que sejam de baixa intensidade.

5.4 Resultado Temporal

Esta seção tem como objetivo analisar e comparar o tempo necessário para o treinamento e a execução das previsões de cada modelo. Para isso, foi realizado o somatório dos tempos individuais de treinamento e previsão de todas as séries temporais, e os resultados foram agrupados por modelo. Posteriormente, calculou-se a porcentagem de contribuição de cada modelo em relação ao tempo total, permitindo identificar quais modelos apresentaram maior eficiência em termos de tempo. Essa análise é fundamental para avaliar a viabilidade prática de cada abordagem, especialmente em cenários que demandam previsões rápidas ou treinamentos frequentes.

A Figura 22 ilustra um *treemap* que permite visualizar a comparação percentual do tempo total de treinamento de cada modelo. Nessa análise, destaca-se que o ARIMA consumiu um tempo de treinamento consideravelmente superior aos demais modelos testa-

Figura 21 – Previsões do Conjunto de Teste dos Produtos por Loja - Olist.

Fonte: Autor

dos. Esse comportamento pode ser explicado pela abordagem adotada pelo 'pmdarima', que realiza uma busca iterativa para identificar os valores ótimos de seus parâmetros. Esse processo tem como objetivo ajustar o modelo da forma mais precisa possível às características específicas da série temporal, mas acaba demandando um custo computacional mais elevado. Em contrapartida, outros modelos, como o LSTM e as CNNs, embora também utilizem recursos computacionais consideráveis, apresentaram tempos de treinamento mais equilibrados em relação ao ARIMA, refletindo a eficiência de suas arquiteturas em lidar com séries temporais complexas. Além disso, os modelos LSTM e CNN empregam *early stopping*, reduzindo o tempo de treinamento e o consumo de recursos computacionais.

Por outro lado, a Figura 23 ilustra a comparação dos tempos de previsão utilizando a mesma abordagem de visualização. Nesse contexto, observa-se que o LSTM apresentou um tempo de previsão significativamente maior em comparação aos outros modelos, evidenciando o custo computacional mais elevado associado ao processo de previsão desse modelo. Isso reflete a complexidade intrínseca das redes neurais recorrentes, especialmente em arquiteturas como o LSTM, que processam sequências de dados de forma iterativa. Em contrapartida, o ARIMA destacou-se como o modelo com o menor tempo de previsão, resultado esperado, dado que é um modelo estatístico mais simples em comparação às abordagens baseadas em aprendizado profundo, o que torna suas previsões computacionalmente mais eficientes.

Figura 22 – *Treemap* de Tempo Total de Treino por Modelo.

Fonte: Autor

Figura 23 – *Treemap* de Tempo Total de Previsão por Modelo.

Fonte: Autor

5.5 Análise Geral de Resultado

Para uma avaliação mais abrangente e detalhada da comparação entre os modelos, torna-se essencial analisar diretamente as métricas gerais consolidadas de todas as séries temporais. Essa abordagem permite identificar com maior precisão o desempenho global de cada modelo, considerando o comportamento médio das previsões.

Para identificar o modelo com melhor desempenho, foram selecionadas as 100 previsões mais precisas com base na métrica MAPE. Em seguida, foi realizada a contagem de ocorrências de cada modelo entre essas melhores previsões. Conforme indicado na Tabela 10, o modelo LSTM destacou-se, apresentando a maior quantidade de previsões entre as mais precisas.

Devido ao problema de explosão na métrica MAPE, foi realizada uma análise

Tabela 10 – Contagem de Modelos nas 100 Melhores Previsões.

Modelo	Contagem
LSTM	33
ARIMA	28
CNN	23
Prophet	16

Fonte: Autor

considerando apenas as 100 melhores séries temporais, em vez de utilizar todas as séries disponíveis. Dessa forma, foi possível avaliar o desempenho dos modelos de maneira mais confiável. Conforme indicado na Tabela 11, o LSTM apresentou o melhor desempenho, superando o segundo colocado, o ARIMA, por uma margem de aproximadamente 15%.

Tabela 11 – MAPE Médio das 100 Melhores Previsões por Modelo.

Modelo	MAPE (%)
ARIMA	8.2
CNN	8.7
LSTM	7.1
Prophet	8.6

Fonte: Autor

Além disso, ao realizar a análise utilizando a métrica RMSE, desta vez considerando todas as previsões realizadas, é possível observar, conforme ilustrado na Tabela 12, que o LSTM novamente se destacou como o modelo mais eficaz. Esse resultado reafirma sua capacidade de capturar padrões complexos e dinâmicos nas séries temporais analisadas, independentemente das características específicas de cada base. O LSTM demonstrou consistência em suas previsões.

Tabela 12 – RMSE Médio por Modelo.

Modelo	RMSE (#)
ARIMA	8705
CNN	13111
LSTM	8082
Prophet	13768

Fonte: Autor

Conforme a Tabela 13, é possível analisar o desempenho médio dos modelos em relação à métrica RMSE para cada base de dados. Observa-se que o modelo LSTM apresentou o melhor desempenho geral, destacando-se nas bases Favorita e M5, com os

menores valores de RMSE: 23.179 e 1.013, respectivamente. Esses resultados reafirmam a capacidade do LSTM de capturar padrões complexos em séries temporais, mesmo em cenários com comportamentos distintos. Por outro lado, na base *Olist*, o desempenho do LSTM foi inferior ao das CNNs, que obtiveram o menor RMSE (19), evidenciando que, para essa base específica, as CNNs foram mais eficazes. Isso pode ser explicado pela natureza das séries temporais da *Olist*, que pode ter favorecido a arquitetura convolucional na captura de padrões locais. Os modelos ARIMA e *Prophet*, embora tenham desempenho competitivo em alguns casos, apresentaram resultados menos satisfatórios na maioria das bases, com RMSEs consideravelmente maiores. Esses resultados reforçam que esses modelos estatísticos podem não ser os mais adequados para lidar com séries temporais complexas e de alta variabilidade.

Tabela 13 – RMSE Médio por Base e Modelo.

Base	Modelo	RMSE (#)
Favorita	ARIMA	24317
Favorita	CNN	37469
Favorita	LSTM	23179
Favorita	Prophet	38698
M5	ARIMA	1770
M5	CNN	1845
M5	LSTM	1013
M5	Prophet	2511
Olist	ARIMA	28
Olist	CNN	19
Olist	LSTM	55
Olist	Prophet	94

Fonte: Autor

Com base nas análises realizadas, conclui-se que o LSTM apresentou o melhor desempenho geral entre os modelos avaliados, evidenciando sua robustez e eficácia na previsão de demandas em diferentes bases de dados. O ARIMA destacou-se como o segundo melhor modelo, ficando logo atrás do LSTM em termos de desempenho. No entanto, as CNNs, embora não tenham demonstrado ser a melhor escolha na maioria dos cenários, mostraram-se particularmente adequadas para a base de dados da *Olist*, caracterizada por um volume reduzido de informações, onde alcançaram o melhor desempenho. O desempenho do *Prophet* ficou abaixo do esperado na maioria dos casos, e em algumas situações foi, inclusive, o modelo com o pior desempenho.

6 Conclusão

Concluimos que, ao longo deste trabalho, foi possível identificar o desempenho de diferentes modelos preditivos aplicados a séries temporais com características distintas, evidenciando suas forças e limitações. De forma geral, o modelo LSTM destacou-se como a abordagem mais robusta, apresentando o melhor desempenho preditivo na maioria das bases analisadas. Superando o segundo melhor modelo em 7% na métrica RMSE e em 13% ao considerar as 100 melhores previsões segundo a métrica MAPE. Sua capacidade de capturar padrões complexos e não lineares nas séries temporais consolidou sua posição como o modelo mais eficaz para as bases da Favorita e M5. Em especial na base M5, o LSTM destacou-se com um desempenho 43% superior ao do segundo colocado de acordo com o RMSE médio. No entanto, na base de dados da Olist, as CNNs demonstraram resultados superiores. Esse comportamento evidencia que o LSTM é particularmente adequado para situações em que há alta variabilidade e padrões temporais intrincados, enquanto as CNNs se mostraram eficientes em cenários mais específicos.

Por outro lado, o modelo estatístico ARIMA destacou-se pela sua consistência, ocupando a segunda posição em todas as análises gerais de desempenho dos modelos. Embora não tenha alcançado a precisão do LSTM, o ARIMA demonstrou ser uma opção robusta, especialmente devido à sua abordagem tradicional e à capacidade de lidar bem com dados estruturados e padronizados. Isso reforça sua relevância como uma alternativa confiável para previsões em cenários menos complexos.

O Prophet, por outro lado, apresentou dificuldades em acompanhar a complexidade das séries temporais analisadas, destacando-se como o modelo de pior desempenho em algumas análises gerais. Esse resultado evidencia que o Prophet pode não ser a opção mais adequada para lidar com dados mais complexos e dinâmicos, especialmente quando se exige maior precisão nas previsões.

Outro aspecto relevante a ser considerado é o custo computacional dos modelos, um fator decisivo na escolha da abordagem mais adequada para diferentes cenários. O ARIMA, por exemplo, apresentou o maior tempo de treinamento, devido à necessidade de otimização de seus parâmetros. Por outro lado, o LSTM destacou-se pelo maior tempo de previsão, uma consequência direta de sua arquitetura complexa e do processamento sequencial. Ainda assim, o LSTM justificou seu elevado custo computacional de previsão ao entregar resultados preditivos superiores na maioria das bases analisadas, tornando-se uma opção viável para aplicações em que a precisão supera a prioridade pela rapidez. Já o ARIMA demonstrou ser o modelo mais eficiente em termos de tempo de previsão, revelando-se uma alternativa interessante para cenários em que a agilidade é mais importante do que a

acurácia e o tempo de treino não importa.

No contexto da base da Olist, as CNNs apresentaram o melhor desempenho, destacando-se como uma alternativa promissora para cenários em que os dados possuem características menos complexas ou que demandam uma abordagem mais eficiente em termos computacionais. Essa flexibilidade torna as CNNs uma opção interessante para aplicações específicas, embora seu desempenho global ainda fique atrás do LSTM em bases mais complexas.

Dessa forma, o estudo reforça a importância de selecionar o modelo preditivo com base nas características da base de dados e nas necessidades específicas da aplicação. Essa análise comparativa contribui para o avanço no uso de modelos de aprendizado de máquina e estatísticos em previsões de séries temporais para o contexto econômico do varejo.

6.1 Limitações

A utilização da métrica MAPE apresentou restrições em bases de dados com poucos registros ou valores zerados, pois essas condições podem gerar distorções significativas nos resultados. Para contornar esse problema, foi necessário recorrer a métricas auxiliares, como o RMSE, que demonstrou maior robustez nesses cenários.

6.2 Trabalhos Futuros

Para trabalhos futuros, algumas direções podem ser exploradas para maximizar a eficiência das previsões de demanda no contexto do varejo. Primeiramente, o uso da métrica SMAPE (*Symmetric Mean Absolute Percentage Error*) em cenários onde a métrica MAPE se mostra limitada, especialmente em bases de dados com valores escassos ou zerados, pode ser necessário para ter uma análise robusta e interpretável, já que supera o problema da explosão com valores reais nulos ou próximos de zero.

Além disso, uma abordagem promissora seria a adoção de redes neurais baseadas em grafos, como o modelo DeepAR, desenvolvido por Salinas, Flunkert e Gasthaus (2019). Esse tipo de arquitetura se destaca por sua capacidade de capturar padrões temporais e sazonais em dados sequenciais de maneira eficiente, podendo oferecer melhorias significativas em precisão e robustez. Ademais, uma das direções mais promissoras é a exploração de arquiteturas baseadas em transformadores, com destaque para o *Temporal Fusion Transformer* (TFT).

Outro aspecto relevante a ser considerado em trabalhos futuros é a inclusão de informações adicionais que impactam diretamente a demanda, como dados sobre condições climáticas, indicadores econômicos e variações no preço dos combustíveis. A incorporação

dessas variáveis contextuais pode enriquecer os modelos preditivos, permitindo uma análise mais detalhada e precisa dos fatores que influenciam o comportamento de consumo.

Esses tópicos apontam possíveis caminhos promissores para pesquisas futuras, buscando melhorar a precisão, a eficiência e a aplicabilidade das soluções de previsão de demanda no contexto do varejo.

Referências

- ADHIKARI, R.; AGRAWAL, R. K. *An Introductory Study on Time Series Modeling and Forecasting*. 2013. Disponível em: <<https://arxiv.org/abs/1302.6613>>. Citado 2 vezes nas páginas 28 e 29.
- ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, v. 8, n. 1, p. 53, 2021. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-021-00444-8>>. Citado na página 32.
- AMORIM, T. et al. Utilizando rede lstm para predição de ações de vários setores econômicos da bolsa de valores. *Gestão.Org – Revista Eletrônica de Gestão Organizacional*, v. 21, p. 1–20, 2023. ISSN 1679-1827. Disponível em: <<https://periodicos.ufpe.br/revistas/index.php/gestaoorg/article/view/257384>>. Citado na página 30.
- BALLOU, R. H. *Gerenciamento da cadeia de suprimentos/logística empresarial*. 5. ed. [S.l.]: Bookman, 2006. Citado na página 21.
- BETTI, V. A. *ESTUDOS TEÓRICOS NA ESTIMAÇÃO EM PROCESSOS K-FACTOR GARMA VIA MODELO DE ESPAÇO DE ESTADOS UTILIZANDO FILTRO DE KALMAN*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul (UFRGS), 2012. Acessado em: 28 dez. 2024. Disponível em: <<https://lume.ufrgs.br/bitstream/handle/10183/140891/000893672.pdf?sequence=1>>. Citado na página 28.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. 1. ed. San Francisco: Holden-Day, 1970. Citado na página 27.
- CAMPOS, B. S. *TCC - Repositório do GitHub*. 2025. Acessado em: 3 fev. 2025. Disponível em: <<https://github.com/hiperbrainer/TCC>>. Citado na página 37.
- FACEBOOK. *Prophet | Forecasting at scale*. 2023. Disponível em: <<https://facebook.github.io/prophet/>>. Citado na página 29.
- FAVORITA, C. et al. *Corporación Favorita Grocery Sales Forecasting*. 2017. <<https://kaggle.com/competitions/favorita-grocery-sales-forecasting>>. Kaggle. Citado na página 38.
- FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, v. 36, n. 4, p. 193–202, 1980. Disponível em: <<https://doi.org/10.1007/BF00344251>>. Citado na página 31.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado 2 vezes nas páginas 31 e 32.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 0899-7667. Disponível em: <<https://doi.org/10.1162/neco.1997.9.8.1735>>. Citado na página 30.

- HOWARD, A. et al. *M5 Forecasting - Accuracy*. 2020. <<https://kaggle.com/competitions/m5-forecasting-accuracy>>. Kaggle. Citado na página 38.
- LUSTOSA, L. et al. *PLANEJAMENTO E CONTROLE DA PRODUÇÃO*. 4. ed. [S.l.]: Elsevier, 2008. Citado 2 vezes nas páginas 21 e 25.
- MATTA, C. E. d. et al. A comparative study of forecasting methods using real-life econometric series data. *Production*, Associação Brasileira de Engenharia de Produção, v. 31, p. e20210043, 2021. ISSN 0103-6513. Disponível em: <<https://doi.org/10.1590/0103-6513.20210043>>. Citado 2 vezes nas páginas 33 e 34.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. São Paulo: Edgard Blucher Ltda., 2018. Citado 2 vezes nas páginas 26 e 27.
- NASSERI, M. et al. Applying machine learning in retail demand prediction—a comparison of tree-based ensembles and long short-term memory-based deep learning. *Applied Sciences*, v. 13, n. 19, 2023. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/13/19/11112>>. Citado na página 35.
- NOVAES, A. G. *Logística e gerenciamento da cadeia de distribuição: estratégia, avaliação e operação*. 5. ed. [S.l.]: GEN LTC, 2021. Citado 3 vezes nas páginas 21, 25 e 26.
- OLAH, C. *Understanding LSTMs*. 2015. <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Accessed: 2024-12-30. Citado 2 vezes nas páginas 30 e 31.
- OLIST; SIONEK, A. *Brazilian E-Commerce Public Dataset by Olist*. Kaggle, 2018. Disponível em: <<https://www.kaggle.com/dsv/195341>>. Citado na página 39.
- OLIVEIRA, J. M.; RAMOS, P. Evaluating the effectiveness of time series transformers for demand forecasting in retail. *Mathematics*, v. 12, n. 17, 2024. ISSN 2227-7390. Disponível em: <<https://www.mdpi.com/2227-7390/12/17/2728>>. Citado na página 36.
- RAGAB, M. G. et al. A novel one-dimensional cnn with exponential adaptive gradients for air pollution index prediction. *Sustainability*, v. 12, n. 23, 2020. ISSN 2071-1050. Disponível em: <<https://www.mdpi.com/2071-1050/12/23/10090>>. Citado 2 vezes nas páginas 32 e 33.
- SALINAS, D.; FLUNKERT, V.; GASTHAUS, J. *DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks*. 2019. Disponível em: <<https://arxiv.org/abs/1704.04110>>. Citado na página 66.
- SHAK, M. S. et al. Optimizing retail demand forecasting: A performance evaluation of machine learning models including lstm and gradient boosting. *The American Journal of Engineering and Technology*, v. 6, n. 09, p. 67–80, 2024. Citado na página 35.
- TAYLOR, S.; LETHAM, B. Previsão em escala. *PeerJ Preprints*, v. 5, n. e3190v2, 2017. Citado 2 vezes nas páginas 29 e 30.
- WANG WOON KIAN CHONG, J. L. J.; HEDENSTIERNA, C. P. T. Retail demand forecasting using spatial-temporal gradient boosting methods. *Journal of Computer Information Systems*, Taylor & Francis, v. 64, n. 5, p. 652–664, 2024. Disponível em: <<https://doi.org/10.1080/08874417.2023.2240753>>. Citado na página 36.

ZHANG, X. et al. Enhancing time series product demand forecasting with hybrid attention-based deep learning models. *IEEE Access*, v. 12, p. 190079–190091, 2024. Citado na página 35.

ZHAO, L.; ZHANG, Z. A improved pooling method for convolutional neural networks. *Scientific Reports*, v. 14, n. 1, p. 1589, 2024. ISSN 2045-2322. Disponível em: <<https://doi.org/10.1038/s41598-024-51258-6>>. Citado na página 32.