

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

MURILO EMANOEL SUDARIO RODRIGUES

**UTILIZAÇÃO DE APRENDIZADO DE MÁQUINA PARA PREVISÃO DE DEMANDA
USANDO ANÁLISE DE SENTIMENTO**

PONTA GROSSA

2025

MURILO EMANOEL SUDARIO RODRIGUES

**UTILIZAÇÃO DE APRENDIZADO DE MÁQUINA PARA PREVISÃO DE DEMANDA
USANDO ANÁLISE DE SENTIMENTO**

**Using machine learning for demand forecasting
using sentiment analysis**

Trabalho de conclusão de curso de Graduação
apresentado como requisito para obtenção do título
de Bacharel em Ciência da Computação da
Universidade Tecnológica Federal do Paraná
(UTFPR).

Orientadora: Simone de Almeida

PONTA GROSSA

2025



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

MURILO EMANOEL SUDARIO RODRIGUES

**UTILIZAÇÃO DE APRENDIZADO DE MÁQUINA PARA PREVISÃO DE DEMANDA
USANDO ANÁLISE DE SENTIMENTO**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título
de Bacharel em Ciência da Computação da
Universidade Tecnológica Federal do Paraná
(UTFPR).

Data de aprovação: 12/Novembro/2024

André Pinz Borges
Doutor
Universidade Tecnológica Federal do Paraná

Helyane Bronoski Borges
Doutora
Universidade Tecnológica Federal do Paraná

Simone de Almeida
Doutora
Universidade Tecnológica Federal do Paraná

**PONTA GROSSA
2025**

Dedico este trabalho à minha família, pelos
momentos de ausência.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço à minha orientadora Professora Simone de Almeida, pela sabedoria com que me guiou nesta trajetória.

Aos meus colegas de sala.

A Secretaria do Curso, pela cooperação.

Gostaria de deixar registrado também, o meu reconhecimento à minha família, pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa.

Eu denomino meu campo de Gestão do
Conhecimento, mas você não pode gerenciar
conhecimento. Ninguém pode. O que você pode
fazer, o que a empresa pode fazer é gerenciar o
ambiente que otimize o conhecimento.
(DAVENPORT; PRUSAK, 2012).

RESUMO

A ciência de dados combina disciplinas como computação, estatística e matemática para tratar, analisar e modelar dados com o objetivo de gerar previsões eficazes para decisões estratégicas. Este projeto foca na integração da análise de sentimentos, uma técnica de processamento de linguagem natural (NLP), com modelos de aprendizado de máquina para prever demandas em setores específicos. A análise de sentimentos identifica emoções em textos, mapeando informações subjetivas que podem aprimorar modelos preditivos, apesar dos desafios como ambiguidades linguísticas e dados ruidosos. O estudo propõe o uso de algoritmos como Naive Bayes, Regressão Logística, LSTM e Transformers para classificar sentimentos, integrando-os a modelos preditivos, como séries temporais e aprendizado profundo. A coleta de dados será feita em fontes confiáveis, como redes sociais e sites de avaliações, visando estruturar dados não estruturados. O projeto aplicará essas técnicas em um setor definido para validar seu impacto, contribuindo com novas abordagens na literatura científica para a previsão de demanda.

Palavras-chave: Análise de Sentimentos; Previsão de Demanda; Aprendizado de Máquina; Processamento de Linguagem Natural.

ABSTRACT

Data science combines disciplines such as computing, statistics, and mathematics to process, analyze, and model data with the goal of generating effective predictions for strategic decision-making. This project focuses on integrating sentiment analysis, a natural language processing (NLP) technique, with machine learning models to forecast demand in specific sectors. Sentiment analysis identifies emotions in text, mapping subjective information that enhances predictive models, despite challenges like linguistic ambiguities and noisy data. The study proposes using algorithms such as Naive Bayes, Logistic Regression, LSTM, and Transformers to classify sentiments, integrating them into predictive models like time series and deep learning. Data will be collected from reliable sources such as social media and review websites to structure unstructured information. The project will apply these techniques to a defined sector to validate its impact, contributing new approaches to the scientific literature on demand forecasting.

Keywords: Sentiment Analysis; Demand Forecasting; Machine Learning; Natural Language Processing.

SUMÁRIO

1	INTRODUÇÃO.....	9
1.1	Objetivos.....	11
1.1.1	Objetivo geral.....	11
1.1.2	Objetivos específicos.....	11
1.2	Justificativa.....	12
1.3	Metodologia.....	13
1.4	Estrutura do trabalho.....	14
2	ANÁLISE DE TRABALHOS SIMILARES.....	15
2.1	Previsão da volatilidade do Bitcoin.....	15
2.2	Comparação de modelos para previsões de preço.....	16
2.3	Previsão de vendas de elétricos.....	16
2.4	Análise de sentimento no mercado.....	17
2.5	Preferências hoteleiras via análise sentimental.....	18
2.6	Análise detalhada de preferências hoteleiras.....	19
2.7	Otimização de vendas no varejo.....	20
2.8	Previsão de sentimentos no turismo.....	21
2.9	Considerações gerais do capítulo.....	22
	REFERÊNCIAS.....	23

1 INTRODUÇÃO

A ciência de dados é um campo interdisciplinar que intersecta áreas como: ciência da computação, estatística e matemática aplicada. Tem como objetivo, a partir da coleta de dados: tratá-los, analisá-los, relacioná-los e, a partir de modelos de inteligência artificial, os quais variam de acordo com o problema proposto e os dados disponíveis, gerar previsões que sejam suficientes para tomadas de decisões nos mais diversos setores. Uma predição baseada em metodologia aplicada pode gerar como resultado uma economia de recursos e uma solução suficientemente eficaz e otimizada (QIANG *et al.* 2019).

Nesse âmbito, a análise de sentimentos (também chamada de mineração de opinião) é uma técnica de processamento de linguagem natural (NLP) que identifica e classifica emoções em textos. No contexto da previsão de demanda, ela possibilita mapear informações subjetivas do consumidor fazendo com que suas satisfações, frustrações e tendências de consumo sejam indicativos mais claros dentro do modelo usado. Atualmente, a análise de sentimentos enfrenta inúmeros desafios para a obtenção de uma precisão eficaz e uma determinação correta da polaridade presente nos sentimentos, tais como a detecção de ambiguidades linguísticas, sarcasmo, expressões culturais, sentimentos mistos, falta de contexto, dados ruidosos, desbalanceados e técnicos, dificultando a precisão e a generalização dos modelos (WANKHADE *et al.* 2022).

Contudo, é inegável que a combinação de processamento de linguagem natural para obtenção de mapeamento de sentimentos, junto a outros modelos de predição baseados em inteligência artificial, pode ajudar a prever as demandas dos consumidores, bem como suas ações (MIRZAALIAN e HALPENNY 2019).

O uso de aprendizado de máquina para prever demandas e o próprio comportamento de áreas da indústria é algo corriqueiro em diversos experimentos detalhados na literatura. Nesse sentido, Azamjon, Sattarov e Cho (2023) destacam que os algoritmos de aprendizado de máquina analisam grandes quantidades de dados históricos em tempo real, identificando padrões e relacionamentos que podem não ser perceptíveis pela análise humana. Além disso, afirmam que os modelos de aprendizado de máquina, como análise de regressão, redes neurais e árvores de

decisão, são hábeis em processar essa multiplicidade de pontos de dados, refinando suas previsões com cada nova informação. Algumas áreas comumente fazem uso de tais técnicas, tais como: Hotelaria (MIRZAALIAN e HALPENNY 2019), Indústria Automotiva (GUAN *et al.* 2022), Mercado de criptomoedas (AZAMJON *et al.* 2023), dentre tantas outras. Em todas, a predição da demanda de recursos é essencial para prever a volatilidade da área de estudo.

O uso de aprendizado de máquina e modelos baseados em inteligência artificial para processar linguagem natural e minerar textos está diretamente relacionado à análise de sentimento que o projeto pretende realizar. Segundo os autores Kejriwal, Garg e Sarin (2022), a análise de sentimentos envolve a utilização de técnicas de mineração de dados, aprendizado de máquina e inteligência artificial para determinar o sentimento de um texto, classificando-o como "positivo", "negativo" ou "neutro". Além disso, eles destacam que a análise de sentimentos é um processo poderoso para analisar dados não estruturados e é fundamental em diversas áreas, como gerenciamento de reputação de marca, análise de mercado, política, avaliação de campanhas de *marketing* e análise de sentimentos financeiros.

Sendo assim, explorar diferentes modelos de classificação de sentimento é de vital importância para transformar textos em dados estruturados, que são fundamentais para o bom funcionamento de modelos de predição. Técnicas como Naive Bayes, Regressão Logística e Redes Neurais (como LSTM e *Transformer*) são usadas para classificar sentimentos em positivo, negativo ou neutro e possibilitam a integração da análise de sentimento a modelos de previsão como séries temporais (ARIMA) ou modelos baseados em aprendizado profundo.

Outro ponto importante na construção de um bom modelo de predição baseado em análise de sentimento encontra-se na coleta de dados em fontes confiáveis. Redes sociais como Twitter, Facebook e Instagram oferecem grande volume de opiniões públicas e frequentemente atualizadas. Outras fontes utilizáveis são *sites* de avaliações. Plataformas como Yelp e Google Reviews contêm avaliações detalhadas dos consumidores. Comentários em *sites* de *e-commerce* também podem revelar padrões de satisfação e insatisfação que influenciam a demanda.

A intercalação da análise de sentimentos com modelos de predição de demanda não é tão expressiva quanto os métodos mais tradicionais que fazem uso exclusivamente de dados estruturados, muito por conta dos desafios acerca do

mapeamento dos sentimentos. Levando isso em consideração, este projeto de pesquisa tem como objetivo explorar e experimentar dentro dos tópicos abordados anteriormente, desenvolvendo uma aplicação que realize a previsão de demanda por meio de diferentes algoritmos e aplique métricas de avaliação para comparar a eficiência de cada um. A determinação de uma área específica como hotelaria, varejo, *e-commerce*, setor financeiro, entre outras possíveis, será feita para validar a aplicação desenvolvida e demonstrar seu potencial impacto.

1.1 Objetivos

A fim de orientar melhor o trabalho de pesquisa, foram definidos um objetivo geral e objetivos específicos, os quais são alcançados ao decorrer do trabalho.

1.1.1 Objetivo geral

O objetivo principal do trabalho é aplicar o aprendizado de máquina para realizar a previsão de demanda em um setor específico da indústria, utilizando uma base de dados que inclua, além de dados estruturados, textos provenientes de avaliações de clientes, os quais serão processados previamente à realização da predição de demanda.

1.1.2 Objetivos específicos

- Definir a base de dados para realização dos experimentos necessários para o desenvolvimento da pesquisa.
- Realizar a análise de sentimento nos textos provenientes dos clientes que estão contidos na base de dados por meio de processamento de linguagem natural.
- Obter a previsão de demanda aplicando de 3 a 4 diferentes algoritmos na realização dos experimentos.
- Fazer uma análise comparativa entre as métricas resultantes de cada algoritmo aplicado na base de dados utilizada nos experimentos, tais como Erro Médio Absoluto, Raiz do Erro Quadrático Médio e Coeficiente de Determinação.

1.2 Justificativa

O presente trabalho busca contribuir para o estudo das técnicas de previsão de demanda na indústria por meio da integração de análise de sentimentos a algoritmos de aprendizado de máquina. Enquanto a ciência de dados já oferece uma ampla gama de métodos para trabalhar com dados estruturados, há um espaço crescente para o desenvolvimento de abordagens que tratem dados qualitativos, especialmente no contexto de sentimentos e preferências de consumidores (COSTA *et al.* 2021). Com o aumento da importância das opiniões e *feedbacks* dos consumidores para a estratégia de negócios, a análise de sentimentos tem se mostrado um campo promissor, oferecendo *insights* valiosos que podem ser aplicados para melhor entender as flutuações da demanda e, com isso, otimizar a alocação de recursos (CARVALHO *et al.* 2023)

A combinação do processamento de linguagem natural e de técnicas de aprendizado de máquina para mapeamento de sentimentos representa uma inovação que pode ser aplicada em múltiplos setores da indústria. A partir dessa metodologia, espera-se alcançar um nível de precisão superior nas previsões de demanda, oferecendo uma solução que não apenas contribui para o aprimoramento das práticas de gestão de estoques e recursos, mas também auxilia empresas a responderem de forma mais eficiente às variações de mercado. Este projeto, portanto, visa preencher uma lacuna na literatura ao propor um modelo que busca aprimorar a capacidade de interpretação de dados qualitativos e não estruturados, proporcionando uma abordagem que transforma esses dados em informações práticas e estratégicas para o setor industrial.

Portanto, este trabalho justifica-se pela necessidade de explorar e aplicar técnicas modernas de IA, especialmente no campo da análise de sentimentos, como uma alternativa inovadora para prever demandas. A proposta busca não apenas contribuir academicamente, mas também oferecer uma solução prática e escalável para empresas interessadas em integrar tecnologias avançadas aos seus processos de planejamento e gestão.

1.3 Metodologia

O desenvolvimento deste trabalho está pautado nas etapas apresentadas a seguir.

Pesquisa documental:

A fundamentação teórica é dada por meio do produto do mapeamento sistemático de literatura realizada com base no Protocolo de Kitchenham e Charters (2007). O mapeamento de literatura possibilitou a seleção de oito trabalhos semelhantes e recentes para contribuição de pesquisas na área de análise de sentimento e previsão de demanda. Livros, teses, dissertações, revistas e sites indicados por docentes da área também haverão de ser consultados ao longo do desenvolvimento do projeto.

Recursos Técnicos:

A fim de implementar uma aplicação que seja capaz de realizar os objetivos propostos, alguns recursos devem ser utilizados. A Linguagem Python (versão 3.11.9) é usada para manipulação das bases de dados a fim de realizar análise, tratamento e abstração de um grande volume de dados. A versatilidade dessa linguagem e seu uso já tradicional no ramo da ciência de dados justificam a escolha (MCKINNEY 2022). São utilizadas bibliotecas diversas, como Pandas, Matplotlib, Numpy, Seaborn e Sklearn.

Bases de dados:

As bases de dados são fundamentais para todo o processo. A consulta de bases em *sites* como Kaggle¹, Google Dataset Search² e IBGE (Instituto Brasileiro de Geografia e Estatística)³ determinam potenciais bases que possam viabilizar o projeto de pesquisa. A base Brazilian E-Commerce⁴ será usada.

Modelos de Aprendizado de Máquina:

Diferentes modelos de aprendizado de máquina são testados ao longo da pesquisa. Os modelos apontados como eficientes na pesquisa documental são testados primeiro. São incluídos na pesquisa modelos que tenham uma precisão razoável em comparação com o que há de atual na literatura em pesquisas semelhantes, ou seja, que consigam realizar a previsão de demanda de modo

¹ KAGGLE. Disponível em: <https://www.kaggle.com>.

² GOOGLE DATASET SEARCH. Disponível em: <https://datasetsearch.research.google.com>.

³ INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Disponível em: <https://www.ibge.gov.br/>.

⁴ BRAZILIAN E-COMMERCE. Disponível em: <https://www.kaggle.com/code/annastasy/brazilian-e-commerce-eda-nlp-ml/notebook>

satisfatório e útil. Após a aplicação dos modelos será feita uma análise comparativa usando diferentes métricas, como Erro Médio Absoluto, Raiz do Erro Quadrático Médio e Coeficiente de Determinação.

1.4 Estrutura do trabalho

O trabalho está dividido em cinco capítulos. No primeiro, é introduzido o tema do trabalho, motivações e escopo geral. No segundo, parte da fundamentação teórica que foi utilizada para o desenvolvimento do trabalho é destrinchada e contextualizada à medida que influenciaram o produto final dessa pesquisa. No terceiro, a metodologia do trabalho é explicada, contendo as bases de dados, especificações dos métodos, tecnologias e ferramentas utilizadas. No quarto, é apresentado uma análise dos resultados, os quais contarão com análises comparativas das métricas obtidas nas execuções dos algoritmos, ao passo que no último capítulo há as considerações e conclusões finais do trabalho.

2 ANÁLISE DE TRABALHOS SIMILARES

Para melhor embasar o projeto de pesquisa, foi realizado um mapeamento sistemático de literatura com base no Protocolo de Kitchenham e Charters (2007). Por meio do mapeamento foi possível encontrar 8 trabalhos similares publicados nos últimos anos e que contemplam um bom estado da arte.

2.1 Previsão da volatilidade do Bitcoin

O artigo "*Forecasting Bitcoin Volatility Through On-Chain and Whale-Alert Tweet Analysis Using the Q-Learning Algorithm*" (AZAMJON *et al.* 2023), explora a volatilidade do Bitcoin com o intuito de prever tendências de mercado. A pesquisa combina dados on-chain, que detalham transações na *blockchain* do Bitcoin, com tweets da conta @whale_alert, que notifica sobre transações significativas feitas por grandes detentores de criptomoedas (crypto-whales). Utilizando o algoritmo Q-learning (uma abordagem de aprendizado por reforço), os autores analisam variáveis como volume de transações e atividade na rede para gerar previsões sobre tendências de preço. Os resultados indicam que a integração desses diferentes tipos de dados permite uma previsão mais precisa das flutuações de preço, oferecendo *insights* valiosos para investidores e gestores de risco no mercado de criptomoedas.

O estudo destaca a importância da análise de sentimento, especialmente a partir de plataformas como o Twitter, para prever a demanda e os preços de ativos voláteis como o Bitcoin. A inclusão dos tweets da @whale_alert oferece não apenas dados quantitativos sobre grandes transações, mas também *insights* sobre a percepção pública em torno delas. Dado o impacto das reações sociais no mercado de criptomoedas, essa dimensão torna-se crucial para entender a demanda. Com o uso de algoritmos de aprendizado de máquina, como Q-learning, a pesquisa demonstra que a combinação de dados quantitativos e qualitativos pode aumentar a precisão das previsões. A correlação entre transações de grandes investidores (crypto-whales) e a volatilidade sugere que o sentimento derivado das redes sociais afeta diretamente as expectativas de mercado, abrindo espaço para futuras pesquisas que integrem fontes diversas de dados e técnicas avançadas de análise.

2.2 Comparação de modelos para previsões de preço

O artigo intitulado "*Performance Comparison of LSTM and XGBOOST for Ether Price Prediction from Spam Filtered Tweets*" (SATHIYAPRIYA *et al.* 2023) compara dois modelos de previsão, o LSTM (*Long Short-Term Memory*) e o XGBoost (*Extreme Gradient Boosting*), com o objetivo de prever o preço da criptomoeda Ether a partir de dados de Twitter processados e filtrados para remoção de spam. Os autores destacam a relevância do Twitter como fonte de informações que influenciam a flutuação do preço do Ether.

O estudo propõe uma abordagem inovadora ao integrar aprendizado de máquina e análise de sentimentos para prever variações no preço de criptomoedas, destacando o impacto das opiniões expressas nas redes sociais na demanda e nos preços. A aplicação de um algoritmo robusto de filtragem de spam aprimora a qualidade dos dados analisados, garantindo que apenas informações relevantes sejam consideradas. Os resultados demonstram que a inclusão de dados sentimentais aumenta a precisão das previsões, sugerindo novas possibilidades para pesquisas futuras que considerem múltiplas plataformas sociais e a interconexão entre criptomoedas.

2.3 Previsão de vendas de elétricos

O artigo "*Sale Forecast and Analysis of Public's Attitude of EV Based on Combination of BP and LSTM Network and Decision Tree*" (GUAN *et al.* 2022) apresenta uma investigação abrangente sobre a previsão de vendas de veículos elétricos (EVs) e a análise da atitude do público em relação a essa tecnologia. Utilizando uma combinação de técnicas de aprendizado de máquina, em particular as redes neurais *backpropagation* (BP) e LSTM, juntamente com árvores de decisão, o estudo busca compreender como diferentes variáveis influenciam a aceitação e a compra de EVs.

A pesquisa destaca-se por fornecer *insights* sobre as preferências dos consumidores e as tendências de mercado, especialmente em um contexto de crescente relevância da mobilidade de veículos elétricos.

No que diz respeito à análise de sentimento, o artigo utiliza questionários e entrevistas para captar as emoções e opiniões dos consumidores em relação aos EVs. Por meio da aplicação de métodos de análise de dados, como árvores de

decisão, os autores examinam como aspectos como preço, infraestrutura de carregamento e consciência ambiental afetam a disposição do público em investir. Essa análise não só oferece uma percepção pública, mas também a vincula a padrões de compra subjacentes ao comportamento do consumidor.

A previsão de demanda para veículos elétricos é um dos pontos centrais do estudo, pois modelos preditivos ajudam a identificar tendências futuras e a formular estratégias de marketing eficazes. O artigo discute a eficácia dos modelos LSTM em comparação com abordagens tradicionais, evidenciando como as redes neurais podem captar padrões complexos em dados de vendas. Os resultados indicam que a integração de diferentes métodos pode aprimorar significativamente a precisão das previsões.

Ao incentivar decisões nas políticas de produção e *marketing* na indústria automotiva, esta pesquisa destaca a necessidade de um entendimento mais profundo dos sentimentos e comportamentos dos consumidores, especialmente no contexto de transição para veículos sustentáveis.

2.4 Análise de sentimento no mercado

O artigo "*Predict financial text sentiment: an empirical examination*" (KEJRIWAL *et al.* 2024) explora a importância da análise de sentimento no contexto dos mercados financeiros e a sua aplicação preditiva em relação ao comportamento das ações. Os autores explicam que, além das análises fundamentais e técnicas, a crescente disponibilidade de dados não estruturados, como tweets e notícias financeiras, torna essa abordagem um componente crucial para prever movimentos de preços. O estudo utiliza técnicas de mineração de texto e modelos de *machine learning*, destacando que o classificador *Kernel Support Vector Machine* apresentou a melhor precisão, com 68% na classificação de sentimentos como "positivo", "negativo" ou "neutro".

No âmbito da análise de sentimento, a pesquisa abrange diversas categorias, como a análise de sentimentos fina, a detecção de emoções e a análise baseada em intenções. Essas abordagens são essenciais para entender não apenas como os investidores reagem às informações, mas também para antecipar tendências de mercado. Além disso, essa técnica desempenha um papel fundamental no processo de tomada de decisão dos investidores, ao automatizar a

interpretação de grandes volumes de dados que, de outra forma, seriam difíceis de processar manualmente.

Os resultados do estudo indicam que a análise de sentimentos pode facilitar previsões de demanda, uma vez que o humor do mercado e a percepção pública exercem influência significativa sobre os preços das ações. Ao classificar rapidamente informações financeiras em categorias de sentimento, investidores podem tomar decisões mais informadas e oportunas, reduzindo riscos e aumentando lucros potenciais. O artigo não apenas contribui para o campo da análise financeira, mas também sugere que a integração dessa abordagem de análises de sentimentos com técnicas preditivas pode criar um modelo robusto para a previsão de tendências no mercado financeiro.

2.5 Preferências hoteleiras via análise sentimental

O artigo "*Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews*" (ZHANG *et al.* 2021) apresenta uma abordagem inovadora ao analisar preferências dos clientes em relação a hotéis, utilizando técnicas de análise de sentimento em um nível detalhado. Os autores, Jing Zhang, Xingchen Lu e Dian Liu, enfatizam a importância de extrair atributos de hotel a partir de comentários gerados por usuários, o que pode ajudar os gestores a compreender melhor as expectativas dos clientes. A pesquisa aborda o desafio de identificar termos de aspecto implícitos em análises de sentimento, propondo uma metodologia que combina a inserção de co-ocorrência e parsing de dependência, permitindo um entendimento mais profundo das sentenças e seus significados.

A interpretação detalhada dos sentimentos é um dos principais focos do estudo, pois permite que as empresas hoteleiras identifiquem não apenas as percepções gerais, mas também as nuances em relação a várias características do hotel, como localização, serviços e instalações. Os resultados indicam que diferentes tipos de clientes priorizam diferentes atributos, com uma ênfase específica em aspectos como higiene, preço e infraestrutura. Essa abordagem não só melhora a eficácia da extração de sentimentos, mas também oferece *insights* valiosos sobre as áreas que demandam melhorias, frequentemente associadas a feedbacks negativos dos hóspedes.

Além disso, a previsão de demanda pode ser aprimorada por meio dos insights derivados da análise de sentimento. Ao entender quais aspectos os clientes

valorizam mais, os hoteleiros podem ajustar suas ofertas e melhorar a satisfação do cliente, criando uma base mais sólida para a previsão de demanda. A implementação de tais análises permitirá que os hotéis não apenas atendam suas necessidades operacionais, mas também se antecipem a futuras demandas, garantindo uma experiência do cliente mais satisfatória e, consequentemente, aumentando a lealdade e a taxa de retorno dos hóspedes. Dessa forma, este estudo oferece uma contribuição relevante para a interseção entre análise de sentimentos e previsão de demanda, destacando seu papel estratégico na gestão hoteleira..

2.6 Análise detalhada de preferências hoteleiras

O artigo "*Customer preference identification from hotel online reviews: A neural network based fine-grained sentiment analysis*" (BIAN et al. 2022) aborda a importância da análise de sentimentos de avaliações *online* para entender as preferências dos clientes no setor hoteleiro. A pesquisa apresenta um modelo aprimorado de rede neural convolucional, que combina recursos não estruturados e estruturados para melhorar a identificação de pares de aspecto-opinião (AOPs). Esse tipo de análise permite que os hotéis capturem informações detalhadas sobre as opiniões dos clientes em relação a atributos específicos, como qualidade do serviço, preço e instalações, ao invés de se basear apenas em análises de sentimento de nível geral.

A análise precisa dos sentimentos, descrita no estudo, envolve múltiplas etapas, incluindo a extração de elementos sentimentais, identificação de AOPs e cálculo da intensidade do sentimento. Esses processos são fundamentais para classificar e quantificar as preferências do cliente em relação a diferentes atributos do hotel, influenciando diretamente as decisões de reserva e a percepção da qualidade do serviço.

A pesquisa conclui que 11 atributos hoteleiros principais foram destacados pelos clientes e que as preferências variam de acordo com o perfil do hóspede e o objetivo de viagem. Esses resultados ilustram a complexidade do comportamento do consumidor, oferecendo insights valiosos para a personalização de serviços e a formulação de estratégias mais eficazes na gestão hoteleira.

Um comentário sobre os achados do artigo é que eles demonstram como a aplicação de técnicas avançadas de análise de dados, como a análise de sentimentos, pode proporcionar *insights* valiosos para a previsão de demanda no

setor hoteleiro. Ao entender melhor as preferências dos clientes, os gerentes de hotéis podem ajustar suas estratégias de *marketing*, alocar recursos de forma mais eficaz e melhorar a satisfação do cliente. Isso não apenas aumenta a competitividade no setor, mas também pode levar a um aumento na taxa de fidelização e satisfação do cliente a longo prazo.

2.7 Otimização de vendas no varejo

O artigo "*Sales Optimization Solution for Fashion Retail*" (GANHEWA et al. 2021) aborda a implementação de técnicas de aprendizado de máquina (ML) para otimizar as vendas na indústria de varejo de moda, com foco especial em vestuário feminino. A pesquisa reconhece as complexidades do setor, como a volatilidade das tendências de moda e as flutuações na demanda do consumidor, que dificultam a previsão de vendas. Para superar esses desafios, o estudo propõe um sistema que integra três componentes principais: previsão de vendas, segmentação de clientes e análise de demanda. Essa abordagem permite que os varejistas possam gerenciar de forma mais eficiente o estoque e ajustem suas estratégias de marketing e atendam melhor às necessidades dos consumidores. O sistema proposto não só melhora a precisão das previsões, mas também contribui para a maximização das receitas e a redução de perdas associadas a estoques excedentes ou insuficientes.

Uma das abordagens mais interessantes discutidas no artigo é a análise de sentimento, que envolve examinar o *feedback* e as avaliações dos consumidores sobre produtos. Esta análise é fundamental para entender as preferências do cliente, ajudando os varejistas a identificar quais características dos produtos (como cor, estilo ou preço) se destacam nas avaliações positivas. Com isso, as empresas podem ajustar sua oferta de produtos conforme as expectativas dos consumidores, promovendo uma relação mais dinâmica e informativa entre o vendedor e o comprador. A categorização dos *feedbacks* em avaliações positivas e negativas também auxilia na identificação rápida de áreas que necessitam de melhorias.

Ademais, a previsão de demanda, uma parte integral do aplicativo proposto, utiliza algoritmos como o *Extra Trees* para estimar a quantidade de vendas futura com base em dados históricos. Esse processo é fundamental para a gestão eficiente do inventário, ajudando a evitar tanto o excesso quanto a falta de produtos. O uso de métodos de *machine learning* para prever essas demandas oferece uma abordagem

mais precisa de responder às flutuações do mercado, fornecendo aos varejistas ferramentas para maximizar as vendas e melhorar a satisfação do cliente.

Em suma, o artigo oferece uma abordagem inovadora para enfrentar os desafios do varejo de moda, destacando a capacidade de adaptação às necessidades do consumidor e o uso estratégico de dados para prever o comportamento do mercado, promovendo uma gestão mais eficiente e orientada por *insight*.

2.8 Previsão de sentimentos no turismo

O artigo "*Predicting sentiment and rating of tourist reviews using machine learning*" de (PUH E BAGIC 2023) explora o uso de modelos de aprendizado de máquina para analisar sentimentos e prever classificações a partir de avaliações de turistas. Utilizando um conjunto de dados do TripAdvisor, os pesquisadores aplicaram métodos como Naïve Bayes, máquinas de vetor de suporte (SVM), redes neurais convolucionais (CNN), redes de memória de longo e curto prazo (LSTM) e LSTM bidirecionais (BiLSTM). Os resultados demonstraram que os modelos de aprendizado profundo, especialmente o BiLSTM, superaram os algoritmos tradicionais em termos de precisão na classificação de sentimentos em avaliações como positivas, negativas ou neutras, além da previsão de classificações em uma escala de um a cinco estrelas.

A análise de sentimentos desempenha um papel crucial na indústria do turismo, em que as opiniões dos clientes influenciam significativamente as decisões de consumo. O artigo destaca como a capacidade de processar grandes volumes de dados de avaliações geradas por usuários pode fornecer informações valiosas às empresas do setor. Esses *insights* contribuem não apenas para a melhoria das experiências dos turistas, mas também para a adaptação de estratégias de *marketing* e no aprimoramento dos serviços oferecidos, criando uma base de *feedback* que pode orientar melhorias contínuas.

Além disso, o estudo aborda a previsão de demanda com base na interpretação de sentimentos. Ao identificar as preferências e insatisfações dos clientes, os modelos preditivos permitem antecipar tendências de fluxo turísticos e padrões de gastos em determinados destinos. Essa informação é valiosa para gestores e investidores na indústria do turismo, pois possibilita um planejamento

mais eficaz e a alocação adequada de recursos, promovendo a sustentabilidade e crescimento do setor.

Assim, o artigo não apenas contribui para o avanço acadêmico na aplicação de técnicas de aprendizado de máquina no turismo, mas também oferece diretrizes práticas para otimizar a *performance* de serviços em um mercado dinâmico em constante evolução.

2.9 Considerações gerais do capítulo

Em conclusão, os estudos analisados demonstram que a integração de fontes diversas de dados e a utilização de técnicas de aprendizado de máquina (especialmente quando combinadas com a análise de sentimentos) podem proporcionar previsões mais robustas e *insights* estratégicos valiosos. Essa abordagem multidisciplinar não só potencializa a precisão dos modelos preditivos, mas também contribui para a adaptação e inovação das empresas em resposta às rápidas mudanças nas preferências dos consumidores e nas condições de mercado.

3 DESENVOLVIMENTO

A fim de desenvolver o projeto de pesquisa proposto, foram usados materiais, ferramentas e métodos que possibilitaram que os objetivos, tanto os gerais quanto os específicos, fossem alcançados. As subseções seguintes têm como objetivo detalhar os materiais utilizados no projeto, bem como todos os métodos em ordem de aplicação.

3.1 Materiais e Ferramentas

Para a implementação do presente estudo, foi utilizado o ambiente de desenvolvimento JupyterLab, que é uma interface web interativa para notebooks, código e dados. O JupyterLab foi instalado e gerenciado por meio da distribuição Anaconda, amplamente utilizada em projetos de ciência de dados por já trazer uma série de bibliotecas e ambientes configurados. A versão utilizada do JupyterLab foi a mais recente disponível no momento da escrita deste trabalho (4.4.1).

A escolha pelo JupyterLab deve-se à sua integração com diversas bibliotecas Python, suporte a células interativas e visualização em tempo real, o que facilita tanto a exploração dos dados quanto a documentação do processo. Os experimentos e análises foram realizados em notebooks do tipo .ipynb, permitindo uma estrutura sequencial e clara para qualquer futura reprodução do estudo.

Para garantir o ambiente de execução idêntico ao utilizado neste trabalho, recomenda-se a instalação da distribuição Anaconda (adicionar nota de rodapé) e a criação de um novo ambiente virtual Python, versão 3.10.7.

3.1.1 Base de dados utilizada

A base de dados utilizada neste trabalho é o Brazilian E-Commerce Public Dataset by Olist, disponibilizada gratuitamente na plataforma Kaggle.

Esta base foi publicada pela própria empresa Olist, uma startup brasileira fundada em 2015 e sediada em Curitiba-PR, que atua como intermediadora entre pequenos lojistas e grandes marketplaces. A proposta da Olist é permitir que vendedores de diversos segmentos possam comercializar seus produtos em grandes plataformas de e-commerce, como Mercado Livre, Amazon, Magalu, entre outras, com maior visibilidade e alcance.

A base de dados foi tornada pública com o intuito de fomentar pesquisas acadêmicas e análises sobre o setor de comércio eletrônico no Brasil, sendo uma das poucas bases nacionais com informações reais e estruturadas sobre compras, pagamentos, logística, avaliações e comportamento do consumidor. O dataset reflete transações ocorridas entre setembro de 2016 e outubro de 2018.

Os responsáveis pela disponibilização e curadoria da base foram os próprios membros da equipe de dados da Olist, que garantiram que todas as informações estivessem anonimizadas, assegurando a privacidade dos usuários e vendedores envolvidos. A publicação da base no Kaggle foi feita sob a conta "olistbr", com licença aberta para fins acadêmicos e não comerciais.

A base está estruturada em múltiplos arquivos .csv, que juntos formam a seguinte relação demonstrada na figura 1.

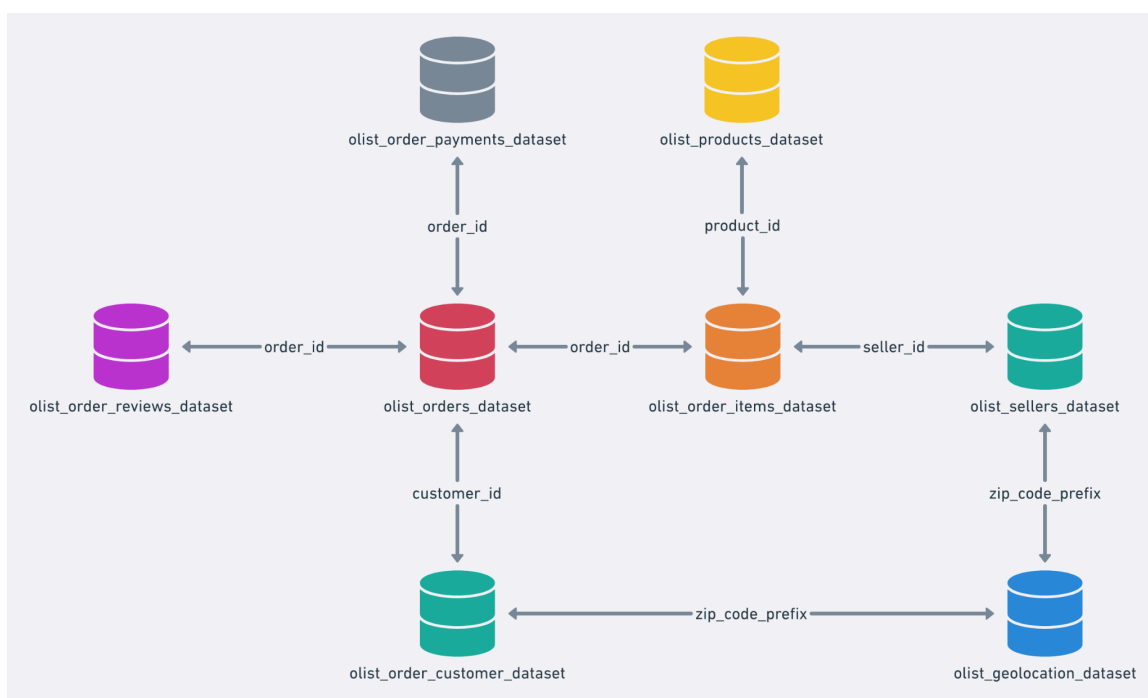


FIGURA 1

Cada arquivo representa uma dimensão específica do comércio eletrônico, como por exemplo:

- list_orders_dataset.csv: informações dos pedidos.

- olist_customers_dataset.csv: dados dos clientes.
- olist_order_items_dataset.csv: itens de cada pedido.
- olist_products_dataset.csv: dados dos produtos.
- olist_sellers_dataset.csv: informações dos vendedores.
- olist_order_reviews_dataset.csv: avaliações e comentários dos pedidos.
- olist_order_payments_dataset.csv: métodos e valores de pagamento.

Somente o olist_geolocation_dataset.csv não foi usado, pois a coluna zip_code_prefix estava ausente e por isso não foi possível concatená-la às outras bases de dados.

REFERÊNCIAS

- AZAMJON, M.; SATTAROV, O.; CHO, J. Forecasting Bitcoin Volatility Through On-Chain and Whale-Alert Tweet Analysis Using the Q-Learning Algorithm. **IEEE Access**, v. 11, p. 108092-108103, set. 2023.
- BIAN, YIWEN; YE, RONGSHENG; ZHANG, JING; YAN, XIN. Customer preference identification from hotel online reviews: A neural network based fine-grained sentiment analysis. **Computers & Industrial Engineering**, v. 172, p. 108648, set. 2022.
- CARVALHO, RENAN SILVA DE; TONIN, JOYCE MENEZES DA FONSECA; SANCHES, SIMONE LETICIA RAIMUNDINI. Previsibilidade do retorno das ações pela análise do sentimento textual: uma revisão. **Revista de Educação e Pesquisa em Contabilidade (REPeC)**, v. 17, n. 2, p. 140-160, jun. 2023.
- COSTA, JOÃO PAULO VIEIRA; BARROS, ROMULO BALDEZ DE; DANTAS, CAIO CÉSAR SILVA; SOUSA, RAQUEL CRISTINA DE; GOUVEIA, CRISTIANO GONÇALVES NASCIMENTO. Análise de Sentimento para Reviews Apresentados em Vídeos: Modelo de Redes Neurais Treinado em Base de Reviews Escritos. **Brazilian Journal of Technology**, v. 4, n. 1, p. 2-19, jan. 2021.
- GANHEWA, N. B.; ABEYRATNE, S. M. L. B.; CHATHURIKA, G. D. S.; DE SILVA, D.; LUNUGALAGE, D. Sales Optimization Solution for Fashion Retail. In: 3rd International Conference on Advancements in Computing (ICAC). 2021, Malabe. **Anais [...]** Malabe: Sri Lanka Institute of Information Technology, 2021. p. 443-448.
- GUAN ,Q.; MA ,Y.; YANG ,S; Sale Forecast and Analysis of Public's Attitude of EV Base on Combination of BP and LSTM Network and Decision Tree. In: International Conference on Machine Learning and Intelligent Systems Engineering. 2022, Guangzhou. **Anais [...]** Guangzhou: IEEE, 2022. p. 46-51.
- KEJRIWAL, R.; GARG, M.; SARIN, G. Predict financial text sentiment:an empirical examination. **Journal of Management** , v. 21, n.1, p. 44-54, mar. 2022.
- KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. 2007. **Technical Report**, EBSE Technical Report EBSE-2007-01, Keele University and University of Durham.
- MCKINNEY, WES. **Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter**. 3. ed. Sebastopol: O'Reilly Media, 2022.
- MIRZAALIAN, F.; HALPENNY, E. Social media analytics in hospitality and tourism. **Journal of Hospitality and Tourism Technology**, v. 10, n. 4, p. 764-790, nov. 2019.
- SATHIYAPRIYA, K.; VANKADARA, Sravya; BABU, Kumaresh Suresh; MURALIDHARAN, Mridula. Performance Comparison of LSTM and XGBOOST for Ether Price Prediction from Spam Filtered Tweets. In: International Conference On Intelligent Systems For Communication, Iot And Security (Iciscois). 2023, Coimbatore. **Anais [...]** Coimbatore: IEEE, 2023.

PUH, KARLO; BAGIC, MARINA. Predicting sentiment and rating of tourist reviews using machine learning. **Journal of Hospitality and Tourism Insights**, v. 6, n. 3, p. 1188-1204, fev. 2023.

QIANG, Z.; DAI, F.; LIN, H.; DONG, Y. Research on the Course System of Data Science and Engineering Major. *In*: IEEE INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND EDUCATIONAL INFORMATIZATION. 2019, Kunming. **Anais [...]** Kunming: IEEE, 2019. p. 90-93.

WANKHADE, M.; RAO, A.; KULKARNI, C. A survey on sentiment analysis methods, applications, and challenges. **Artificial Intelligence Review**, v. 55, p. 5731-5780, fev. 2022.

ZHANG, JING; LU, XINGCHEN; LIU, DIAN. Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews. **Electronic Commerce Research and Applications**, v. 49, p. 101094, set. 2021.