

**Universidade Presbiteriana Mackenzie**  
**Faculdade de Computação e Informática - Ciência de Dados**

Projeto Aplicado III – Sistema de recomendação em site de compras

Membros do grupo: Carlos Oliveira, Felipe Ferraz, Erick Isidoro

São Paulo, 01 de Abril de 2025

## Resumo

## Sumário

Resumo.....	2
Introdução .....	4
Metodologia .....	7
Resultado.....	8
Conclusões e trabalhos Futuros .....	9
GitHub.....	10
Link para acesso ao projeto no GitHub e Youtube .....	10
Tabela de imagens.....	11
Bibliografia .....	12

## Introdução

Neste projeto, trabalhamos com sistemas de recomendação para sites de vendas, uma ferramenta que contribui significativamente para o aumento das vendas, podendo, em muitos casos, representar a sobrevivência da empresa. Esses sistemas evitam o acúmulo de produtos, otimizam a alocação de espaço e direcionam esforços para produtos que, de fato, trarão resultados relevantes.

O objetivo principal do sistema é compreender os hábitos de compra dos clientes em sites de comércio eletrônico e sugerir produtos frequentemente adquiridos em conjunto. Por exemplo, ao adicionar farinha ao carrinho, o sistema pode sugerir automaticamente ingredientes como açúcar e ovos, que são comumente comprados juntos para o preparo de uma receita. Além disso, o sistema também pode sugerir receitas completas com base nos ingredientes já selecionados ou recomendar ingredientes adicionais de acordo com a preferência culinária do usuário, como pratos típicos italianos, japoneses ou brasileiros.

Este projeto aplicado considera um objetivo extensionista, atendendo às necessidades dos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, especialmente no que diz respeito ao consumo e produção responsáveis (ODS 12). Através da recomendação de produtos que otimizam o uso de recursos e reduzem o desperdício, o sistema contribui para práticas comerciais mais sustentáveis. Além disso, ao priorizar produtos de pequenos comerciantes e negócios locais, o projeto também apoia o crescimento econômico inclusivo (ODS 8) e promove a inovação na gestão de vendas e estoques (ODS 9).

A integração desses objetivos no sistema de recomendação possibilita não apenas aumentar as vendas e a satisfação do cliente, mas também impulsionar práticas comerciais responsáveis e socialmente conscientes, fortalecendo o impacto positivo do projeto tanto no âmbito econômico quanto social.

Os dados utilizados foram obtidos na plataforma Kaggle, [Análise do e-Commerce no Brasil - Olist Dataset](#), abrangendo o período de 2016 a 2018. Os dados foram disponibilizados pela Olist, uma loja de

departamentos com mais de 100 mil pedidos registrados. Por se tratar de dados reais, foram preservados o anonimato dos indivíduos, parceiros e empresas envolvidas.

Os dados foram organizados em oito conjuntos distintos: Consumidores, Vendedores, Produtos, Pedidos, Artigos dos Pedidos, Pagamento e Geolocalização.

## Referencial Teórico

O presente estudo apresenta a aplicação prática do algoritmo de K-Vizinhos Mais Próximos (KNN) em um sistema de recomendação de produtos, utilizando a linguagem Python e bibliotecas como **pandas** e **scikit-learn**. Inicialmente, realiza-se a importação das bibliotecas necessárias: **pandas** para manipulação de dados estruturados (MCKINNEY, 2010), e **scikit-learn** para divisão dos dados em conjuntos de treino e teste, além da implementação do KNN (GÉRON, 2019). A base de dados, em formato CSV, é carregada em um DataFrame, permitindo a manipulação tabular dos atributos dos produtos e das vendas realizadas.

Posteriormente, os dados são organizados, separando-se os atributos preditores do alvo (produto). A divisão dos dados entre treino e teste é feita para garantir a avaliação do modelo e evitar o sobreajuste. O modelo KNN é treinado com 10 vizinhos, sendo considerado um algoritmo não-paramétrico e baseado na distância entre pontos no espaço das características, comumente utilizando as métricas Euclidiana, Manhattan ou Minkowski (COVER; HART, 1967).

Na fase de recomendação, o sistema identifica os produtos mais semelhantes ao item escolhido, iterando sobre os vizinhos mais próximos e retornando sugestões de produtos. Por fim, é realizada uma análise dos itens mais vendidos, agrupando e ordenando os produtos de acordo com o volume de vendas.

O KNN se destaca pela simplicidade e eficiência em conjuntos de dados menores, sendo amplamente utilizado em sistemas de recomendação pela capacidade de avaliar a similaridade entre itens com base em atributos comuns (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; RICCI; ROKACH; SHAPIRA, 2015).

## Metodologia

## Resultado



## **Conclusões e trabalhos Futuros**

## GitHub

### Link para acesso ao projeto no GitHub e Youtube

Abaixo tem-se o link do GitHub onde estão compartilhados, dataset, cronograma, script outros documentos atualizados do projeto:

Demonstração da imagem do cronograma em tabelas de imagem ( Imagem 1 – Cronograma de Atividade), arquivo disponibilizado em link Github.

[https://github.com/Ferraz0Felipe/Projeto\\_Aplicado\\_III.git](https://github.com/Ferraz0Felipe/Projeto_Aplicado_III.git)

## Tabela de imagens

Imagem 1 – Cronograma de Atividade (disponível na integra link Github)

									Concluído	Programado	Atrasado															
									Fevereiro																	
Etapa	Título	Atividades	Responsável	Início	Término	Duração (dias)	Milestones	Status	16	17	18	19	20	21	22	23	24	25	26	27	28	29	01	02	03	04
1	Concepção do Produto	Organizar o grupo	Carlos	2-fev-2025	4-mar-2025	15	6-mar-2024	Concluído																		
		Escolher Tema	Erick	2-fev-2025	4-mar-2025	3	6-mar-2024	Concluído																		
		Escolher a base de dados	Felipe	2-fev-2025	4-mar-2025	5	6-mar-2024	Concluído																		
		definir o cronograma do projeto	Erick	2-fev-2025	4-mar-2025	4	6-mar-2024	Concluído																		
		Elabora o documento inicial	Carlos	2-fev-2025	4-mar-2025	4	6-mar-2024	Concluído																		
2	Definição do produto	Analisar a base de dados	Erick	11-mar-2025	10-abr-2025	24	3-abr-2024	Concluído																		
		limpar e preparar a base de dados	Felipe	11-mar-2025	10-abr-2025	16	3-abr-2024	Concluído																		
		Escolher a técnica para treinamento do modelo	Carlos	11-mar-2025	10-abr-2025	4	3-abr-2024	Concluído																		
		construir uma prova de conceito	Carlos	11-mar-2025	10-abr-2025	10	28-abr-2024	Concluído																		
		Definir métricas de avaliação de desempenho	Erick	11-mar-2025	10-abr-2025	12	28-abr-2024	Concluído																		
3	Metodologia	Implementar a técnica proposta	Felipe	3-abr-2025	3-mai-2025	8	28-abr-2024	Concluído																		
		Ajustar o pipeline de dados	Carlos	3-abr-2025	3-mai-2025	5	28-abr-2024	Concluído																		
		Documentar os passos implementados	Erick	3-abr-2025	3-mai-2025																					
4	Resultados e Conclusão	Organizar e Documentar os resultados	Carlos	2-mai-2025	1-jun-2025	11	17-mai-2024	Concluído																		
		Finalizar Documento do projeto	Erick	2-mai-2025	1-jun-2025	10	17-mai-2024	Concluído																		
		Entregar Materiais produzidos e apresentação	Felipe	2-mai-2025	1-jun-2025	8	17-mai-2024	Concluído																		

## Bibliografia

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21-27, 1967.

GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2. ed. Sebastopol: O'Reilly Media, 2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009.

MCKINNEY, W. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference*. 2010.

RICCI, F.; ROKACH, L.; SHAPIRA, B. *Recommender Systems Handbook*. 2. ed. Springer, 2015.