

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Desenvolvimento de um sistema para coleta,
armazenamento e análise nos dados de estoque,
produção e venda de salgados**

Michelly Fontoura da Silva

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Michelly Fontoura da Silva

**Desenvolvimento de um sistema para coleta,
armazenamento e análise nos dados de estoque, produção
e venda de salgados**

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientadora: Profa. Dra. Mirela Teixeira Cazzolato

Versão original

São Carlos

2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

F 677 Fontoura da Silva , Michelly
d Desenvolvimento de um sistema para coleta,
armazenamento e análise nos dados de estoque,
produção e venda de salgados. / Michelly Fontoura
da Silva ; orientadora Mirela Teixeira Cazzolato. -
- São Carlos, 2024.
72 p.

Trabalho de conclusão de curso (MBA em
Inteligência Artificial e Big Data) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2024.

1. Mineração de dados. 2. Inteligência
artificial. 3. Gestão de estoque. 4. Visualização de
dados. 5. Análise de vendas. I. Teixeira Cazzolato,
Mirela , orient. II. Título.

Michelly Fontoura da Silva

**Development of a system for collecting, storing, and
analyzing data on inventory, production, and sales of
snacks**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Profa. Dra. Mirela Teixeira Cazzolato

Original version

São Carlos

2024

RESUMO

SILVA, M. F. **Desenvolvimento de um sistema para coleta, armazenamento e análise nos dados de estoque, produção e venda de salgados.** 2024. 72 p.

Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Este trabalho de conclusão de curso explora a implementação de técnicas analíticas e a gestão eficiente de dados para otimizar operações em empresas do setor alimentício, com foco na loja "Salgados do Quadrado", que se especializou em salgadinhos para festas. O uso crescente de tecnologias modernas, como mineração de dados e inteligência artificial, tornou-se crucial para a tomada de decisões estratégicas. No entanto, a complexidade dessas ferramentas muitas vezes representa um desafio para os proprietários de empresas, especialmente no setor alimentício, onde a aplicação prática e a compreensão de algoritmos podem ser difíceis. A motivação para este estudo surge da experiência do autor como proprietário e gerente da "Salgados do Quadrado", uma empresa que está em operação há quase três anos. A empresa enfrenta desafios na coleta e análise de dados de pedidos, particularmente devido a métodos manuais e à falta de integração entre os pedidos feitos na loja e aqueles realizados por meio de plataformas digitais como o iFood. Essa desconexão dificulta uma visão unificada do desempenho da empresa, complicando a identificação de padrões de vendas e o desenvolvimento de estratégias baseadas em dados. Para abordar essas questões, o projeto propõe a criação de um sistema de gestão de vendas e estoque para digitalizar a coleta e o armazenamento dos dados de pedidos e controle de estoque da empresa. Esse sistema permitirá que a "Salgados do Quadrado" melhore a logística interna, desenvolva estratégias de vendas informadas e aprimore o atendimento ao cliente. A implementação dessa tecnologia visa aumentar a eficiência, reduzir os custos operacionais, minimizar o desperdício e elevar a experiência do cliente. Além disso, o projeto inclui o desenvolvimento de uma plataforma analítica que integra os dados internos com aqueles provenientes do iFood. Essa plataforma otimizará processos empresariais como gestão de estoque, fluxo financeiro e atendimento personalizado ao cliente, ao mesmo tempo, em que possibilitará campanhas de marketing mais eficazes e práticas sustentáveis. Ao incorporar técnicas de visualização de dados, a plataforma busca simplificar a análise e a tomada de decisões, capacitando a "Salgados do Quadrado" a atender às demandas do mercado moderno e alcançar sucesso a longo prazo.

Palavras-chave: Mineração de dados. Ramo alimentício. Gerenciamento de dados. Visualização.

ABSTRACT

SILVA, M. F. **Development of a system for collecting, storing, and analyzing data on inventory, production, and sales of snacks.** 2024. 72 p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

This final paper explores the implementation of analytical techniques and efficient data management to optimize operations in companies within the food industry, with a focus on the “Salgados do Quadrado” shop, which specializes in party snacks. The growing use of modern technologies such as data mining and artificial intelligence has become crucial for strategic decision-making. However, the complexity of these tools often poses challenges for business owners, particularly in the food sector, where practical application and understanding of algorithms can be difficult. The motivation for this study comes from the author’s experience as the owner and manager of “Salgados do Quadrado,” a business that has been in operation for nearly three years. The company faces challenges in collecting and analyzing order data, particularly due to manual methods and a lack of integration between in-store orders and those made through digital platforms like iFood. This disconnect hinders a unified view of the company’s performance, complicating the identification of sales patterns and the development of data-driven strategies. To address these issues, the project proposes creating a sales and inventory management system to digitize the collection and storage of order and inventory data. This system will enable “Salgados do Quadrado” to enhance internal logistics, develop informed sales strategies, and improve customer service. The implementation of this technology aims to increase efficiency, reduce operational costs, minimize waste, and elevate the customer experience. Additionally, the project includes developing an analytical platform that integrates internal data with that from iFood. This platform will optimize business processes such as inventory management, financial flow, and personalized customer service while enabling more effective marketing campaigns and sustainable practices. By incorporating data visualization techniques, the platform aims to simplify analysis and decision-making, empowering “Salgados do Quadrado” to meet modern market demands and achieve long-term success.

Keywords: Data mining. Food business. Data management. Visualization.

LISTA DE FIGURAS

Figura 1 – Fluxo das fases do KDD englobando a etapa de Data Mining	22
Figura 2 – Regra de associação obtida através do programa Orange Data Mining.	24
Figura 3 – Exemplos de modelos de classificação obtidos através do programa knime	26
Figura 4 – Exemplo de Agrupamento (<i>Clustering</i>)	27
Figura 5 – Visualização de dados obtida por meio do programa Power BI	28
Figura 6 – Modelo Entidade Relacional - MER	41
Figura 7 – Arquitetura do sistema proposto.	44
Figura 8 – Componentes do sistema proposto.	44
Figura 9 – Workflow construído para gerar regras de associação a partir do arquivo contendo dados binários de produtos.	53
Figura 10 – Resultado obtido através da aplicação do modelo Apriori utilizado na mineração de dados para busca identificar regras de associação em grandes conjuntos de dados.	54
Figura 11 – Workflow construído cruzando arquivos contendo palavras positivas e negativa e concatenadas ao documento de feedback para gerar um dicionário de palavras de sentimento.	56
Figura 12 – Resultado de saída da análise das palavras positivas e negativas identi- ficadas pelo dicionário.	56
Figura 13 – Workflow TF-DF construído para ajuda a ranquear os documentos mais relevantes para a consulta, considerando a frequência das palavras-chave nos documentos.	57
Figura 14 – Nuvem de palavras gerada como resultado do node TF-IDF, onde a palavra goiabada aparece num tamanho maior comparado a outras palavras, indicando uma maior citação da mesma dentre os textos analisados.	57
Figura 15 – Dashboard de dados dos salgados	58
Figura 16 – Workflow de Análise de Tendências: Usando o nó Time Series Generator para identificar padrões ao longo do tempo, como sazonalidade ou tendências nos pedidos.	58
Figura 17 – Coleta de dados por gênero	59
Figura 18 – TOP 20 dados analisados pelos node “Statistics”	60
Figura 19 – Menu principal do site que serve como porta de entrada para a navegação do usuário escolher a tarefa que deseja executar.	62
Figura 20 – Dashboard contendo gráficos dos dados registrados pela aplicação. . . .	63
Figura 21 – Página para preenchimento e salvamento dos dados de cadastro de cliente	64
Figura 22 – Página para preenchimento e salvamento dos dados de cadastro de produto	64

Figura 23 – Página para preenchimento e salvamento dos dados de cadastro de insumo	64
Figura 24 – Página para preenchimento e salvamento dos dados de Pedidos realizados.	65
Figura 25 – Página de exibição dos Relatórios de Clientes, Produtos Insumos e Pedidos cadastrados no sistema.	65
Figura 26 – Cardápio contendo as opções de menu da loja.	66
Figura 27 – Redes sociais da loja Salgados do Quadrado	66
Figura 28 – MER	67

LISTA DE TABELAS

Tabela 1 – Comparação das ferramentas levantadas considerando suas funcionalidades (primeira parte).	34
Tabela 2 – Comparação das ferramentas levantadas considerando suas funcionalidades (continuação).	35
Tabela 3 – Resumo dos resultados obtidos.	50

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Contextualização	17
1.2	Motivação	17
1.3	Lacunas	17
1.4	Cenário de Contribuição	18
1.5	Objetivos	18
1.6	Estrutura do trabalho	18
2	CONCEITOS FUNDAMENTAIS	21
2.1	Regras de Associação	21
2.2	Classificação	24
2.3	Agrupamento	26
2.4	Considerações Finais	29
3	TRABALHOS RELACIONADOS	31
3.1	Ferramentas Analíticas	31
3.2	Trabalhos de Análise de Padrões de Compras	36
3.3	Discussão Sobre os Trabalhos Relacionados	37
4	METODOLOGIA	39
4.1	Levantamento dos requisitos do projeto	39
4.2	Estudo de ferramentas para análise e visualização	39
4.3	Estudo de algoritmos de mineração de dados	39
4.4	Modelagem e implementação do banco de dados do problema	40
4.5	Estudo de tecnologias para implementação do protótipo do sistema	40
4.6	Validação com base nos dados já coletados (históricos) do negócio	42
5	PROPOSTA E DESENVOLVIMENTO	43
5.1	Vertente de Análise de Dados	43
5.2	Vertente do Sistema Proposto	43
5.2.1	Usuário	43
5.2.2	Streamlit (Interface)	45
5.2.3	MySQL (Dados)	45
5.2.4	Python (Backend)	45
5.2.5	Matplotlib/Plotly (Visualização)	46
5.2.6	Scikit-learn (Análise)	46
5.2.7	Testes	47

5.3	Fluxo de Interação Entre Componentes	47
6	RESULTADOS	49
6.1	Análise de Cenário e Organização dos Dados	49
6.2	Base de Dados do SquareSys	49
6.3	Cenários de Análise	51
6.4	Análise de Dados	53
6.4.1	Criação de um dashboard	60
6.5	Implementação do SquareSys	61
6.6	Considerações Finais	63
7	CONCLUSÕES	69
7.1	Contribuições	69
7.2	Benefícios Práticos para o Negócio Analisado	69
7.3	Trabalhos Futuros	69
7.4	Considerações Finais	70
	REFERÊNCIAS	71

1 INTRODUÇÃO

1.1 Contextualização

A utilização de ferramentas analíticas e o eficiente gerenciamento de dados coletados são fundamentais no funcionamento e otimização de diversos setores relacionados a negócios, incluindo empresas do ramo alimentício.

Existe uma plethora de algoritmos de Mineração de Dados e Inteligência Artificial, mas por vezes seu uso é de difícil aplicação e entendimento para usuários não familiarizados. Além disso, com a disseminação de aplicativos de venda do meio alimentício, muitas vezes os dados de empresas são coletados e gerenciados de diversas formas ao mesmo tempo, o que dificulta a análise do negócio de maneira completa.

1.2 Motivação

A aluna é proprietária e administradora da empresa “Salgados do Quadrado”, uma empresa que atua no setor de salgados para festas. A empresa existe há 2 anos e 9 meses, atuando desde a produção dos salgados até a venda final dos produtos, além de cuidar do marketing e do atendimento aos clientes. Atualmente, o iFood é utilizado como plataforma complementar da loja, para receber pedidos e realizar as entregas. O iFood é uma plataforma digital que usa a tecnologia para conectar consumidores, entregadores, restaurantes e varejistas a levar refeições e compras à casa dos clientes.

Os pedidos feitos diretamente pelos clientes na loja são anotados em blocos de notas. A forma manual de anotação dos pedidos é prejudicial ao negócio, pois há o risco de várias formas de danificação dos papéis onde as informações podem se perder, além de não ser possível buscar por *insights* de padrões de vendas e dados personalizados de clientes, pois não há armazenamento de dados históricos ou integração das vendas feitas de forma direta e pelo iFood.

1.3 Lacunas

Plataformas como iFood fornecem um conjunto de funcionalidades para auxiliar os proprietários na tomada de decisão. No entanto, muitas das informações gerenciadas (por exemplo, avaliação de usuários, motivos de cancelamentos) não são transparentes para os negócios. Além disso, pedidos feitos fora da plataforma não podem ser analisados com os dados que passam por ela, fazendo com que a análise fornecida pelo aplicativo seja, de certa forma, enviesada e incompleta.

A literatura fornece diversas técnicas de análise de dados. No entanto, muitas vezes a

aplicação dessas técnicas em empresas que atuam em nichos personalizados pode ser pouco trivial para proprietários de negócios. Dessa forma, há a necessidade do desenvolvimento de técnicas analíticas intuitivas, que permitam aos donos de negócios entender como os dados se comportam, e tomar decisão a partir deles.

1.4 Cenário de Contribuição

Com a implementação de um sistema de gerenciamento de vendas e estoque, será possível não apenas aprimorar a logística da loja e traçar estratégias de venda e atendimento ao cliente com base em dados, mas também melhorar a eficiência na preparação e nas entregas dos salgados. Isso resultará em uma redução de tempo, custos e desperdício de produtos.

1.5 Objetivos

O objetivo deste projeto é implementar técnicas analíticas de regras de negócio em cima de dados coletados por meio de encomendas, uso de estoque e entregas. Combinado a isso, objetiva-se utilizar os dados disponibilizados por pedidos diretos e pela plataforma iFood, fazendo com que a análise englobe todos os pedidos e gerenciamentos de dados da empresa.

Dessa forma, os objetivos do presente projeto estão alinhados em duas vertentes:

1. Um sistema de gerenciamento de dados internos, para informatização da coleta e armazenamento dos dados de pedidos e estoque da empresa.
2. Uma plataforma analítica, que visa otimizar os processos relacionados ao negócio, que pode incluir: a otimização de entregas; melhora no controle e gerenciamento de estoque; campanhas e atendimentos personalizados para os clientes; fluxo financeiro; melhora no fluxo de venda de produtos; e medidas sustentáveis que possam ser incorporadas ao negócio. Essa plataforma trabalhará com os dados internos (coletados pelo item 1) e os dados coletados do iFood.

A solução desenvolvida terá como base trabalhos que extraem padrões relevantes de textos, séries temporais e dados transacionais. Além disso, a plataforma analítica será apoiada por técnicas de visualização de dados, visando reduzir a complexidade de entendimento dos padrões e apoiar usuários na tomada de decisões por meio de ferramentas intuitivas.

1.6 Estrutura do trabalho

Os demais capítulos desta monografia estão organizados da seguinte forma. O Capítulo 2 apresenta a fundamentação teórica. O Capítulo 3 descreve os trabalhos relacionados.

O Capítulo 3 apresenta e detalha a proposta deste projeto. O Capítulo 6 descreve a análise experimental e validação da proposta. Finalmente, o Capítulo 7 apresenta as conclusões deste trabalho.

2 CONCEITOS FUNDAMENTAIS

Neste capítulo são apresentados os principais conceitos relacionados a este trabalho de conclusão de curso.

Os conceitos relacionados a este projeto incluem tarefas analíticas de dados transacionais, como regras de associação entre itens comprados juntos e outras coocorrências. Além disso, considera-se o contexto de algoritmos de reconhecimento de padrões e tendências em conjuntos de dados, que incluem as tarefas de classificação e agrupamento. Finalmente, o objetivo do trabalho engloba também o emprego de ferramentas de visualização para suporte ao entendimento dos padrões e a tomada de decisão. Tais ferramentas têm o papel de dar suporte ao entendimento dos padrões descobertos, bem como fornecer meios intuitivos de entendimento dos dados coletados e armazenados pela aplicação que está ligada ao domínio de vendas.

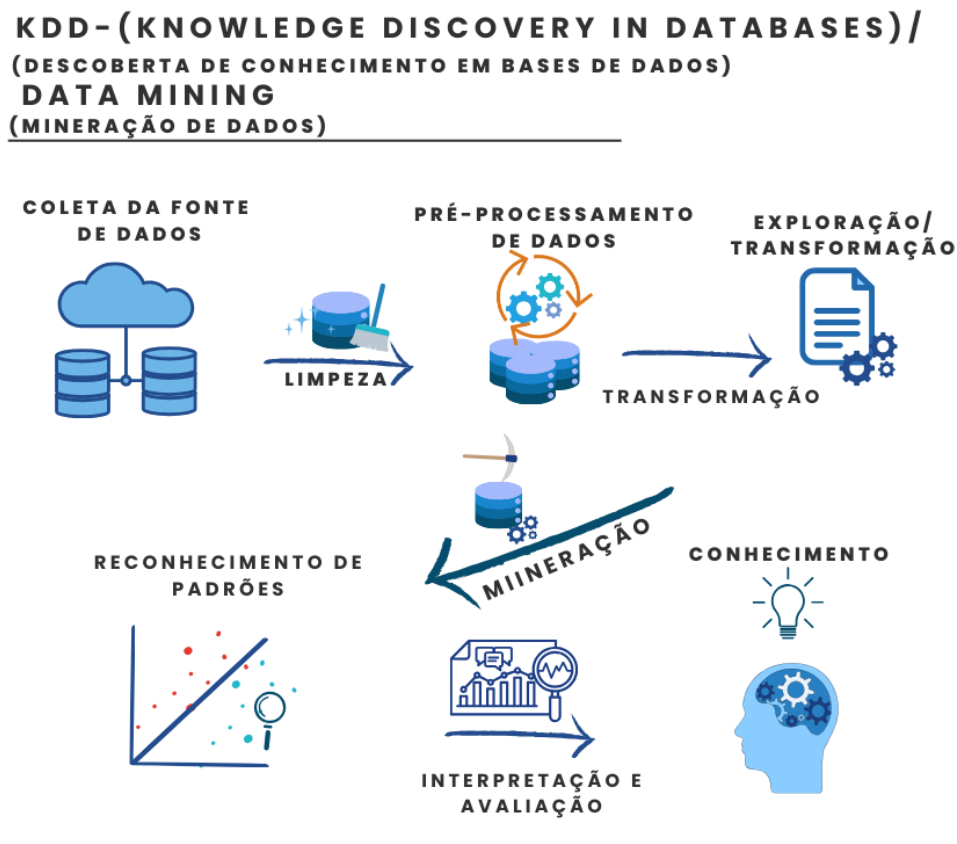
2.1 Regras de Associação

Todos os dias, milhares de dados estruturados e não estruturados são gerados e armazenados em bancos de dados. No começo da era digital, os dados eram armazenados em fitas magnéticas, disquetes, CD, DVDs, Pendrives, máquinas locais, dentre outros. Mas, com a evolução da internet e o crescimento do volume gerado diariamente por meios transacionais, mídias sociais e IOT (do inglês, *Internet of Things* – Internet das coisas), os dados começaram a ser armazenados em bancos de dados escalonáveis e em nuvem (Servidores Remotos) para poder não só armazenar uma quantidade maior, mas também acessá-los de qualquer lugar e a qualquer momento (Han; Kamber; Pei, 2011).

Se antigamente o objetivo de guardar dados era somente como forma de armazenagem de controle de registro, hoje em dia os dados passaram a ser informações valiosas dentro das organizações, pois através deles podemos tirar *insights* de negócios importantes que podem agregar valores estratégicos para as organizações. Uma forma de utilizar os dados de maneira estratégica é através da Mineração de Dados (do inglês, *Data Mining*), que faz uso de vários algoritmos e técnicas de extração de conjuntos de dados, onde é possível identificar padrões, detectar anomalias e otimizar o suporte a decisões (Mitchell, 1997; Han; Kamber; Pei, 2011).

A mineração de dados é parte do processo amplamente conhecido como KDD (do inglês, *Knowledge Discovery in Databases* – Descoberta de Conhecimento em Bancos de Dados), na qual a Mineração de Dados está inserida e faz parte de uma etapa do processo (Han; Kamber; Pei, 2011). A Figura 1 mostra os passos desse processo: Coleta da fonte de dados; Limpeza; Pré-Processamento; Transformação; Exploração e Transformação dos

Figura 1 – Fluxo das fases do KDD englobando a etapa de Data Mining



Fonte: Adaptado de (Han; Kamber; Pei, 2011).

dados; Mineração de dados para Reconhecimento de Padrões, Interpretação e Avaliação dos padrões obtidos, buscando o conhecimento.

O KDD extrai conhecimentos úteis ao analisar padrões e modelos de grande volumes de dados. O processo aplica os conhecimentos descobertos em *insights* de tomadas de decisão. Durante a etapa de Mineração de Dados, é possível aplicar as regras de associação (do inglês, *Association Rules Mining* – ARM). Regras de associação servem para identificar associações entre diferentes variáveis de um conjunto de dados (Han; Kamber; Pei, 2011).

Para descobrir as regras de associação de um conjunto de dados transacionais, um dos algoritmos que podem ser utilizados é o Apriori, amplamente conhecido na literatura e utilizado em aplicações práticas (Han; Kamber; Pei, 2011; Agrawal; Imielinski; Swami, 1993). O Apriori analisa um conjunto de transacionais e conta quantas ocorrências cada item da lista teve. Analisando a frequência dos itens, é possível calcular duas medidas de confiança elementares para o Apriori: o Suporte e a Confiança (Santos; Nascimento, 2023).

A ARM consiste em identificar regras de associação frequentes e válidas em um conjunto de dados (Han; Kamber; Pei, 2011). Uma regra de associação R é uma implicação

da forma:

$$X \rightarrow Y,$$

onde X e Y são conjuntos de itens tais que $X \cap Y = \emptyset$. Um item é uma condição que pode assumir valor verdadeiro ou falso em função do registro de dados selecionado, e essa representação é inerente aos conjuntos de dados transacionais. Por exemplo:

$$R1 : \text{Sexo} = M, \text{JogaFutebol} = S \rightarrow \text{Saude} = \text{Boa}$$

é uma regra de associação onde $\text{Sexo}=M$ (masculino), $\text{JogaFutebol}=S$ (sim) e $\text{Saude}=Boa$ são itens. Satisfazem à regra $R1$ todos os registros do conjunto de dados que satisfazem (ou seja, tornam verdadeiros) esses três itens ao mesmo tempo.

Uma regra de associação $R : X \rightarrow Y$ é dita frequente (ou válida) se, e somente se:

$$\text{Sup}(R) = \frac{|X \cup Y|}{|D|} \geq \text{MinSup}$$

e

$$\text{Conf}(R) = \frac{|X \cup Y|}{|X|} \geq \text{MinConf},$$

onde $X \cup Y$ é o conjunto de registros que satisfazem os itens em X e Y simultaneamente; $\text{Sup}(R)$ e $\text{Conf}(R)$ são, respectivamente, o suporte e a confiança de uma regra R ; $|D|$ representa a quantidade total de registros de dados disponíveis no conjunto de dados D , ou seja, seu tamanho; e MinSup e MinConf são o suporte mínimo e a confiança mínima, respectivamente, que consistem em parâmetros definidos pelo usuário.

Exemplo: Considerando a regra $R1$ exemplificada anteriormente, se o conjunto de dados tem 100 registros no total, dos quais 50 satisfazem simultaneamente aos itens $\text{Sexo}=M$ e $\text{JogaFutebol}=S$, mas apenas 45 desses satisfazem também ao item $\text{Saude}=Boa$, tem-se que $\text{Sup}(R1)=45\%$ e $\text{Conf}(R1)=90\%$. Assumindo, por exemplo, $\text{MinSup}=10\%$ e $\text{MinConf}=80\%$, $R1$ seria considerada uma regra frequente e válida.

Em geral, o processo de ARM ocorre em duas etapas (Agrawal; Imielinski; Swami, 1993). A primeira, de maior custo computacional, *busca por conjuntos de itens frequentes* que ocorrem simultaneamente no conjunto de dados. A segunda consiste em *identificar as regras válidas a partir de cada conjunto de itens considerado frequente*, resultantes da etapa anterior.

No contexto do presente projeto de pesquisa, a regra de associação é empregada na análise dos pedidos, contendo as informações dos salgados, como quantidade e sabores que mais foram vendidos nos últimos meses. Essa análise tem como objetivo aumentar o poder de personalização dos combos de salgados com os sabores de maior relevância para os clientes, baseado na frequência de vendas.

A Figura 2 mostra um exemplo de aplicação de ARM. No exemplo, o algoritmo Apriori foi aplicado sobre um conjunto de dados de vendas de salgados. É possível observar,

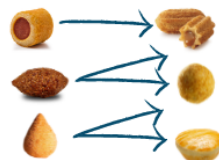
na parte esquerda da imagem, as diferentes métricas geradas pelas regras. Nas duas primeiras colunas estão presentes o Suporte (Supp) e Confiança (Conf), discutidos nesta seção. Outras métricas relacionadas a ARM são exibidas, como a cobertura, *lift*. Mais detalhes sobre elas podem ser encontrados em (Han; Kamber; Pei, 2011). Nas duas colunas da direita são exibidas as regras encontradas. Elas são divididas em Antecedente e Consequente, que dizem respeito às duas partes de uma regra gerada.

Figura 2 – Regra de associação obtida através do programa Orange Data Mining.

REGRA DE ASSOCIAÇÃO

DADOS OBTIDOS ATRAVÉS DO PROGRAMA ORANGE DATA MINING

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	CONSEQUENT
0.359	0.933	0.385	2.333	1.040	0.014	Bolinha de Queijo=1, Empada de Alho-Poró=0	→ Coxinha=1
0.410	1.000	0.410	2.188	1.114	0.042	Bolinha de Queijo=1, Empada de Frango=1	→ Coxinha=1
0.385	0.938	0.410	2.188	1.045	0.016	Bolinha de Queijo=1, Enroladinho de Salsicha=1	→ Coxinha=1
0.462	0.947	0.487	1.842	1.056	0.024	Bolinha de Queijo=1, Kibe=1	→ Coxinha=1
0.436	1.000	0.436	2.059	1.114	0.045	Empada de Frango=1, Kibe=1	→ Coxinha=1
0.462	0.900	0.513	1.350	1.300	0.107	Coxinha=1, Mini Churros=1	→ Bolinha de Queijo=1
0.462	0.947	0.487	1.842	1.056	0.024	Bolinha de Queijo=1, Mini Churros=1	→ Coxinha=1
0.359	0.933	0.385	1.800	1.348	0.093	Kibe=1, Mini Churros=1	→ Bolinha de Queijo=1
0.359	0.933	0.385	2.333	1.040	0.014	Kibe=1, Mini Churros=1	→ Coxinha=1
0.410	0.941	0.436	2.059	1.049	0.019	Empada de Alho-Poró=0, Risole de Milho=0	→ Coxinha=1
0.359	0.933	0.385	1.533	1.583	0.132	Coxinha=1, Mini Churros=0	→ Risole de Milho=0
0.436	0.944	0.462	1.944	1.052	0.022	Bolinha de Queijo=1, Quadrado de Calabresa com Queijo=1	→ Coxinha=1
0.410	1.000	0.410	2.188	1.114	0.042	Empada de Frango=1, Quadrado de Calabresa com Queijo=1	→ Coxinha=1
0.359	0.933	0.385	1.800	1.348	0.093	Kibe=1, Quadrado de Calabresa com Queijo=1	→ Bolinha de Queijo=1
0.359	0.933	0.385	2.333	1.040	0.014	Kibe=1, Quadrado de Calabresa com Queijo=1	→ Coxinha=1



Fonte: Elaborada pela autora.

2.2 Classificação

O Aprendizado de Máquina (AM, ou ML do inglês *Machine Learning*) pode ser definido como o campo de estudo se preocupa em como fornecer ao computador a habilidade de aprender sem ser explicitamente programado (Wiederhold; McCarthy, 1992). É o ramo da Inteligência Artificial que utiliza técnicas e algoritmos com o intuito de reconhecer padrões ou de melhorar seu desempenho por meio de sua experiência (Mitchell, 1997; Russell; Norvig, 2020). De forma geral, existem três formas de aquisição de conhecimento pelas técnicas de Aprendizado de Máquina, a saber: Aprendizado Supervisionado, Aprendizado

não Supervisionado e Aprendizado por Reforço (Russell; Norvig, 2020; Santos; Nascimento, 2023).

A classificação é uma parte importante na mineração de dados e na aprendizagem de máquinas, pois é a forma de encontrar um modelo para distinguir e descrever classe de dados ou conceitos através da atribuição rótulos com base em características específicas. (Han; Kamber; Pei, 2011)

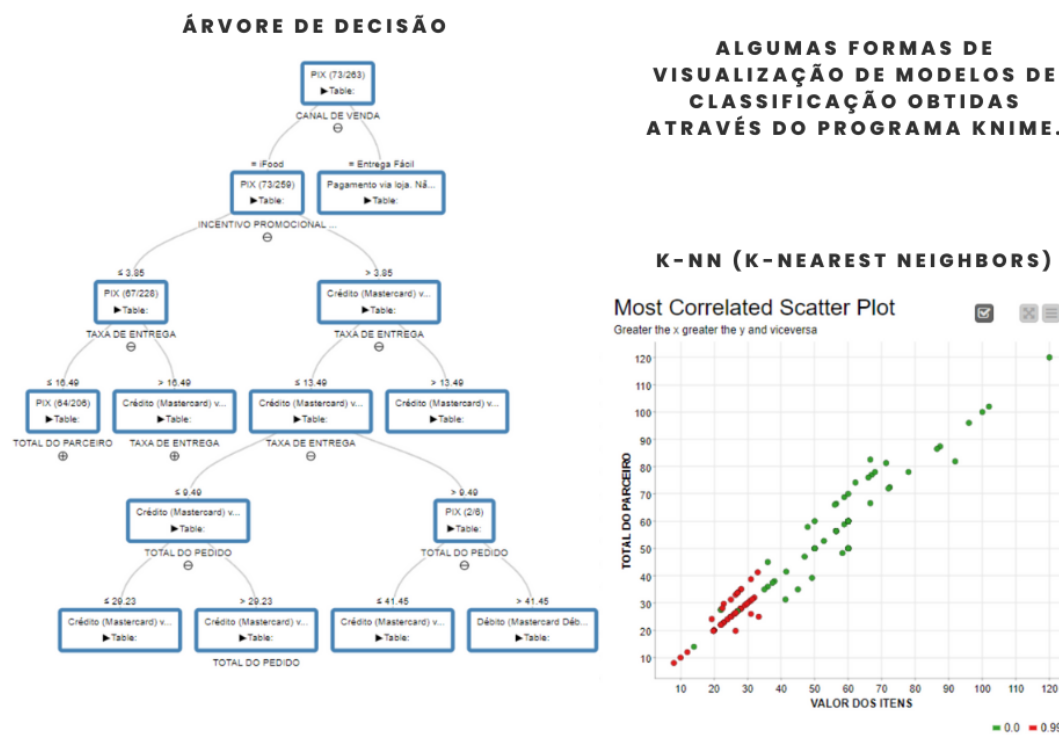
Classificação é uma forma de Aprendizado Supervisionado, em que os dados são enviados juntamente com os rótulos (ou classes), ou seja, o algoritmo já possui informações prévias sobre como os dados são classificados. Para esse tipo de aprendizado, são fornecidos aos algoritmos os dados de “treinamento” e “teste”. Desta forma, é necessário que os dados sejam divididos nessas duas bases de dados distintas para o classificador “aprender” e depois “validar” seus resultados, ou seja, prever a qual classe uma nova entrada de dados pertence. Após a classificação dos dados, é necessário verificar a verdadeira capacidade do classificador em reconhecer as classes apresentadas. Um dos métodos mais utilizados e recomendados para estimar a verdadeira predição dos classificadores no Aprendizado Supervisionado é o método de validação cruzada de *k-folds* (do inglês, *k-fold cross-validation*). Tal método consiste basicamente em dividir a base de dados em k partes, utilizando-se $k-1$ partes para a etapa de treinamento e 1 parte para a etapa de teste, repetindo-se este processo k vezes, e modificando-se os conjuntos de treinamento e teste a cada vez. De forma geral utiliza-se $k = 10$, mas outros valores para k também podem ser adotados (Berrar, 2019; Mitchell, 1997). No Aprendizado não Supervisionado, os algoritmos não possuem informações sobre as classes dos dados, ou seja, eles não possuem informação prévia que os influencie a prever os novos dados. Neste caso, o próprio algoritmo é, portanto, o responsável por analisar os dados com o intuito de separá-los conforme a similaridade e os padrões identificados, agrupando-os em classes ou clusters distintos. Por fim, no Aprendizado por Reforço, os algoritmos aprendem por meio de reforços positivos ou negativos. Caso o algoritmo forneça uma resposta “correta”, recebe uma recompensa (reforço positivo), e caso forneça uma resposta “incorreta”, recebe uma punição (reforço negativo) (Russell; Norvig, 2020). (Santos; Nascimento, 2023)

Existem vários modelos de classificação que podem ser empregados no aprendizado de máquina, como, por exemplo, Árvores de Decisão (do inglês, *Decision Trees*), Florestas Aleatórias (do inglês, *Random Forests*), Máquinas de Vetores de Suporte (SVM), Redes Neurais Artificiais (ANN), Naïve Bayes (NB), K-Vizinhos Mais Próximos (do inglês, *K-Nearest Neighbors* - KNN) dentre outros (Han; Kamber; Pei, 2011).

Nesse contexto, diferentes algoritmos fornecem diferentes representações dos modelos criados. Por exemplo, na Figura 3, é exibida uma árvore de decisão com as regras geradas pelo classificador (lado esquerdo). Do lado direito da imagem, é possível observar o resultado de um classificador k -NN, com duas classes (vermelha e verde), com relação

Figura 3 – Exemplos de modelos de classificação obtidos através do programa knime

CLASSIFICAÇÃO



Fonte: Elaborada pela autora.

a duas variáveis de interesse (“total do parceiro” e “valor dos itens”). A avaliação da classificação é feita por meio de métricas. Alguns exemplos amplamente empregados são: Acurácia, Precisão, Revocação, curva ROC e tempo de execução dos algoritmos (Han; Kamber; Pei, 2011).

2.3 Agrupamento

Diferente da classificação, que analisa um conjunto de dados (de treinamento) pré-rotulado com as classes desejadas, o agrupamento (do inglês, *Clustering*) agrupa objetos de dados sem consultar rótulos de classe. O Agrupamento (*Clustering*) é voltado ao aprendizado não supervisionado, em que as classes do conjunto de dados não são previamente conhecidas. Os dados são agrupados em grupos similares, de maneira que os objetos pertencentes a um mesmo grupo tenham maior similaridade entre os demais do mesmo grupo, e menor similaridade em relação aos elementos de grupos diferentes (Han; Kamber; Pei, 2011).

Os métodos de clusterização podem ser categorizados em métodos de Particiona-

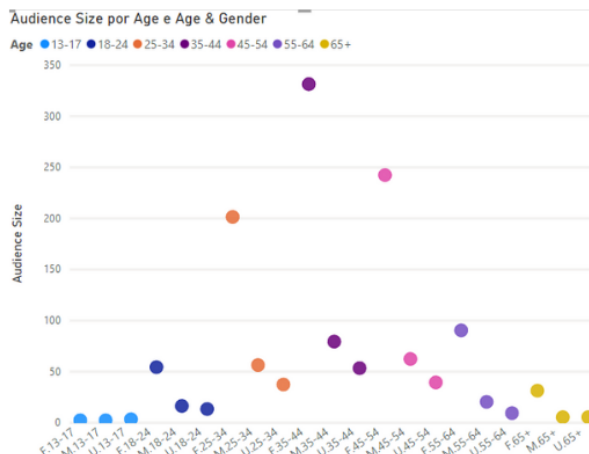
mento (do inglês, *Partitioning*), Métodos Hierárquicos (do inglês, *Hierarchical*), Métodos Baseados em Densidade (do inglês, *Density-based*) e Métodos Baseados em Grade (do inglês, *Grid-based*). Cada abordagem possui suas vantagens e limitações, e a escolha do método adequado depende da natureza dos dados e dos objetivos da análise. Dentre os algoritmos de agrupamento (*clustering*) mais usados estão o *K-Means*, Agrupamento Hierárquico Aglomerativo e Divisivo, e DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), dentre outros (Han; Kamber; Pei, 2011).

A Figura 4 ilustra um exemplo de agrupamento considerando as variáveis “*Audience Size*”, “*Age*” e “*Age & Gender*”. As cores ilustram o resultado dos grupos formados. Já o lado direito da figura ilustra um agrupamento de salgados por tipos. No contexto deste projeto, o agrupamento poderá ser empregado, por exemplo, sobre dados de vendas de salgados, buscando identificar possíveis padrões de compras ou perfis de clientes.

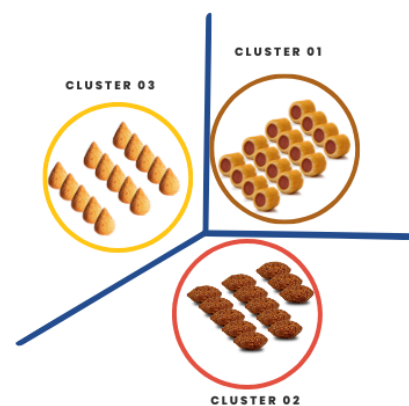
Figura 4 – Exemplo de Agrupamento (*Clustering*)

AGRUPAMENTO (CLUSTERING)

EXEMPLO DE VISUALIZAÇÃO DE AGRUPAMENTO NO PROGRAMA POWER BI. DADOS DE QUANTIDADE DE SEGUIDORES DO INSTAGRAM POR IDADE E GÊNERO.



DESENHO EXEMPLIFICANDO O AGRUPAMENTO DE TIPOS DE SALGADOS



Fonte: Elaborada pela autora.

A visualização de informações é uma forma de manipular dados complexos e representá-los de maneira visual e interativa, facilitando assim não só o entendimento de um profissional já acostumado a lidar com dados, mas também simplificando as informações de forma visual para que usuários comuns que não possuem familiaridade com termos

Figura 5 – Visualização de dados obtida por meio do programa Power BI

VISUALIZAÇÃO DE INFORMAÇÕES



Fonte: Elaborada pela autora.

técnicos posam compreender o que está sendo representado. A Figura 5 mostra exemplos de visualizações de dados geradas por meio do programa Power BI.

Atualmente, existem vários programas de visualização como Tableau, Knime, Power BI, Google Chart, entre outros, que são muito utilizados por possuírem vários recursos, tais como gráficos, mapas, diagramas e infográficos. Esses tipos de recursos auxiliam o usuário a identificar de forma mais simplificada a presença de padrões, tendências, *outliers*, dentre outras técnicas de análise de dados fundamentais na geração de *insights* que auxiliam na tomada de decisão. Todas essas técnicas podem ser empregadas em vários setores, como, por exemplo, áreas de negócios, educação e saúde. No entanto, é preciso também manter cuidados na hora de criar modelos de visualizações, pois visões mal elaboradas podem levar à interpretação incorreta dos dados e, consequentemente, gerar desvantagens para essas organizações.

2.4 Considerações Finais

Neste capítulo foram apresentados os principais conceitos relacionados a esse trabalho de conclusão. As abordagens discutidas envolvem Regras de Associação, Classificação, Agrupamento e Visualização de Informações. Ao longo do desenvolvimento do presente projeto, serão empregadas tais técnicas para a análise de dados de venda, gerenciamento e entrega de salgados.

3 TRABALHOS RELACIONADOS

Neste capítulo são apresentados os principais trabalhos relacionados a esse trabalho de conclusão de curso. Para isso, foi realizada uma pesquisa na literatura buscando identificar ferramentas analíticas que possam ser utilizadas no desenvolvimento deste projeto de conclusão.

3.1 Ferramentas Analíticas

As ferramentas selecionadas foram: *iFood* (Costa; Velloso, 2021), *Knime* (Ihrmark; Tyrkkö, 2023), *Orange* (Demsar; Zupan, 2013), *Power BI* (Pearson *et al.*, 2020). Apesar de não ser uma ferramenta analítica, o *iFood* foi escolhido por ser a principal plataforma de vendas eletrônicas do setor alimentício no mercado nacional. A plataforma digital conecta milhões de consumidores a restaurantes, lojas e entregadores de modo simples e prático. O *iFood* é uma empresa brasileira, referência em *delivery* online na América Latina (Costa; Velloso, 2021). A plataforma fornece, a usuários, ferramentas para realização de vendas, gerenciamento de itens, entregas e um conjunto, mesmo que limitado, de indicadores analíticos aos usuários da plataforma.

A seguir, as Tabelas 1 e 2 mostram uma comparação das ferramentas, considerando, diferentes funcionalidades, relacionadas ao escopo do presente trabalho. As funcionalidades são descritas e explicadas na sequência. As funcionalidades que possuem notas de avaliação receberam valores que variam de 0 (mais baixo, pior avaliação) até 5 (melhor avaliação).

Complexidade de funções que o programa pode executar em termos de Mineração de Dados (Nota de 0 a 5)

Essa funcionalidade refere-se à capacidade que os programas de mineração possuem de manipular e processar dados, utilizando técnicas de inteligência artificial (IA) para descobrir padrões e gerar *insights* a partir dos dados minerados. Entre algumas técnicas que podem ser empregadas, estão análise preditiva, análise descritiva, processamento de linguagem natural (PLN) e outras abordagens de IA.

Geração e download de relatórios (Nota de 0 -5)

Refere-se à capacidade que os programas possuem de gerar relatórios para os usuários, sejam eles informativos, analíticos ou operacionais. Tais relatórios são de grande importância para a regra de negócio analisada no escopo deste projeto.

Funções nativas do programa

As funções nativas de um programa estão presentes em seu código-fonte, e são as funcionalidades básicas que o programa possui para executar tarefas essenciais. Alguns programas já possuem todas as ferramentas necessárias instaladas, enquanto em outros é preciso fazer a instalação de extensões para ampliar seus recursos.

Interface gráfica intuitiva (Nota de 0 -5)

Uma interface gráfica do usuário (do inglês, *Graphical User Interface – GUI*) intuitiva é uma interface que permite ao usuário de software interagir com um programa ou dispositivo de forma natural e automática. Essa característica é avaliada considerando a possibilidade do usuário poder utilizar a ferramenta sem a necessidade de instruções detalhadas ou treinamento prévio. Uma interface gráfica intuitiva é essencial para proporcionar uma experiência de usuário agradável e eficiente, reduzindo a curva de aprendizado e aumentando a produtividade do usuário.

Notificações

As notificações do sistema servem para alertar os usuários sobre possíveis erros e problemas no sistema, além de gerar *feedbacks* e informar quando há atualizações no mesmo.

Obtenção de dados estatísticos

Os dados a serem analisados podem ser obtidos de diversas fontes. Alguns programas permitem que o usuário conecte a ferramenta diretamente a um banco de dados ou incluem a possibilidade de exportar para dentro da plataforma tabelas contendo os dados a serem analisados.

Necessidade de conhecimento em programação para utilizar os recursos de ferramentas do programa (Nota de 0 -5)

Os programas e plataformas citados (*Knime*, *iFood*, *Orange* e *Power BI*) possuem opções para executar tarefas utilizando linguagem de programação para os usuários com um conhecimento mais aprofundado em tecnologia da informação e análise de dados. Porém, os 4 programas caracterizam-se por possuírem interface gráfica intuitiva, onde o usuário consegue utilizar os recursos e manipular os dados sem a necessidade de saber programar.

Preço

Knime e *Orange* são softwares com código-fonte aberto, que permite qualquer pessoa a inspecionar, modificar e aprimorar o sistema. Os recursos adicionais e atualizações também são gratuitos. O *iFood* e o *Power BI* são de código fechado, porém o download

dos programas é gratuito. Caso o usuário queira fazer upgrade dos recursos, precisará pagar.

O *iFood* cobra para fazer toda a parte de marketing e os serviços adicionais oferecidos pela plataforma.

- **Plano Básico:** Por padrão, as taxas cobradas pelo iFood para o plano Básico são:
 - 12% de comissão sobre o valor total dos pedidos feito na plataforma;
 - Taxa de Pagamento Online (3,5% do valor) para pedidos pagos na plataforma;
 - Mensalidade de R\$100,00 para restaurantes que venderem acima de R\$1.800,00 no mês.
- **Plano Entrega:** Por padrão, as taxas cobradas pelo Plano Entrega são:
 - 23% de comissão sobre o valor total dos pedidos feitos na plataforma;
 - Taxa de Pagamento Online (3,5% do valor) para pedidos pagos na plataforma;
 - Mensalidade de R\$130,00 para restaurantes que venderem acima de R\$1.800,00 no mês;

O *Power BI* possui dois planos, o Pro e o Premium,

- O *Power BI* Pro custa R\$ 59,50 por mês e o *Power BI* Premium R\$119,00 por mês.

Suporte ao usuário

O suporte ao usuário tem a finalidade de ajudar o usuário tanto na utilização do programa, quanto na resolução de possíveis problemas que surjam no manuseio da plataforma. O contato com a central de serviço geralmente pode ser feito via chat, telefone ou e-mail.

Tutoriais e manuais de uso da ferramenta em Português (Nota de 0 -5)

Tutoriais são formas de auxiliar e ensinar os passo a passos de como o usuário vai interagir com o programa, bem como auxiliar usuários na resolução de problemas. A documentação correspondente pode ser disponível por meio de vídeos, textos ou softwares interativos.

Tabela 1 – Comparação das ferramentas levantadas considerando suas funcionalidades (primeira parte).

Funcionalidades	<i>iFood</i>	<i>Knime</i>	Programas		<i>Power BI</i>
			<i>Orange</i>		
Complexidade de Funções que o programa pode executar em termos de mineração (Nota de 0 a 5) Download de relatórios (Nota de 0 a 5)	2	5	4		3
	5	5	5		5
Funções Nativas do Programa	Todas as funções vêm habilitadas. Você só paga a mais por serviços oferecidos	Programa com Nodes básicos instalados, para ter acesso a outras funções é preciso fazer o download da função.	Programa vem com Nodes básicos instalados, para ter acesso a outras funções é preciso fazer o download da função.		Todas as funções vêm habilitadas. Você só paga a mais por serviços oferecidos.
Interface gráfica intuitiva (Nota de 0 a 5)	5	2	3		4
Notificações	<i>Feedback</i> , atualizações do sistema ou problemas na loja, o contato é feito através do WhatsApp, email, aplicativo da loja e site de gerenciamento da loja.)	<i>Feedback</i> e atualizações do sistema feito através do email pessoal e tela inicial do programa.	<i>Feedback</i> e atualizações do sistema feito através do email pessoal e tela inicial do programa.		<i>Feedback</i> e atualizações do sistema feito através do email Corporativo e tela inicial do programa.
	Por padrão, o programa disponibiliza dados gerais de venda de pedidos dos últimos 90 dias, mas não fornece dados de usuários.	Permite calcular e ter <i>insights</i> dos dados analisados de maneira mais profunda, mas os dados precisam ser importados de alguma fonte de dados.	Permite calcular e ter <i>insights</i> dos dados analisados de maneira mais profunda, mas os dados precisam ser importados de alguma fonte de dados.		Permite calcular e ter <i>insights</i> dos dados analisados de maneira mais profunda, mas os dados precisam ser importados de alguma fonte de dados.

Fonte: Elaborado pela autora.

Tabela 2 – Comparação das ferramentas levantadas considerando suas funcionalidades (continuação).

Funcionalidades	iFood	Knime	Programas	Orange	Power BI
Necessidade de saber programar para utilizar os recursos de ferramentas do programa (Nota de 0 a 5)	0	0		0	0 (Porém algumas funções dentro do programa podem ser feitas com as seguintes linguagens DAX, M, R e Python)
Preços	Pago	Software de código aberto (Gratuito)	Software de código aberto (Gratuito)		Download gratuito, mas algumas funcionalidades só estão disponíveis na versão paga.
Suporte ao usuário	Chat e materiais de suporte ao usuário (Canal no YouTube, Site com tutorial e aulas ao vivo de técnicas de marketing)	Chat e materiais de suporte ao usuário (Canal no YouTube, Site com tutorial e ajuda de membros da comunidade)	Chat e materiais de suporte ao usuário (Canal no YouTube, Site com tutorial e ajuda de membros da comunidade)		Chat e materiais de suporte ao usuário (Site com tutorial e ajuda de membros da comunidade)
Tutoriais e Manuais de uso da ferramenta em português (Nota de 0 a 5)	5	5		5	5
Usabilidade (Nota de 0 a 5)	5	3		4	4
Visualização de gráficos	Conta com 3 tipos de gráficos (Barra, Linha e Coluna)	Exibe todos os tipos de gráficos e nuvem de palavras	Exibe todos os tipos de gráficos e nuvem de palavras		Exibe todos os tipos de gráficos e nuvem de palavras

Fonte: Elaborado pela autora.

Usabilidade (Nota de 0 -5)

A usabilidade do sistema refere-se à maneira como o usuário executa as tarefas dentro do programa e como interage com as ferramentas. Para que o sistema possua uma boa usabilidade, é necessário ser fácil de aprender, eficiente, satisfatório, memorável e acessível.

Visualização de gráficos

Existem vários tipos de gráficos, com gráficos de barras, de pizza, de linhas, dentre outros tipos de representações visuais. Eles servem para fornecer uma representação visual simplificada de dados complexos, agilizando assim o entendimento do problema a ser analisado. Programas que fornecem este recurso somado a outras técnicas podem ser mais vantajosos para o usuário. A etapa de visualização é de grande importância no contexto deste projeto. Ferramentas visuais podem potencializar a análise de algoritmos de mineração visual e o entendimento das informações analíticas adquiridas pela aplicação. Dessa forma, é de suma importância possuir opções visuais para as diferentes análises e padrões envolvidas no contexto do projeto.

3.2 Trabalhos de Análise de Padrões de Compras

A inteligência artificial (IA) emergiu como uma ferramenta poderosa que pode ser empregada em campos distintos, trazendo otimização de processos, precisão na análise de grandes volumes de dados e retornos financeiros para as organizações por meio de técnicas aplicadas para tomada de decisões. Dentre os trabalhos citados na pesquisa, é possível observar alguns benefícios que podem ser extraídos aplicando as técnicas corretas para cada problema proposto.

Tanto nos trabalhos relacionados quanto no contexto da loja de salgados, o emprego de algumas técnicas são fundamentais na experiência de compra do cliente. Em (Gino *et al.*, 2023), por exemplo, uma das técnicas empregada é o uso de grafos, uma estrutura matemática usada para representar relações entre objetos. Consiste em um conjunto de vértices (ou nós) e um conjunto de arestas (ou arcos) que conectam esses vértices. A utilização dos grafos pode ocorrer tanto na cesta de compras, onde os nós geralmente representam produtos e as arestas representam a correlação entre eles, quanto na parte de Logística e Transporte, pois a mineração de dados em estruturas de grafos pode ajudar na otimização de rotas e entregas, além de detectar gargalos em redes de transporte (Oliveira *et al.*, 2017; Oliveira *et al.*, 2020).

Entender o comportamento do cliente também é uma parte fundamental dentro da área comercial. E saber as reais opiniões dos consumidores sobre produtos e serviços é importante para poder ter um *feedback* do que precisa ser melhorado e ajustado. Apesar dos desafios por conta dos diferentes aspectos linguísticos, em (Ihrmark; Tyrkkö, 2023) os

autores destacam que o emprego da mineração de texto é também um poderoso recurso a ser utilizado. Com ela é possível analisar sentimentos através da opinião de vários usuários e extrair dados comuns entre eles, sejam eles positivos ou negativos em relação a produtos, ou serviços.

No trabalho (Espinoza *et al.*, 2015), os autores também ressaltam a importância de técnicas de mineração de texto, assim como a possibilidade de otimizar estratégias de negócios com base nesse entendimento. Esses são apenas alguns exemplos de um vasto leque de onde técnicas de (IA) que foram somadas ao emprego da regra de associação e outros algoritmos de Mineração de Dados, e que puderam ser aplicadas como ferramenta essencial na previsão de falhas, no melhoramento da qualidade e na otimização do desempenho, fornecendo *insights* valiosos para as diferentes organizações (Ribeiro; Plastino; Murta, 2013).

3.3 Discussão Sobre os Trabalhos Relacionados

Dentre os softwares de análise de dados utilizados para o desenvolvimento desta pesquisa, estão o Portal do *iFood*, que embora seja uma ferramenta que permite a gestão e configuração da loja e do cardápio, também disponibiliza dados de monitoramento e desempenho da loja no *iFood*. A plataforma também fornece a visualização de gráficos de vendas e dicas de marketing para o impulsionamento de vendas (Costa; Velloso, 2021).

O KNIME é uma plataforma de software gratuita utilizada para manipulação de dados complexos e aprendizado de máquina. A ferramenta disponibiliza várias formas de visualização de dados e geração de relatórios de maneira detalhada e interativa (Ihrmark; Tyrkkö, 2023). Por este motivo, o KNIME foi escolhido dentre as outras ferramentas para manipular e analisar os dados referentes à presente pesquisa.

O Orange Data Mining é uma ferramenta muito semelhante ao KNIME, pois também utiliza o “Building Blocks of Analytics” (blocos de construção da análise). Com tais blocos, é possível criar os *workflows* (fluxos de trabalho), manipular e visualizar os dados apenas arrastando e soltando os componentes pré-construídos (nós). No entanto, embora os softwares sejam similares, o KNIME ainda possui mais recursos, fornecendo métricas melhores e uma interface mais robusta (Demsar; Zupan, 2013).

Por fim, o Power BI também foi selecionado para utilização na presente pesquisa. Após a análise de suas principais características, optou-se por empregar a ferramenta na fase de visualização dos dados e padrões obtidos. Essa escolha foi feita porque o Power BI possui diversos recursos, com interatividade, para criação dos *dashboards* de visualização (Pearson *et al.*, 2020).

Os trabalhos relacionados também evidenciam a importância de técnicas de mineração de dados, incluindo a análise de padrões de compras e análise de sentimentos,

bem como a mineração de textos em campos de escrita livre. No desenvolvimento do presente projeto, tais abordagens serão utilizadas e combinadas para a mineração de dados de venda de salgados, bem como a análise de *feedbacks* de clientes, visando melhorar o gerenciamento do negócio e apoiar tomadas de decisão.

4 METODOLOGIA

A metodologia adotada para o desenvolvimento deste trabalho envolve os passos descritos nas seções listadas a seguir. Importante ressaltar que tais passos são complementares, e caminham juntas ao longo do desenvolvimento do projeto.

4.1 Levantamento dos requisitos do projeto

Os primeiros passos na construção do projeto foram a definição do problema e a identificação dos requisitos. O problema escolhido está diretamente relacionado com o negócio da aluna, trazendo assim benefícios para sua formação e profissão. A tarefa de levantamento de requisitos envolveu a coleta e análise dos dados necessários no dia a dia do negócio. Foram realizadas análises de processos internos e levantamento das necessidades de armazenamento e processamento de dados. O objetivo foi identificar quais informações seriam cruciais para a solução proposta e como esses dados poderiam ser coletados e utilizados de maneira eficiente para agregação de valor ao negócio e otimização dos processos de produção.

4.2 Estudo de ferramentas para análise e visualização

Nesta etapa foi realizado um levantamento das ferramentas disponíveis no mercado para análise e visualização de dados. Após selecionadas, foram realizados testes guiados destas ferramentas, focando na familiarização de uso, identificação de tarefas que podem ser realizadas e as contribuições que cada uma pudesse trazer para o projeto.

Foram considerados diversos critérios de avaliação, como usabilidade, funcionalidades, integração com outras ferramentas e custo-benefício. Dentre as opções avaliadas, destacaram-se ferramentas como Knime, Power BI, Orange Data Mine e o portal de parceiros online do iFood. Cada ferramenta foi analisada quanto à sua capacidade de atender às necessidades específicas do projeto, ajustando onde a funcionalidade de cada uma poderia se encaixar melhor na solução final.

4.3 Estudo de algoritmos de mineração de dados

Após a identificação das análises que serão incorporadas à solução proposta, foram utilizadas técnicas de mineração de dados, tais como as regras de associação, mineração de grafos, dentre outras citadas anteriormente (veja no Capítulo 3), que foram instanciadas conforme o problema específico. Foram analisados algoritmos de aprendizado supervisionado e não supervisionado, agrupamento, regressão, dentre outros, considerando suas aplicações práticas e resultados esperados no contexto do projeto. A expectativa é que as técnicas

utilizadas sejam eficazes para extrair informações valiosas dos dados coletados, nas próximas etapas do projeto.

4.4 Modelagem e implementação do banco de dados do problema

Foram identificados as informações a serem armazenadas, e com isso foi modelada a base de dados do projeto. A Figura 6 apresenta o Modelo de Entidade-Relacionamento (MER) desenvolvido para o projeto, ilustrando como os dados estão interligados e estruturados para suportar as análises propostas. Esse modelo ainda está sendo elaborado, e sofrerá alterações até o estágio final do presente projeto.

A partir do MER criado, o banco de dados está sendo criado no MySQL Workbench, uma escolha fundamentada pela sua facilidade de uso, robustez, escalabilidade e compatibilidade com as ferramentas de análise escolhidas, além de ser um software livre.

4.5 Estudo de tecnologias para implementação do protótipo do sistema

Para a implementação do protótipo do sistema, foi realizado um estudo das tecnologias mais adequadas. Dentre as que mais se ajustaram na aplicação e construção do sistema estão as seguintes ferramentas e bibliotecas, que estão sendo utilizadas no desenvolvimento do projeto:

- **Python:** Linguagem de programação principal, selecionada devido à sua versatilidade, facilidade de prototipagem, e vasta gama de bibliotecas para análise de dados.
- **Streamlit:** Framework para construção de aplicações web interativas, facilitando a visualização e interação com os dados.
- **Matplotlib:** Biblioteca de plotagem para criar gráficos estáticos, animados e interativos.
- **Plotly:** Ferramenta de visualização gráfica que oferece suporte para gráficos interativos e de alta qualidade.
- **Scikit-learn (Sklearn):** Biblioteca para aprendizado de máquina em Python, incluindo uma ampla variedade de algoritmos de mineração de dados, que serão úteis ao longo do desenvolvimento do projeto.
- **MySQL:** Sistema de gerenciamento de banco de dados relacional escolhido para armazenar e gerenciar os dados do projeto.

Estas tecnologias foram selecionadas com base em suas capacidades de atender às necessidades específicas do projeto, integrando-se de maneira eficiente para formar uma solução coesa.

Figura 6 – Modelo Entidade Relacional - MER



Fonte: Elaborado pela autora.

4.6 Validação com base nos dados já coletados (históricos) do negócio

Os dados previamente coletados contendo dados de clientes e produtos vendidos foram digitados em planilhas do Excel. Assim como os dados extraídos da plataforma de vendas do iFood.

5 PROPOSTA E DESENVOLVIMENTO

O principal objetivo deste projeto de conclusão de curso consiste em um sistema capaz de armazenar, gerenciar e analisar dados provenientes de venda de salgados e dos processos de produção relacionados. Dessa forma, as contribuições têm duas vertentes principais: a primeira consiste na identificação e definição das análises adequadas a serem realizadas sobre os dados. Com as análises, espera-se dar suporte à tomada de decisão do negócio. A segunda vertente consiste na materialização de um sistema, que englobe as análises definidas, mas que traga funcionalidades para fomentar o gerenciamento dos dados e a visualização dos resultados das análises propostas.

5.1 Vertente de Análise de Dados

Para análise de dados, serão considerados métodos de classificação, regras de associação e agrupamento. Os dados utilizados são provenientes das vendas históricas do negócio, feitas dentro e fora do aplicativo iFood. A ideia dessa parte do projeto é tirar proveito das ferramentas analíticas existentes, que foram discutidas anteriormente: Knime, Orange e PowerBI. Assim, serão aplicados os algoritmos com base nos dados disponíveis, buscando fornecer padrões de compra, melhores percepções de controle de estoque, e outras informações que possam fomentar uma melhor tomada de decisões.

5.2 Vertente do Sistema Proposto

Com o intuito de sanar os problemas atuais já descritos ao longo do projeto, a criação e implementação do sistema proposto passará por várias etapas. Especificamente, a etapa do desenvolvimento do protótipo do sistema é composta por uma arquitetura de diversos componentes que interagem entre si. A arquitetura é ilustrada na Figura 7. Este sistema visa proporcionar uma interface interativa para análise e visualização de dados de clientes, produtos vendidos, e controle de estoque.

A Figura 8 mostra os diversos componentes do sistema, e como eles interagem entre si. Nas subseções a seguir os componentes são descritos com mais detalhes.

5.2.1 Usuário

No contexto do sistema proposto, o usuário é o ponto de partida e o responsável por toda interação entre os processos. Busca-se desenvolver uma interface amigável, utilizando o framework com **Streamlit** de modo a permitir que o usuário possa interagir com os dados do sistema de forma intuitiva e eficiente. Além disso, a interface fornecerá ao usuário

Figura 7 – Arquitetura do sistema proposto.

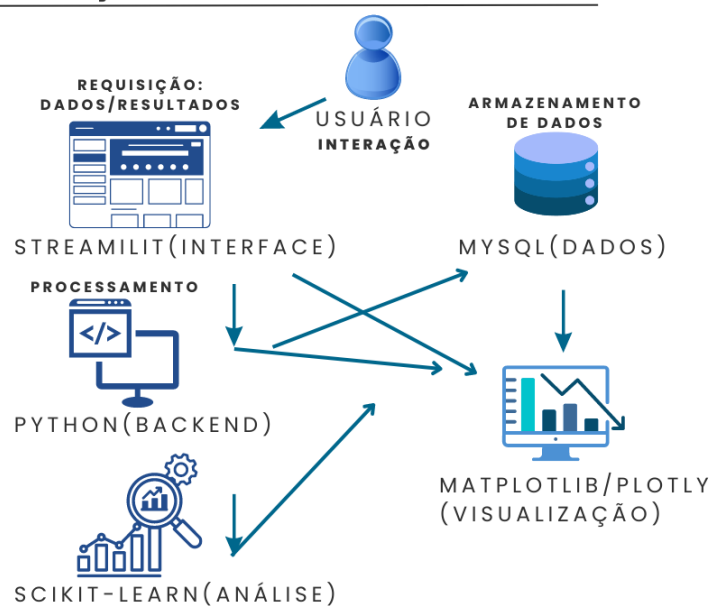
INTERAÇÃO DOS COMPONENTES DO SISTEMA



Fonte: Elaborado pela autora.

Figura 8 – Componentes do sistema proposto.

INTERAÇÃO DOS COMPONENTES DO SISTEMA



Fonte: Elaborado pela autora.

a opção de inserir e consultar a base de dados, bem como executar e visualizar os resultados de análise.

5.2.2 Streamlit (Interface)

O **Streamlit** é um framework de código aberto, com API intuitiva que permite a criação de uma interface web interativa para sistemas em Python. Com o Streamlit, o usuário pode utilizar a interface criada para visualizar e interagir com os dados.

No contexto deste projeto, a interface desenvolvida deve ser capaz de fornecer:

- Acesso e cadastro dos dados dos clientes.
- Acesso e cadastro dos dados de produtos.
- Acesso e cadastro dos ingredientes.
- Solicitação e gerenciamento de pedidos.
- Exibição de relatórios, gráficos e dashboards interativos.
- Navegação intuitiva para facilitar a análise dos dados.
- Execução de algoritmos de mineração de dados e análise dos resultados obtidos.

5.2.3 MySQL (Dados)

MySQL é o sistema de gerenciamento de banco de dados relacional escolhido para armazenar e gerenciar os dados do. No contexto deste projeto, o banco de dados será responsável por:

- Facilitar consultas eficientes para análise e visualização de dados.
- Armazenamento estruturado dos dados de clientes e produtos.
- Garantir a segurança dos dados do sistema.
- Garantir a integridade e consistência dos dados, evitando duplicidades e inconsistências nos dados para garantir a confiabilidade.

5.2.4 Python (Backend)

O **Python** é a linguagem de programação escolhida para o desenvolvimento do sistema. Os códigos criados em Python têm como função conectar todos os componentes e permitir o processamento e análise dos dados.

As responsabilidades do *backend* em Python, no contexto deste projeto, incluem:

- Criação e alteração do Layout da interface do sistema.
- Conexão com o banco de dados MySQL para buscar e inserir dados.
- Processamento dos dados para preparar a visualização e análise.
- Implementação de lógica de negócios para suportar funcionalidades da interface.
- Implementação das análises utilizando algoritmos de mineração de dados existentes e ferramentas de análise escolhidas para dar suporte a essa tarefa.

5.2.5 Matplotlib/Plotly (Visualização)

As bibliotecas **Matplotlib** e **Plotly** são utilizadas para criar gráficos e visualizações de alta qualidade e interatividade.

As funcionalidades das ferramentas de visualização, relacionadas a este projeto, incluem:

- Criação de gráficos estáticos e interativos para análise visual dos dados.
- Integração com Streamlit para exibir gráficos diretamente na interface web.
- Suporte para diversos tipos de gráficos, como linhas, barras, e gráficos de dispersão.

5.2.6 Scikit-learn (Análise)

Scikit-learn é a biblioteca utilizada para implementar algoritmos de aprendizado de máquina e análise de dados.

As funcionalidades de análise dessa biblioteca, dentro do contexto deste projeto, incluem:

- Treinamento de modelos preditivos utilizando dados históricos.
- Análise de desempenho e geração de insights a partir dos dados.
- Suporte para diversas técnicas de aprendizado de máquina, como regressão, classificação e clustering.

Além dos algoritmos disponíveis em Python, espera-se integrar o sistema proposto a ferramentas já existentes, mencionadas anteriormente nesta monografia, tais como: Orange, Knime e PowerBI.

5.2.7 Testes

Na fase de testes, será abordado o processo de verificação e validação do protótipo para garantir que todas as funcionalidades estejam operando corretamente e que o sistema atenda aos requisitos especificados no projeto. Serão feitos testes de unidade, nas diferentes funcionalidades, além de validação do desempenho do sistema e validação dos resultados. Essa etapa será realizada para garantir que o sistema está entregando tudo que foi solicitado e se está rodando sem erros graves de lógica e execução, que possam comprometer o propósito do sistema.

5.3 Fluxo de Interação Entre Componentes

Conforme ilustrado na Figura 8, o fluxo de interação entre os componentes do sistema acontece da seguinte forma:

1. O usuário interage com a interface Streamlit, solicitando o cadastro, visualização ou análise de dados.
2. A interface Streamlit envia solicitações para o backend Python, que processa os pedidos do usuário. A interface também acessa a base de dados criada, podendo inserir, consultar e alterar dados de dentro do MySQL.
3. O backend Python se conecta ao banco de dados MySQL para buscar os dados necessários.
4. Os dados recuperados são processados pelo backend Python e, se necessário, analisados utilizando Scikit-learn.
5. Os resultados processados são visualizados utilizando Matplotlib ou Plotly.
6. As visualizações são exibidas na interface Streamlit para o usuário.

6 RESULTADOS

Neste capítulo, são apresentados os resultados obtidos ao longo do desenvolvimento deste projeto de TCC. Na Seção 6.1 é apresentada uma visão geral do cenário de aplicação e organização de dados. Na Seção 6.2 é apresentada a coleta de dados e a modelagem criada para dar suporte à coleta e gerenciamento dos dados do sistema proposto, o SquareSys. Na Seção 6.3 são descritos diversos cenários de análise que foram gerados, por meio da exploração de algoritmos de mineração de dados existentes e ferramentas analíticas selecionadas. Finalmente, na Seção 6.6 são apresentadas as considerações finais deste capítulo.

6.1 Análise de Cenário e Organização dos Dados

O primeiro passo realizado foi a análise do cenário de aplicação, considerando a rotina do negócio, as ferramentas utilizadas e plataformas disponíveis. Buscou-se identificar os seguintes aspectos:

- Disponibilidade de dados históricos para análise e criação de uma base de dado local para coleta;
- Organização dos dados de pedidos feitos fora da plataforma;
- Cenários de análise, de modo a obter *insights* dos dados para tomada de decisão no cenário de aplicação;
- Automatização do cálculo de reposição de insumos com base na produção e venda de salgados.

6.2 Base de Dados do SquareSys

Para viabilizar uma análise inicial dos dados e identificação de possíveis técnicas e algoritmos a serem empregados no cenário em questão, foi realizada uma coleta de dados iniciais de pedidos, feitos diretamente na loja, ao longo do desenvolvimento do projeto. Dessa forma, os dados a serem analisados referem-se às vendas diretas realizadas na loja Salgados do Quadrado no período de 1º de janeiro de 2024 até agosto do mesmo ano. Durante este período, todas as transações foram, inicialmente, registradas manualmente em talões de pedidos. Esses registros foram posteriormente transferidos para uma planilha de Excel.

Foi modelada e criada uma base de dados para o negócio. Com isso, permite-se gerenciar os dados coletados no domínio de aplicação para pedidos futuros, assegurando assim uma base histórica para possíveis análises futuras e apoio à tomada de decisão.

Tabela 3 – Resumo dos resultados obtidos.

Tarefa realizada	Resultados
Entendimento do problema (Seção 6.2)	Identificadas informações (atributos) a serem coletadas de pedidos e dados do negócio a serem armazenados e gerenciados pela aplicação.
Coleta de dados (Seção 6.2)	Foram coletados dados de 198 clientes e um total de 223 pedidos, populando a base criada de maneira apropriada. Já na plataforma iFood, somente dados sumarizados são apresentados, e não há separação entre itens conforme o pedido ou identificador de cliente. Dessa forma, os dados coletados não poderão ser analisados com os algoritmos de mineração de dados, mas serão utilizados nas visualizações de agregações disponibilizadas pela ferramenta proposta, a SquareSys.
Modelagem da base de dados (Seção 6.2)	Foi modelada e implementada a base de dados da aplicação, considerando os dados coletados em pedidos diretamente pelo negócio, e dados coletados e disponibilizados a respeito de pedidos realizados pelo aplicativo iFood.
Cenários de análise (Seção 6.3)	Foi feito um levantamento de cenários de análise adequados com base nos dados disponíveis e nos objetivos do cenário de aplicação.
Análise de dados (Seção 6.4)	Os dados coletados foram pré-processados e analisados usando os seguintes algoritmos: Foram utilizados diversos algoritmos durante esta etapa, como o algoritmo Apriori para regra de associação, e outros para mineração de texto e análise de sentimentos. Com base nessas técnicas, foram identificados padrões significativos para geração de insights valiosos para auxiliar tomada de decisão. Estes achados destacam, por exemplo, a relevância de determinados comportamentos dos consumidores e a predominância de sentimentos específicos em relação aos produtos analisados.
Dashboard de visualização (Seção 6.4.1)	Foi criado um dashboard para visualização das informações agregadas de pedido. Nessa etapa, tanto os dados disponibilizados pelo aplicativo do iFood como os dados coletados pela aplicação foram considerados. Como resultado, foi gerado um conjunto de visualizações.
Aplicação SquareSys	Foi criada a aplicação SquareSys para implementação da coleta, busca, análise e visualização dos dados, padrões e agregações geradas nas tarefas listadas nessa tabela.

Fonte: Elaborada pela autora.

O objetivo dessa transferência foi permitir a soma das quantidades de salgados vendidos e o cálculo do valor total das vendas no período fornecido. Essa metodologia garantiu uma maior precisão dos dados, possibilitando uma análise detalhada e confiável das vendas realizadas.

No entanto, mesmo com a transcrição dos pedidos para a planilha, há um risco de que alguma informação tenha sido perdida ao longo do caminho. Isso pode ter ocorrido devido ao extravio de algum papel ou à falha na transcrição de alguns pedidos, especialmente em dias de maior movimento. Alguns pedidos poderiam não ter sido transcritos corretamente ou poderiam ter sido perdidos no processo de registro manual.

A partir da planilha de Excel, foi possível identificar padrões de consumo, picos de vendas e produtos mais populares. Esses insights foram essenciais para otimizar o estoque, melhorar o atendimento ao cliente e planejar estratégias de marketing mais eficazes. No entanto, é importante considerar os possíveis erros de registro ao analisar os dados para garantir que as conclusões sejam as mais precisas possíveis. Embora a planilha tenha dado um bom suporte a loja. As informações não tinham dados gráficos para um melhor entendimento dos dados coletados.

Estes mesmos dados serão integrados a um banco de dados MYSQL que é onde a aplicação está ligada. Isso irá garantir ainda mais uma confiabilidade na aplicação.

Paralelamente, foi feito um levantamento dos dados disponibilizados pelo iFood, uma vez que se trata da plataforma utilizada atualmente pelo negócio, em que a maioria dos pedidos é realizada.

6.3 Cenários de Análise

Com a implementação do SquareSys os usuários que vão interagir com o sistema, poderão obter na aba de visualização de relatórios do sistema da loja 'Salgados do Quadrado' diferentes tipos de gráficos para facilitar a análise e a tomada de decisões estratégicas. Os gráficos terão como principal objetivo fornecer insights detalhados sobre os dados dos clientes, produtos e insumos.

Primeiramente, para visualizar a quantidade de produtos vendidos em um período específico, categorizados por sabores de salgados, será utilizado um *gráfico de barras*. Esse gráfico facilita a visualização e pode mostrar a quantidade de salgados vendida durante um período especificado, permitindo uma análise detalhada das preferências dos clientes em diferentes momentos. Alinhado a um *gráfico de linhas*, ele pode mostrar os picos de venda de cada produto nos dias, horários e meses de maior movimento, além dos dias da semana e dos meses de maior faturamento.

Ainda se tratando da análise de produtos, será utilizado um *gráfico de pizza* para mostrar a quantidade percentual de cada tipo de salgado, que será representado cada

um com um cor diferente para facilitar a leitura, os dados tem como intuito mostrar o total de vendas, destacando os produtos mais populares. Um *histograma* será usado para mostrar a quantidade de produtos ainda disponíveis em estoque. Essa visualização facilita o gerenciamento e a reposição de produtos, permitindo que a loja mantenha um controle eficiente da quantidade de cada sabor de salgado e evite faltas ou excessos de estoque. O mesmo se aplica aos insumos, onde é crucial manter um controle de estoque não só baseado na quantidade dos produtos, mas também na data de validade de cada insumo.

Para identificar e categorizar os clientes mais compram do estabelecimento, o sistema utiliza um *gráfico de barras lateral* que exibe os 10 clientes que mais recorrentes da loja. Esse gráfico mostra a quantidade de compras realizadas por comprador, permitindo a implementação de estratégias de fidelização personalizadas, como criar combos personalizados ou aplicar promoções para fidelizar ainda mais esses clientes. O mesmo vale para os top 10 clientes que menos compram, possibilitando mapear esses indivíduos e criar estratégias para incentivá-los a voltar a consumir no estabelecimento.

Além disso, o sistema utiliza outro *gráfico de barras* para mostrar os 5 sabores de salgados que mais vendem, permitindo foco nos produtos de maior aceitação. Isso é útil tanto na criação de combos para venda quanto na manutenção de estoque, garantindo que esses produtos nunca falem. O sistema também deve emitir um alerta quando o estoque desses produtos estiver abaixo do aceitável. Para os produtos que vendem menos, o alerta será emitido com uma métrica diferente.

Atualmente as vendas por clientes são divididos em homens, mulheres e empresas. Para determinar a quantidade de salgados vendida em um mês específico, será utilizado um *gráfico de dispersão* para identificar padrões ou tendências e revelar diferenças de comportamento de compra de cada grupo, sendo uma boa forma de analisar picos de demanda e sazonalidade.

A utilização de um *mapa* também será aplicada para demarcar as regiões que concentram a maior quantidade de clientes, possibilitando a criação de promoções como, por exemplo, entrega grátis para determinadas regiões.

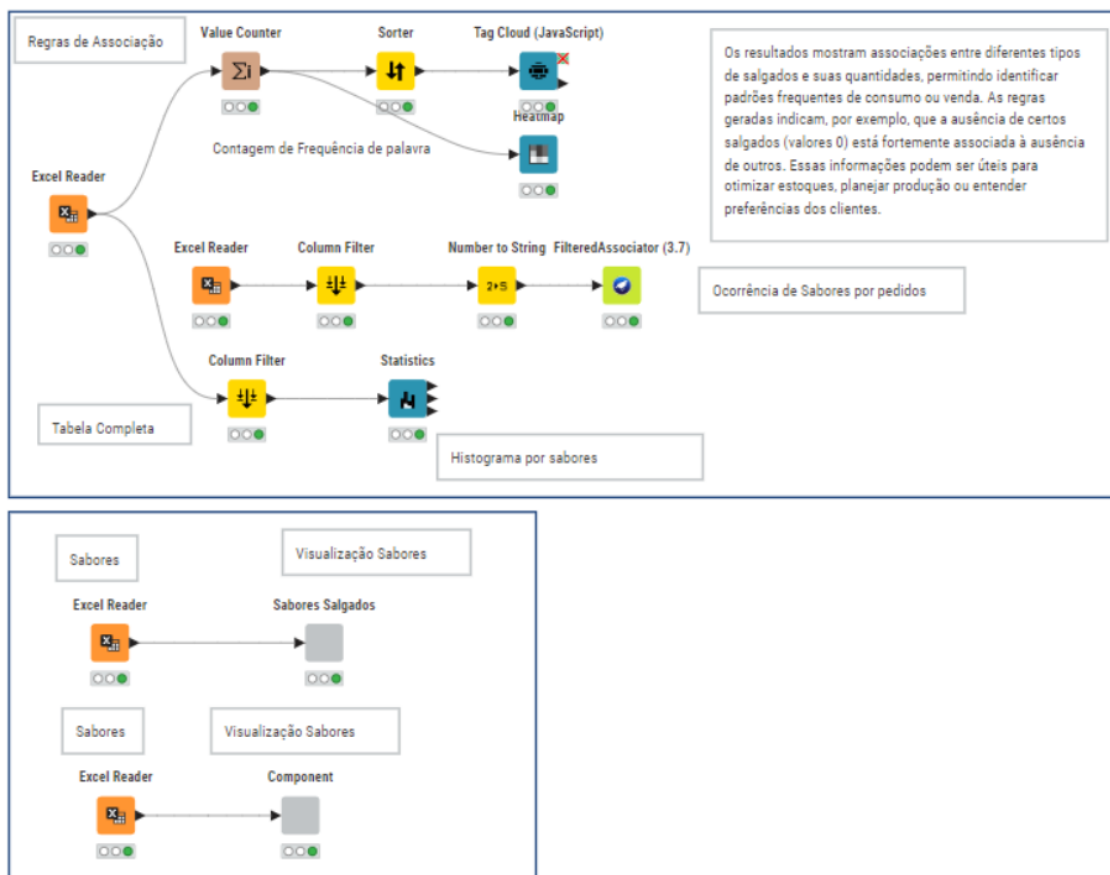
Com essas e outras ferramentas de visualização, o SquareSys proporciona uma visão clara e detalhada do desempenho geral da loja 'Salgados do Quadrado', permitindo identificar oportunidades de melhoria e maximizar a satisfação dos clientes e a eficiência operacional. A análise visual é um grande facilitador para a gestão estratégica da loja, pois pode possibilitar um melhor entendimento do comportamento do mercado e das preferências dos clientes.

6.4 Análise de Dados

Embora o SquareSys conte com uma aba de Dashboard para análise dos dados cadastrados, as informações já obtidas passaram por análises mais detalhadas no KNIME. Essa ferramenta permitiu o cálculo de métricas como regras de associação de produtos, análise de sentimentos, avaliação de desempenho, entre outras.

As análises realizadas no KNIME oferecem uma visão mais profunda e estratégica sobre o desempenho dos produtos e das operações da loja. A utilização de regras de associação, por exemplo, possibilitou a identificação de padrões de compra entre os produtos, permitindo a criação de promoções e ofertas direcionadas que podem aumentar as vendas e melhorar a experiência do cliente. Saber quais produtos são frequentemente comprados juntos, é possível criar campanhas de marketing combinando estes produtos, gerando maior eficiência e retorno de vendas. A Figura 9 mostra o workflow criado no KNIME para essa análise, e a Figura 10 mostra as regras de associação encontradas.

Figura 9 – Workflow construído para gerar regras de associação a partir do arquivo contendo dados binários de produtos.



Fonte: Elaborado pela autora.

Com base nos principais resultados obtidos, foi possível notar que a Coxinha é frequentemente comprada com outros salgados como Bolinha de Queijo, Empada de Frango,

Figura 10 – Resultado obtido através da aplicação do modelo Apriori utilizado na mineração de dados para busca identificar regras de associação em grandes conjuntos de dados.

```

Weka Node View - 5:37 - FilteredAssociator (3.7)
File
FilteredAssociator using weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 on data filtered through weka.filters.MultiFilter -F "weka.filters.unsupervised.attribute.ReplaceMissingValues

Filtered Header
@relation Weka-Instances-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.MultiFilter-Fweka.filters.unsupervised.attribute.ReplaceMissingValues

@attribute BolinhadeQueijo {0,1}
@attribute Coxinha {0,1}
@attribute EmpadadeAlho-Por {0,1}
@attribute EmpadadeFrango {0,1}
@attribute EnroladinhodeSalsicha {0,1}
@attribute Kibe {0,1}
@attribute MiniChurros {0,1}
@attribute RisoledMilho {0,1}
@attribute QuadradinhodeCalabresacomQueijo {0,1}

@data

Associator Model
Apriori
=====
Minimum support: 0.4 (89 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15

Size of set of large itemsets L(2): 16

Size of set of large itemsets L(3): 4

Best rules found:

1. BolinhadeQueijo=1 EmpadadeFrango=1 94 ==> Coxinha=1 94 <conf:(1)> lift:(1.1) lev:(0.04) [8] conv:(8.85)
2. BolinhadeQueijo=1 EnroladinhodeSalsicha=1 92 ==> Coxinha=1 89 <conf:(0.97)> lift:(1.07) lev:(0.03) [5] conv:(2.17)
3. EnroladinhodeSalsicha=1 108 ==> Coxinha=1 104 <conf:(0.96)> lift:(1.06) lev:(0.03) [6] conv:(2.03)
4. BolinhadeQueijo=1 MiniChurros=1 101 ==> Coxinha=1 97 <conf:(0.96)> lift:(1.06) lev:(0.02) [5] conv:(1.9)
5. BolinhadeQueijo=1 150 ==> Coxinha=1 144 <conf:(0.96)> lift:(1.06) lev:(0.04) [8] conv:(2.02)
6. BolinhadeQueijo=1 QuadradinhodeCalabresacomQueijo=1 95 ==> Coxinha=1 90 <conf:(0.95)> lift:(1.05) lev:(0.02) [3] conv:(1.49)
7. Kibe=1 120 ==> Coxinha=1 113 <conf:(0.94)> lift:(1.04) lev:(0.02) [4] conv:(1.41)
8. QuadradinhodeCalabresacomQueijo=1 109 ==> Coxinha=1 102 <conf:(0.94)> lift:(1.03) lev:(0.01) [3] conv:(1.28)
9. EmpadadeFrango=1 138 ==> Coxinha=1 129 <conf:(0.93)> lift:(1.03) lev:(0.02) [3] conv:(1.3)
10. MiniChurros=1 125 ==> Coxinha=1 115 <conf:(0.92)> lift:(1.02) lev:(0.01) [1] conv:(1.07)

```

Fonte: Elaborado pela autora.

e Enroladinho de Salsicha. A confiança elevada nas regras sugere que esses produtos têm uma forte associação nas compras. Foram levadas em consideram somente as regras de maior confiança para ocorrer. Por exemplo, a seguir são listados resultados das Regras de Associação:

- Tamanho dos Conjuntos de Itemsets (L):

L(1) = 15: Foram encontrados 15 conjuntos de itens frequentes de tamanho 1 (itens únicos).

L(2) = 16: Foram encontrados 16 conjuntos de itens frequentes de tamanho 2 (pares de itens).

L(3) = 4: Foram encontrados 4 conjuntos de itens frequentes de tamanho 3 (trios de itens).

Esses números indicam quantos conjuntos de itens diferentes atingiram o suporte mínimo em cada categoria.

- Regra 1:

BolinhadQueijo=1 e EmpadadeFrango=1 ==> Coxinha=1

Confiança: 1 (100%)

Lift: 1.1

Leverage: 0.04

Convicção: 8.85

- **Interpretação:** Sempre que um cliente compra uma Bolinha de Queijo e uma Empada de Frango juntos, ele também compra uma Coxinha. Essa regra tem 100% de confiança, o que significa que todas às vezes que esse par de itens foi comprado, a Coxinha também foi comprada. O lift de 1.1 sugere que a compra da Coxinha é 10% provável na presença desses itens do que em uma situação aleatória. A alta convicção de 8.85 indica que essa regra tem uma forte relação entre os itens.

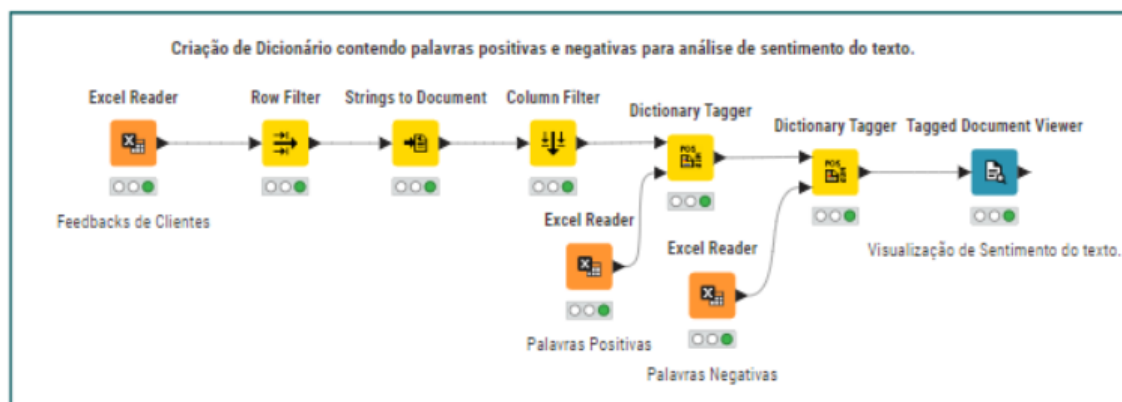
A análise de sentimentos aplicada aos *feedbacks* dos clientes oferece *insights* valiosos sobre a percepção da marca e dos produtos. Ao identificar e categorizar as opiniões como positivas, negativas ou neutras, a loja pode ajustar suas estratégias para melhorar a satisfação do cliente. Isso não apenas fortalece o relacionamento com os consumidores, mas também ajuda a loja a se antecipar a problemas potenciais, resolvendo-os antes que se tornem críticas mais amplas. A Figura 11 mostra o *workflow* criado para essa análise, e a Figura 12.

A avaliação de desempenho das vendas e dos produtos ajuda a identificar áreas de sucesso e oportunidades de melhoria. Com essa análise, a loja pode tomar decisões mais informadas sobre o gerenciamento de estoques, o desenvolvimento de novos produtos e o planejamento de campanhas de vendas. As métricas detalhadas fornecem uma base sólida para a criação de estratégias que não apenas mantêm a loja competitiva, mas também promovem seu crescimento a longo prazo.

Na parte de análise de sentimento e mineração de foram feitas duas etapas, a primeira foi a criação de um dicionário contendo palavras positivas e negativas para facilitar a interpretação de *feedbacks* com as palavras mais recorrentes em relação à avaliação de salgados, pois são palavras específicas para o tipo de produto, como, por exemplo, quentinho, dourado, cru, etc. Foi feito também a multiplicação dos termos TF e IDF. A Figura 13 mostra o *workflow* criado. Frequência do Termo (TF), que conta quantas vezes uma palavra aparece em um documento e a Frequência Inversa do Documento (IDF), que mede quão rara essa palavra é em uma coleção de documentos. Ao mutiplicar o TF pelo IDF, é possível obter o TF-IDF, que equilibra esses dois fatores. Isso significa que o TF-IDF ajuda a encontrar palavras que são não apenas frequentes, mas também significativas para diferenciar um documento dos outros, tornando-o uma ferramenta útil para identificar o conteúdo mais relevante em textos.

A utilização do método TF-IDF para analisar os *feedbacks* dos clientes foi uma

Figura 11 – Workflow construído cruzando arquivos contendo palavras positivas e negativa e concatenadas ao documento de feedback para gerar um dicionário de palavras de sentimento.



Fonte: Elaborado pela autora.

Figura 12 – Resultado de saída da análise das palavras positivas e negativas identificadas pelo dicionário.

☐ **Salgados maravilhosos mas teve um problema , faltou o meu refrigerante .**

	Positive	Negative
1	Salgados maravilhosos	mas teve um problema , faltou o meu refrigerante .
2	Salgados maravilhosos	mas teve um problema , faltou o meu refrigerante .

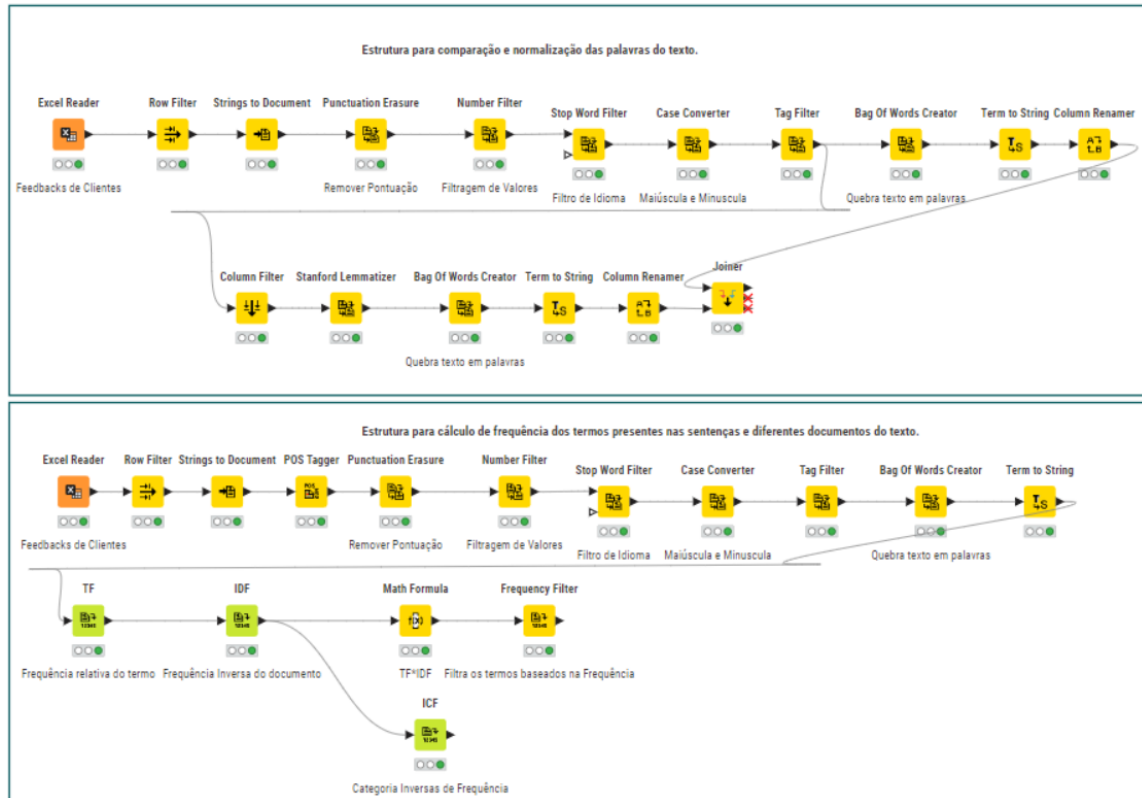
Fonte: Elaborado pela autora.

excelente forma de identificar palavras-chave importantes e entender as áreas que precisam de atenção. A Figura 14 mostra um exemplo de resultado dessa análise. Por exemplo, ao aplicar essa técnica, descobrimos que a palavra “Goiabada” se destacou como sendo especialmente relevante nos comentários dos clientes. Isso se refere ao sabor “Romeu e Julieta”, que, apesar de ter sido descontinuado por questões logísticas na fabricação, recebeu muitos elogios. Esse feedback positivo sugere que a loja poderia ou reintroduzir esse sabor ao cardápio, ou criar algum sabor semelhante poderia ser uma estratégia interessante para atender às expectativas dos clientes e melhorar a oferta da loja.

Foram feitas Análise de Tendências usando o nó Time Series Generator para identificar padrões ao longo do tempo, como sazonalidade ou tendências nos pedidos, conforme mostra a Figura 16. O node “Statistics” retorna como resultado de visualização uma tabela contendo vários dados, mas os resultados obtidos também podem ser vistos por meio de um dashboard interativo que facilita de forma visual a leitura e interpretação dos dados.

A Figura 15 mostra uma visão geral do dashboard criado.

Figura 13 – Workflow TF-DF construído para ajuda a ranquear os documentos mais relevantes para a consulta, considerando a frequência das palavras-chave nos documentos.



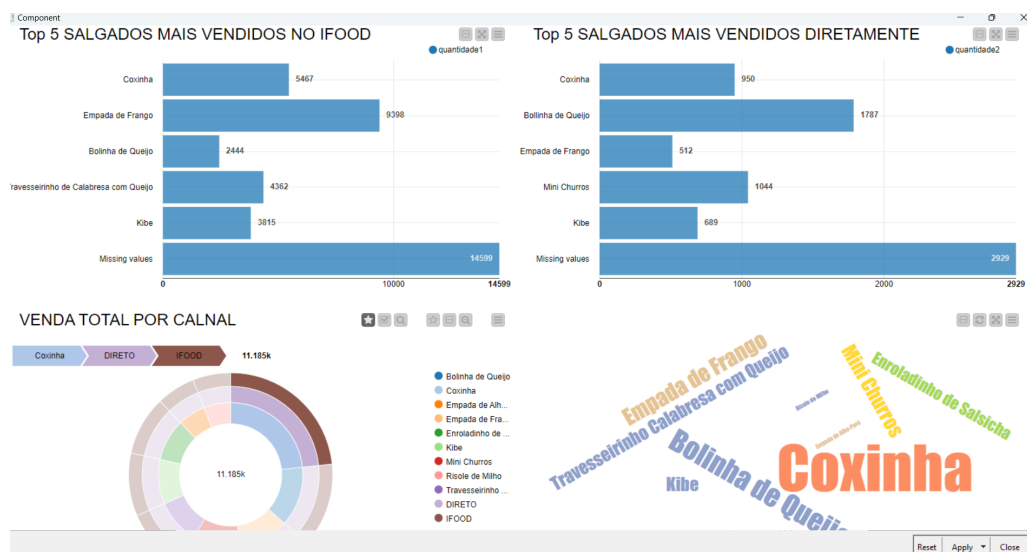
Fonte: Elaborado pela autora.

Figura 14 – Nuvem de palavras gerada como resultado do node TF-IDF, onde a palavra goiabada aparece num tamanho maior comparado a outras palavras, indicando uma maior citação da mesma dentre os textos analisados.



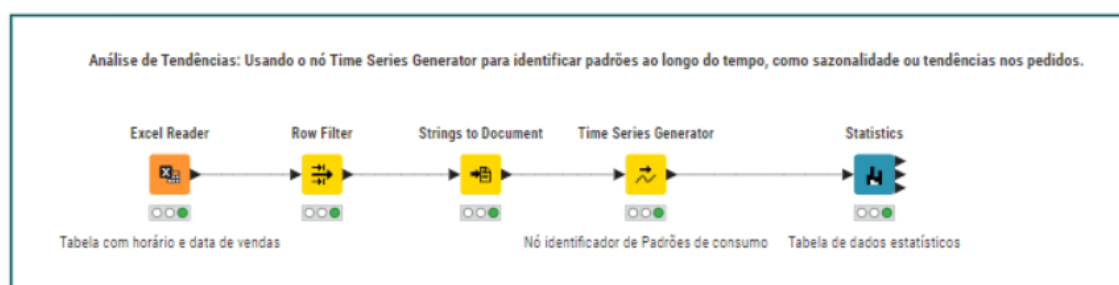
Fonte: Elaborado pela autora.

Figura 15 – Dashboard de dados dos salgados



Fonte: Elaborado pela autora.

Figura 16 – Workflow de Análise de Tendências: Usando o nó Time Series Generator para identificar padrões ao longo do tempo, como sazonalidade ou tendências nos pedidos.



Fonte: Elaborado pela autora.

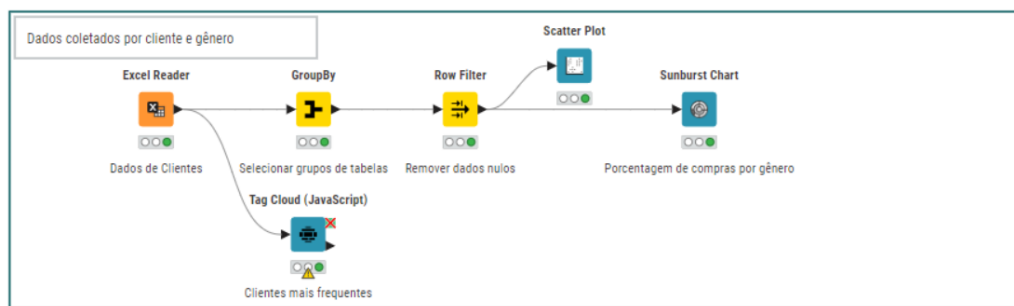
O relatório de análise de pedidos na loja apresenta uma visão detalhada dos dados coletados, com foco nos clientes que mais compraram, dias e horários de maior movimento, consumo por gênero e uma análise individual dos sabores de salgados. A tabela de pedidos inclui informações como nome do cliente, ID do pedido, data, hora, gênero, itens pedidos (como empada de frango, bolinha de queijo, coxinha, kibe, risole de milho, enroladinho de salsicha, empada de alho-poró, mini churros e travesseirinho de calabresa com queijo), quantidade e valor total do pedido. A análise destaca os top 20 clientes que mais compraram durante o período, e a quantidade total de pedidos realizados por cada um.

Os dados da tabela analisada também mostram os dias e horários com maior volume de vendas, organizados de forma decrescente, o que permite identificar os padrões de movimento na loja. Essas informações são essenciais para otimizar a alocação de pessoal nos horários de pico e para identificar quais os dias com movimentação incomum, como

feriados ou eventos sazonais, como, por exemplo, um aumento ocasionado pelas festas juninas que ocorrem entre junho e julho. Nesses períodos, é recomendável não apenas reforçar o estoque com sabores típicos dessas festividades, como coxinha de frango e risole de milho, mas também criar kits de festas temáticos para aumentar as vendas.

A análise de consumo por gênero foi importante para direcionar as campanhas de marketing de forma mais eficaz, alcançando tanto os grupos que mais consomem quanto criando campanhas que possam atrair aqueles menos consumistas, ampliando o alcance da marca. A Figura 17 mostra o workflow correspondente.

Figura 17 – Coleta de dados por gênero



Fonte: Elaborado pela autora.

Utilizando o node Statistics, foi realizada uma análise detalhada dos 223 pedidos registrados, focando nos sabores de forma individual. Por exemplo, o campo “Bolinha de Queijo” possui dois dados separados por (:) onde o primeiro dado se refere a quantidade pedida e o segundo em quantos pedidos o sabor estava presente.

- A análise Top 20 (veja na Figura 18) mostra a frequência dos valores mais comuns no campo “Bolinha de Queijo”:
 - 0: 73 pedidos não incluíram bolinhas de queijo.
 - 25: 39 pedidos incluíram 25 bolinhas de queijo.
 - 50: 22 pedidos incluíram 50 bolinhas de queijo.
 - 20: 19 pedidos incluíram 20 bolinhas de queijo.
 - 12: 11 pedidos incluíram 12 bolinhas de queijo.
 - 100: 8 pedidos incluíram 100 bolinhas de queijo.

Esses dados podem ser visualizados em dashboards interativos, o que facilita a interpretação. Saber a média de salgados consumidos é importante para promover vendas em maior quantidade, considerando a possibilidade de aplicar promoções ou descontos para pedidos intermediários de bolinhas de queijo, ou combinação com outros sabores usando como base combos contendo 25 ou 50 unidades o qual foi a média de consumo obtida.

Figura 18 – TOP 20 dados analisados pelos node “Statistics”

Cliente	ID Pedido	Gênero	Data	Hora	Bolinha de Queijo	Coxinha	Empada de Alho-Poró	Empada de Frango	Enroladinho de Salsicha	Kibe	Mini Churros	Risole de Milho
No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 1	No. missings: 1	No. missings: 0	No. missings: 0
Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:
Bruno Smart Fir : 8	1 : 1	Feminino : 159	2024-06-28 : 9	19:00 : 32	0 : 73	50 : 43	0 : 133	0 : 86	0 : 115	0 : 103	0 : 99	0 : 131
Edmilson : 4	3257 : 1	Masculino : 55	2024-06-26 : 7	18:00 : 29	25 : 39	25 : 37	25 : 16	50 : 31	25 : 20	25 : 21	50 : 24	25 : 23
Amanda cs12 : 4	6513 : 1	Outros : 9	2024-07-02 : 7	15:00 : 20	50 : 22	0 : 21	11 : 14	25 : 27	11 : 14	50 : 19	25 : 21	11 : 14
Bruna : 3	9769 : 1		2024-07-05 : 7	17:00 : 15	20 : 19	100 : 20	50 : 11	20 : 15	20 : 14	11 : 14	11 : 14	50 : 8
Erica : 3	13025 : 1		2024-04-10 : 6	10:00 : 12	12 : 11	20 : 19	10 : 9	11 : 14	50 : 11	20 : 10	10 : 10	33 : 7
Sagaz : 3	16281 : 1		2024-04-06 : 5	13:00 : 12	100 : 8	11 : 13	33 : 8	33 : 9	33 : 7	33 : 9	20 : 9	10 : 6
Adriana Capela : 3	19537 : 1		2024-06-20 : 5	20:00 : 9	33 : 6	33 : 8	20 : 5	10 : 6	100 : 7	10 : 7	33 : 8	20 : 6
Daniela Reis : 3	22793 : 1		2024-06-27 : 5	09:00 : 9	10 : 6	10 : 7	15 : 3	30 : 5	10 : 5	100 : 4	100 : 7	55 : 3
Maria Lenes : 3	26049 : 1		2024-02-01 : 4	16:00 : 9	40 : 4	40 : 6	22 : 3	100 : 4	17 : 4	30 : 4	30 : 6	40 : 3
Gladiene : 3	28905 : 1		2024-03-03 : 4	14:00 : 8	11 : 4	30 : 6	16 : 3	16 : 3	16 : 3	40 : 3	16 : 4	28 : 2
Kelly : 3	32561 : 1		2024-04-13 : 4	08:00 : 8	22 : 4	60 : 4	100 : 2	22 : 3	55 : 2	17 : 3	17 : 3	17 : 2
Lari : 3	35817 : 1		2024-04-14 : 4	18:30 : 7	17 : 3	70 : 3	30 : 2	40 : 3	30 : 2	16 : 3	22 : 3	16 : 2
Andreia Moema : 3	39073 : 1		2024-05-05 : 4	14:30 : 6	30 : 3	17 : 3	55 : 2	5 : 3	22 : 2	22 : 3	28 : 2	22 : 2
Karine : 2	42329 : 1		2024-05-24 : 4	12:00 : 6	34 : 3	150 : 3	5 : 2	55 : 2	12 : 2	15 : 3	55 : 2	12 : 2
Castelo Branco : 2	45585 : 1		2024-06-21 : 4	07:30 : 5	60 : 2	55 : 3	28 : 1	15 : 2	40 : 2	55 : 2	40 : 2	18 : 2
Pedro Sanchez : 2	48841 : 1		2024-06-29 : 4	11:00 : 5	16 : 2	22 : 3	17 : 1	28 : 1	5 : 2	44 : 2	60 : 1	5 : 2
José : 2	52097 : 1		2024-01-01 : 3	19:30 : 5	28 : 1	75 : 2	67 : 1	60 : 1	28 : 1	60 : 2	75 : 1	67 : 1
Raiane : 2	55353 : 1		2024-02-17 : 3	17:30 : 4	29 : 1	67 : 2	43 : 1	17 : 1	29 : 1	5 : 2	67 : 1	43 : 1
Ana Paula : 2	58609 : 1		2024-02-27 : 3	16:30 : 3	150 : 1	15 : 2	14 : 1	62 : 1	62 : 1	28 : 1	78 : 1	100 : 1
Felipe Sallem : 2	61865 : 1		2024-03-26 : 3	15:30 : 3	66 : 1	12 : 2	40 : 1	43 : 1	7 : 1	70 : 1	44 : 1	14 : 1

Fonte: Elaborado pela autora.

6.4.1 Criação de um dashboard

No iFood, foram criados combos separados por quantidades, como o “Combo para 1 ou 2 Pessoas”, composto por dois combos que permite ao cliente comprar a quantidade exata para consumo próprio. Entre os combos mais vendidos estão o “Combinado Não Sai Caro”, que inclui 25 salgados sortidos,1 suco e 10 mini churros, e o “Tudo pra Mim”, que oferece 25 salgados de sal sortidos.

A análise de resultados foi estruturada com base nas diversas fontes de dados coletadas, que incluem: dados de seguidores e interações do Instagram, dados de vendas do iFood, dados de vendas diretas da loja (sem utilização de canais intermediários) e feedback dos clientes, obtidos por meio de comentários no Instagram, no portal do iFood, em pesquisas no Googleads (quando as pessoas buscam o nome da loja) e feedbacks recebidos diretamente pelo WhatsApp. Para isso, utilizamos os programas KNIME e Power BI, conhecidos por oferecerem uma interface gráfica interativa que facilita a exibição e criação de dashboards.

Na parte de análise dos Dados do Instagram feita no Power BI, foram analisados diversos aspectos dos dados do Instagram. A quantidade de seguidores foi segmentada por gênero (feminino e masculino) e também foram identificados perfis conhecidos como “unowned”, que geralmente são perfis profissionais ou de organizações. Além disso, analisamos as regiões de origem dos seguidores. Esta informação é crucial para a criação de promoções geograficamente focadas. Por exemplo, podemos oferecer entrega grátis em um raio de quilometragem predefinido, o que pode aumentar a atratividade para seguidores de determinadas regiões.

A análise dos dados de vendas realizadas através do iFood foram cruzados com as vendas diretas da loja. Isso nos permitiu analisar, por período, data e hora, quais produtos são mais consumidos e identificar tendências de consumo dos usuários. Essa análise revelou diferenças significativas no comportamento de compra entre os canais. No

iFood, os cinco salgados mais vendidos foram, em ordem decrescente: Coxinha, Empada de Frango, Bolinha de Queijo, Travesseirinho de Calabresa e Kibe. Em contraste, nas vendas diretas, a ordem dos salgados mais vendidos variou um pouco. Coxinha continuou sendo o mais vendido, seguido por Bolinha de Queijo, Empada de Frango e Mini Churros. A venda superior de Mini Churros nas transações diretas pode ser atribuída ao fato de que, no iFood, ele é categorizado como sobremesa, sendo vendido a um preço mais alto do que os salgados tradicionais.

Em termos de produtos com menor número de vendas, tanto no canal direto quanto pelo iFood, Risole de Milho e Empada de Alho-Poró ocupam as penúltimas posições. Esta informação é vital para ajustar o estoque e planejar futuras promoções ou descontinuar produtos de baixa rotatividade.

Na parte de análise das avaliações dos Feedbacks que foram colhidos durante o período de 1º de janeiro de 2024 a 30 de junho de 2024, coletamos e analisamos feedbacks recebidos de várias fontes: Instagram, portal do iFood, Googlads e WhatsApp. Para gerenciar e interpretar esses dados de maneira eficiente, utilizamos o KNIME para criar um workflow de tratamento de texto. Este workflow incluiu a segmentação das palavras, classificação semântica, remoção de pontuação e espaços em branco e o cálculo do peso de cada palavra, permitindo identificar a frequência e a raridade com que cada termo apareceu nos textos analisados.

Esta análise semântica nos ajudou a determinar se os comentários eram predominantemente positivos ou negativos. Além disso, construímos um dicionário de palavras associadas a avaliações de produtos, como “gostoso”, “queimado”, “cru”, “delicioso”, “amassado”, “bonito”, etc. Estas palavras foram previamente categorizadas como positivas ou negativas, facilitando a classificação automática dos sentimentos expressos nos comentários.

A implementação do nó OpenNLP NE Tagger no KNIME, foi utilizada, pois realiza a análise dos textos para identificar entidades como nomes de pessoas, cidades ou organizações. Este processo é essencial para compreender melhor o contexto dos feedbacks e identificar menções específicas a produtos, serviços ou locais, proporcionando uma visão mais detalhada e precisa das opiniões dos clientes.

6.5 Implementação do SquareSys

O SquareSys foi criado utilizando Streamlit¹ em conjunto com MySQL Workbench², que foi utilizado como banco de dados para salvar as informações preenchidas no sistema. O sistema conta com um menu autoexplicativo e de fácil interação para o usuário, que, mesmo com baixos conhecimentos técnicos, pode manusear a ferramenta para os fins propostos (veja Figura 19). A finalidade principal do sistema é registrar pedidos e cadastrar

¹ Framework Streamlit: <https://streamlit.io/>

² MySQL Workbench: <https://www.mysql.com/products/workbench/>

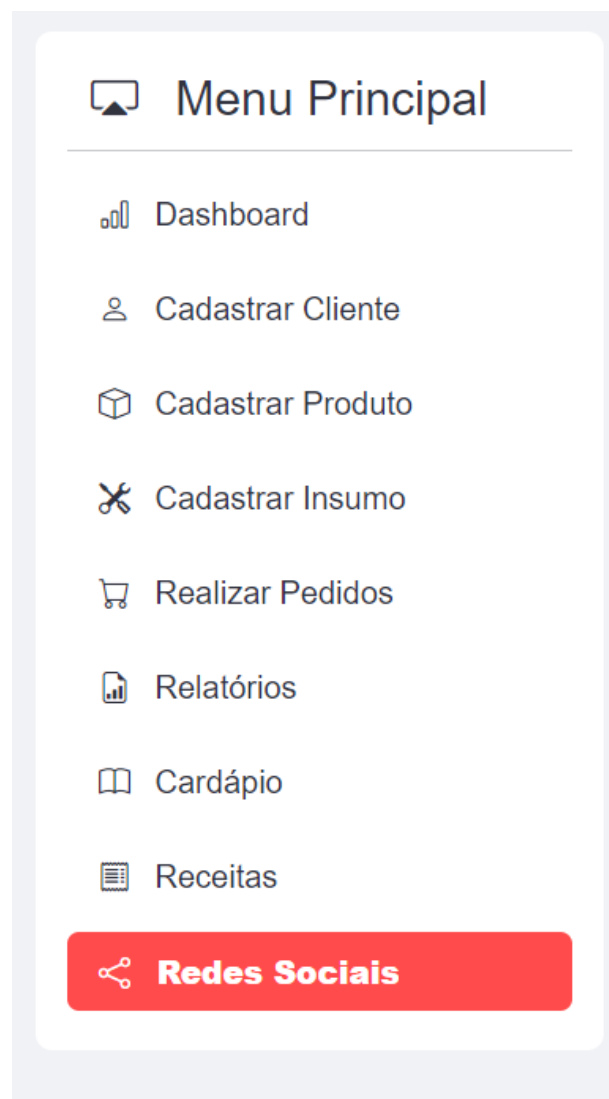


Figura 19 – Menu principal do site que serve como porta de entrada para a navegação do usuário escolher a tarefa que deseja executar.

Fonte: Elaborado pela autora.

insumos e produtos no estoque. As telas do sistema foram desenhadas de maneira sucinta, armazenando as informações principais necessárias para cada função a ser executada.

O banco de dados foi projetado para espelhar fielmente os campos necessários para a gravação das informações, garantindo uma organização clara e eficiente. Ele é local, ou seja, fica armazenado na própria máquina onde o sistema está instalado, proporcionando maior controle sobre os dados. O sistema, desenvolvido em Python utilizando o PyCharm, integra perfeitamente as funções de cadastro e gestão com o banco de dados, oferecendo uma solução robusta e confiável para a administração do estoque e pedidos da loja.

Além disso, o sistema foi criado com escalabilidade em mente, permitindo futuras expansões e a adição de novos módulos conforme as necessidades do negócio evoluem. Com

o banco de dados estruturado de forma relacional, as informações podem ser facilmente acessadas e manipuladas, o que garante agilidade nos processos de consulta e atualização de dados. Futuramente, quando o programa tiver todos os seus objetivos propostos totalmente finalizados, é estudado hospedá-lo na nuvem para ser acessado de qualquer lugar.

Abaixo seguem as telas do sistema e as linhas de códigos utilizadas para o desenvolvimento dos mesmos: A Figura 20 mostra os gráficos registrados; a Figura 21 mostra a tela de cadastro de clientes; a Figura 22 mostra a tela de cadastro de produtos; a Figura 23 mostra o cadastro de insumos; a Figura 24 mostra o cadastro de pedidos; a Figura 25 mostra os relatórios disponíveis; a Figura 26 mostra o cardápio disponível; e a Figura 27 mostra as redes sociais do negócio.

Figura 20 – Dashboard contendo gráficos dos dados registrados pela aplicação.



Fonte: Elaborado pela autora.

A parte de RECEITAS não foi finalizada, mas irá conter todos os dados de insumos utilizados para fabricação dos salgados, cruzando com as informações de estoque para controle de quantidade de cada ingrediente a ser utilizado na fabricação dos produtos. Na parte de SQL foi feito no workbench a engenharia reversa do código para geração do MER facilitando a compreensão de forma visual a interação das tabelas. A Figura 28 detalha a modelagem dessa parte do projeto.

6.6 Considerações Finais

A maioria dos workflows criados para gerar insights e desenvolver estratégias de marketing a partir dos resultados obtidos utilizando as métricas derivadas do modelo Apriori e técnicas de mineração de texto foram construídos com o auxílio de modelos

Figura 21 – Página para preenchimento e salvamento dos dados de cadastro de cliente

Deploy

Cadastrar Cliente

Nome Completo

Endereço

CEP

Telefone

Gênero

Feminino

Salvar

Fonte: Elaborado pela autora.

Figura 22 – Página para preenchimento e salvamento dos dados de cadastro de produto

Deploy

Cadastrar Produto

Tipo Produto

Salgado

Nome Produto

Bolinha de Queijo

Quantidade do Produto

Valor

Salvar

Fonte: Elaborado pela autora.

Figura 23 – Página para preenchimento e salvamento dos dados de cadastro de insumo

Deploy

Cadastrar Insumo

Nome Insumo

Data da Compra

2024/08/13

Data Vencimento

2024/08/13

Quantidade

Valor Unidade

Medida

Kg

Salvar

Fonte: Elaborado pela autora.

Figura 24 – Página para preenchimento e salvamento dos dados de Pedidos realizados.

Fonte: Elaborado pela autora.

Figura 25 – Página de exibição dos Relatórios de Clientes, Produtos Insumos e Pedidos cadastrados no sistema.

Fonte: Elaborado pela autora.

semelhantes aos apresentados no livro *From Words to Wisdom - Intro to Text Mining with KNIME*, escrito por Vincenzo Tursi e Rosaria Silipo (Tursi; Silipo, 2010). Este livro pode ser obtido no próprio site da empresa Knime, e é muito útil mesmo para quem não domina as técnicas de programação e mineração, pois ensina passo a passo, de forma detalhada, como utilizar a ferramenta, explicando o funcionamento de cada node e os tipos de resultados que eles podem gerar para o usuário.

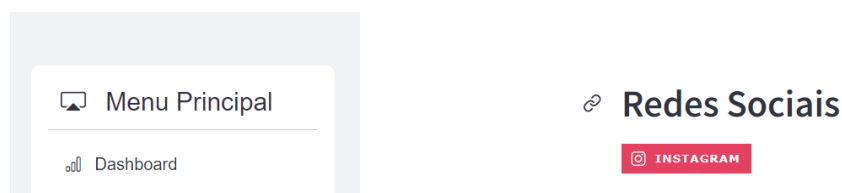
A construção de workflows eficientes e a obtenção de bons resultados em termos de insights exigem não apenas o conhecimento das ferramentas, mas também a habilidade de tratar e preparar os dados de forma adequada. Saber quais informações são realmente relevantes para o negócio e como analisá-las é essencial para que as estratégias geradas a partir desses dados sejam eficazes e impactem positivamente os resultados da empresa.

Figura 26 – Cardápio contendo as opções de menu da loja.



Fonte: Elaborado pela autora.

Figura 27 – Redes sociais da loja Salgados do Quadrado



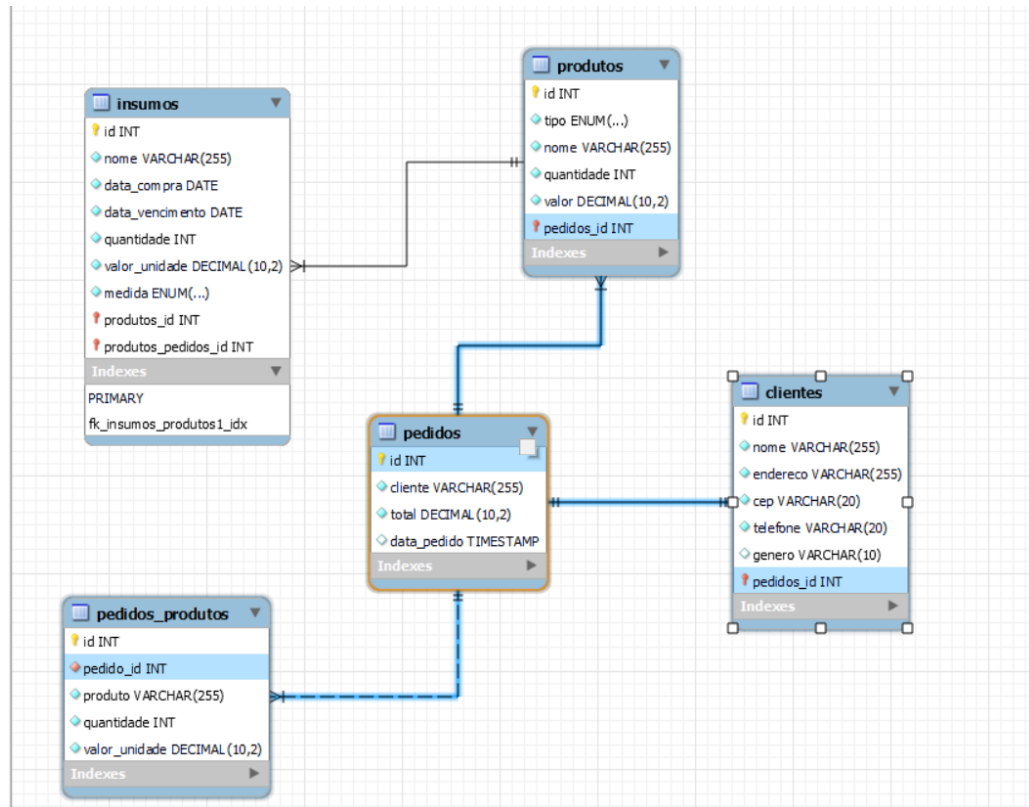
Fonte: Elaborado pela autora.

Os resultados obtidos com a criação e implementação do sistema SquareSys foram extremamente promissores, evidenciando que a combinação de inteligência artificial e big data pode transformar significativamente a gestão de pequenas e médias empresas, como a loja “Salgados do Quadrado”. Embora seja nova no mercado, a loja demonstrou que, com a aplicação correta de técnicas de análise de dados, é possível alcançar ganhos significativos na forma como se gerenciam as operações e se atende aos clientes.

O sistema desenvolvido não apenas aumentou a eficiência operacional da loja, como também estabeleceu uma base sólida para futuras expansões e adaptações, à medida que o mercado e as preferências dos clientes evoluem. A flexibilidade do SquareSys permite que ele seja incrementado conforme as necessidades futuras possam surgir no negócio, garantindo que a loja continue a crescer e se adaptar às demandas do mercado.

Em suma, este trabalho demonstrou que o uso de tecnologia de ponta, como inteligência artificial e big data, pode ser uma ferramenta poderosa para pequenas empresas, permitindo-lhes competir de maneira mais eficaz em um mercado cada vez mais competitivo. As metodologias aplicadas neste projeto não apenas solucionaram os desafios que eram

Figura 28 – MER



Fonte: Elaborado pela autora.

enfrentados pela loja, como, por exemplo, substituindo o registro de dados de forma manual para digital, como também a solução de outros problemas, que irão servir como base para futuras pesquisas e o desenvolvimento de sistemas ainda mais robustos, oferecendo soluções personalizadas e altamente eficientes para empresas de diversos setores. Este projeto destaca a importância de integrar a inovação tecnológica na gestão de negócios, tornando a análise de dados uma ferramenta essencial para apoiar as decisões estratégicas das organizações. A utilização eficaz dessas tecnologias não apenas impulsiona o crescimento e resolve problemas imediatos, mas também gera retorno financeiro e otimiza o tempo, contribuindo para a criação de valor a longo prazo.

7 CONCLUSÕES

Neste capítulo são apresentadas as conclusões deste trabalho de conclusão de curso, resumindo as principais descobertas obtidas através da análise do negócio “Salgados do Quadrado”. A pesquisa permitiu identificar diferentes fatores que podem afetar o desempenho do negócio e as estratégias mais eficazes para superá-los. A integração das soluções propostas com as práticas da empresa resultou em impactos significativos, como um aumento nos lucros do negócio e na geração de eficiência dos processos. Esses resultados evidenciam a importância de alinhar o negócio a técnicas de análise de dados para otimizar recursos, melhorar a competitividade e garantir o crescimento organizacional.

7.1 Contribuições

Este trabalho contribuiu significativamente para o entendimento, buscando a melhoria e otimização processos operacionais da loja “Salgados do Quadrado”, pois ajudou a desenvolver estratégias de mercado focadas no retorno financeiro para o negócio. Além de modernizar a forma de anotação de pedidos e criar uma base de dados robusta para o cadastro de clientes e insumos na loja. A implementação dos métodos de pesquisa, como a mineração de dados e a análise de padrões, resultou em um aumento notável na eficiência da coleta e análise de dados. A utilização de algoritmos como Apriori, somada a técnicas de mineração de texto e análise de sentimentos, aprimorou também a capacidade de identificar padrões de consumo e preferências dos clientes, possibilitando uma melhor tomada de decisões estratégicas e a personalização das ofertas de produtos.

7.2 Benefícios Práticos para o Negócio Analisado

As descobertas deste trabalho têm aplicações práticas diretas para a loja “Salgados do Quadrado”. A otimização dos processos de anotação e a criação de uma base de dados mais estruturada permitem uma gestão mais eficaz dos pedidos e um controle mais preciso do estoque. A capacidade de identificar padrões de compra e preferências dos clientes possibilita a criação de combos e ofertas mais alinhadas com as necessidades e desejos dos consumidores. Isso não só melhora a satisfação dos clientes, mas também pode aumentar a eficiência operacional e a rentabilidade da loja.

7.3 Trabalhos Futuros

Para aprimorar ainda mais os processos e expandir as capacidades analíticas, futuros trabalhos podem explorar a integração de técnicas avançadas de IA para prever tendências de demanda e otimizar a gestão de inventário. Além disso, a aplicação de métodos de

análise mais sofisticados para a segmentação de clientes e a personalização das ofertas pode proporcionar *insights* ainda mais valiosos. Investigar o impacto da implementação dessas técnicas em diferentes contextos de negócios também pode oferecer oportunidades para generalizar e aplicar os métodos desenvolvidos a outras áreas.

7.4 Considerações Finais

O trabalho realizado teve um impacto positivo significativo na loja “Salgados do Quadrado”, melhorando a eficiência operacional e a capacidade de análise de dados. A implementação dos métodos de pesquisa não apenas aprimorou a coleta e análise de dados, mas também ofereceu uma base sólida para decisões estratégicas mais assertivas. As descobertas contribuem para o campo de estudo ao demonstrar como aplicação de técnicas de análise de dados podem ser eficientes para resolver problemas específicos e melhorar processos de negócios. A relevância e o impacto dessas descobertas destacam a importância da análise de dados e da otimização de processos na gestão de negócios, desde empresas de pequeno porte como a loja de salgado, analisada, até grandes organizações.

REFERÊNCIAS

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Mining association rules between sets of items in large databases. *In*: BUNEMAN, P.; JAJODIA, S. (ed.). **Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, May 26-28, 1993**. ACM Press, 1993. p. 207–216. Disponível em: <https://doi.org/10.1145/170035.170072>.
- BERRAR, D. Cross-validation. *In*: RANGANATHAN, S. *et al.* (ed.). **Encyclopedia of Bioinformatics and Computational Biology - Volume 1**. Elsevier, 2019. p. 542–545. Disponível em: <https://doi.org/10.1016/b978-0-12-809633-8.20349-x>.
- COSTA, A. C. R. D.; VELLOSO, L. M. R. Interfaces digitais de serviços de entrega de alimentos - análise semiótica dos aplicativos da plataforma ifood. **InfoDesign : Brazilian journal of information design**, Brazilian Information Design Society (SBDI), v. 18, n. 1, 2021. ISSN 1808-5377.
- DEMSAR, J.; ZUPAN, B. Orange: Data mining fruitful and fun - A historical perspective. **Informatica (Slovenia)**, v. 37, n. 1, p. 55–60, 2013. Disponível em: <http://www.informatica.si/index.php/informatica/article/view/434>.
- ESPINOZA, P. *et al.* Metodología basada en grafos para la identificación de perfiles de usuario. **Res. Comput. Sci.**, v. 97, p. 127–139, 2015.
- GINO, H. *et al.* Exploratory analysis on market basket data using network visualization. *In*: **Anais do XII Brazilian Workshop on Social Network Analysis and Mining**. Porto Alegre, RS, Brasil: SBC, 2023. p. 19–30. ISSN 2595-6094.
- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques, 3rd edition**. Morgan Kaufmann, 2011. ISBN 978-0123814791. Disponível em: <http://hanj.cs.illinois.edu/bk3/>.
- IHRMARK, D.; TYRKKÖ, J. Learning text analytics without coding? an introduction to KNIME. **Educ. Inf.**, v. 39, n. 2, p. 121–137, 2023. Disponível em: <https://doi.org/10.3233/EFI-230027>.
- MITCHELL, T. M. **Machine learning, International Edition**. [S.l.: s.n.]: McGraw-Hill, 1997. (McGraw-Hill Series in Computer Science).
- OLIVEIRA, F. A. de *et al.* Mining multirelation association rules on the web of data. **iSys - Brazilian Journal of Information Systems**, v. 13, n. 4, p. 77–100, Jul. 2020. Disponível em: <https://sol.sbc.org.br/journals/index.php/isys/article/view/830>.
- OLIVEIRA, F. de *et al.* Mineração de regras de associação multirrelação em grafos: Direcionando o processo de busca. *In*: **Anais do XXXII Simpósio Brasileiro de Bancos de Dados**. Porto Alegre, RS, Brasil: SBC, 2017. p. 270–275. ISSN 2763-8979.
- PEARSON, M. *et al.* Introduction to power bi. *In*: _____. **Pro Microsoft Power Platform: Solution Building for the Citizen Developer**. Berkeley, CA: Apress, 2020. p. 173–189. ISBN 978-1-4842-6008-1. Disponível em: https://doi.org/10.1007/978-1-4842-6008-1_16.

RIBEIRO, D.; PLASTINO, A.; MURTA, L. Ostra: uma abordagem para análise da qualidade de software por meio de regras de associação de métricas. *In: Anais do XII Simpósio Brasileiro de Qualidade de Software*. Porto Alegre, RS, Brasil: SBC, 2013. p. 351–365. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbqs/article/view/15299>.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach (4th Edition)**. Pearson, 2020. ISBN 9780134610993. Disponível em: <http://aima.cs.berkeley.edu/>.

SANTOS, M. M.; NASCIMENTO, D. C. do. Avaliando fatores de influência sobre algoritmos de aprendizado de máquina na etapa de classificação da resolução de entidades. *In: Proceedings of the 38th Brazilian Symposium on Databases, SBBD 2023, Belo Horizonte, MG, Brazil, September 25-29, 2023*. SBC, 2023. p. 63–75. Disponível em: <https://sol.sbc.org.br/index.php/sbbd/article/view/25517>.

TURSI, V.; SILIPO, R. **From Words to Wisdom: Introduction to Text Mining with KNIME**. [S.l.: s.n.]: KNIME Textbooks, 2010. ISBN 978-3-9523926-5-2.

WIEDERHOLD, G.; MCCARTHY, J. Arthur samuel: Pioneer in machine learning. **IBM J. Res. Dev.**, v. 36, n. 3, p. 329–332, 1992. Disponível em: <https://doi.org/10.1147/rd.363.0329>.