REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

AVALIAÇÃO DE DESEMPENHO DE ALGORITMOS DE MINERAÇÃO DE DADOS E SIMULAÇÃO DE MONTE CARLO NA DESCOBERTA DE TENDÊNCIAS NO APP *HAMBRE DELIVERY*

PERFORMANCE EVALUATION OF DATA MINING ALGORITHMS AND MONTE CARLO SIMULATIONS FOR TREND DISCOVERY IN THE HAMBRE DELIVERY APP

EVALUACIÓN DEL RENDIMIENTO DE ALGORITMOS DE MINERÍA DE DATOS Y SIMULACIONES DE MONTE CARLO PARA EL DESCUBRIMIENTO DE TENDENCIAS EN LA APLICACIÓN HAMBRE DELIVEY

Laciene Melo Garcia[1], Keventon Rian Guimarães Gonçalves[2], Joiner dos Santos Sá[3], Elton Rafael Alves[4], Jasmine Priscyla Leite de Araújo[5], Fabricio de Souza Farias[6]

e686669
https://doi.org/10.47820/recima21.v6i8.6669

PUBLICADO: 8/2025

RESUMO

O setor *food service* vivencia um crescimento contínuo, decorrente da valorização de compras *online* via *marketplace*, que agilizam as transações e contribuem para melhorar a qualidade dos produtos e serviços oferecidos. A digitalização da gestão comercial de empresas proporciona o conhecimento necessário para enfrentar crises de saúde e econômicas, tendo como suporte a incorporação de *apps* especializados. A adesão a ferramentas baseadas em Inteligência Artificial (IA) mudou a forma como os negócios operam, mas representa um desafio para empresas menores e mais jovens que ainda não oferecem serviços *online*. Motivado pela necessidade de sistematizar a análise de tendências de vendas provenientes das lojas parceiras do *app Hambre Delivery*, esta pesquisa propõe uma solução para avaliar o desempenho computacional, combinando o método Monte Carlo e algoritmos de mineração de dados para identificar o modelo mais adequado para suportar a decisão na gestão de vendas via aplicativo. A partir das simulações de Monte Carlo, foram analisados os algoritmos FP-Growth, FP-Max, Apriori e Eclat, considerando a escalabilidade, o tempo de execução e o uso de memória como critérios de desempenho. Os resultados revelaram que o algoritmo Eclat é mais indicado para conjuntos de dados pequenos e de baixa complexidade, enquanto FP-Growth e FP-Max são escaláveis para lidar com grandes volumes de dados, sendo mais eficientes quanto ao tempo de execução e ao uso de memória. Além disso, as 27 regras de associação geradas revelaram tendências relevantes, mostrando que a aplicação do Monte Carlo resulta em padrões mais precisos e confiáveis.

PALAVRAS-CHAVE: Mineração de Dados. Monte Carlo. Food Service. Descoberta de Conhecimento. Regras de Associação. Tendências de Venda.

ABSTRACT

The food service sector is experiencing continuous growth, driven by the rise of online shopping. Online shopping streamlines transactions and improves the quality of products and services offered.

[1] Estudante de Mestrado em Computação Aplicada pelo PPCA/UFPA.
[2] Bacharel em Sistemas de Informação (UFPA). Professor Substituto na Faculdade de Sistemas de Informação FASI/UFPA.
[3] Estudante de Doutorado em Engenharia Elétrica pelo PPGEE/UFPA.
[4] Doutor em Engenharia Elétrica (UFPA). Professor Adjunto na Faculdade de Engenharia da Computação FAEC/UNIFESSPA.
[5] Doutora em Engenharia Elétrica. Professora Adjunta do Instituto de Tecnologia ITEC/UFPA.
[6] Doutor em Engenharia Elétrica. Professor Adjunto na Faculdade de Sistemas de Informação FASI/UFPA.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Digitizing commercial management provides companies with the knowledge necessary to withstand health and economic crises, supported by specialized apps. While the adoption of AI-based tools has changed the way businesses operate, it poses a challenge for smaller, younger companies that do not yet offer online services. Motivated by the need to systematize the analysis of sales trends from partner stores of the Hambre Delivery app, this study proposes a computational performance evaluation solution that combines the Monte Carlo method and data mining algorithms to identify the most appropriate model for strategic sales management support. Through Monte Carlo simulations, the FP-Growth, FP-Max, Apriori, and Eclat algorithms were assessed in terms of scalability, execution time, and memory usage. The results showed that the Eclat algorithm is better suited to small, low-complexity data sets. FP-Growth and FP-Max, on the other hand, are scalable and can handle large volumes of data more efficiently in terms of execution time and memory usage. Additionally, the 27 generated association rules revealed relevant trends, showing that applying Monte Carlo results in more accurate and reliable patterns.

KEYWORDS: *Data Mining. Monte Carlo. Food Service. Knowledge Discovery. Association Rules. Sales Trends.*

RESUMEN
El sector del food service experimenta un crecimiento sostenido, impulsado por la valorización de las compras en línea a través de marketplaces, que agilizan las transacciones y mejoran la calidad de productos y servicios. La digitalización de la gestión comercial proporciona el conocimiento necesario para enfrentar crisis sanitarias y económicas, respaldada por la incorporación de aplicaciones especializadas. La adopción de herramientas basadas en Inteligencia Artificial (IA) ha transformado la operación de los negocios, aunque representa un desafío para empresas pequeñas que aún no ofrecen servicios digitales. Motivada por la necesidad de sistematizar el análisis de tendencias de ventas de tiendas asociadas a la aplicación Hambre Delivery, esta investigación propone una solución para evaluar el rendimiento computacional, combinando el método de Monte Carlo con algoritmos de minería de datos. El objetivo es identificar el modelo más adecuado para respaldar la toma de decisiones en la gestión de ventas a través de la aplicación. Mediante simulaciones de Monte Carlo, se analizaron los algoritmos FP-Growth, FP-Max, Apriori y Eclat, considerando la escalabilidad, el tiempo de ejecución y el uso de memoria. Los resultados muestran que Eclat es más adecuado para conjuntos de datos pequeños y simples, mientras que FP-Growth y FP-Max se destacan por su eficiencia y escalabilidad frente a grandes volúmenes de datos. Además, las 27 reglas de asociación generadas revelaron tendencias relevantes, confirmando que el uso de Monte Carlo permite identificar patrones más precisos y confiables.

PALABRAS CLAVE*: Minería de Datos. Monte Carlo. Servicio de Alimentación. Descubrimiento de Conocimiento. Reglas de Asociación. Tendencias de Ventas.*

INTRODUÇÃO

A digitalização do setor *food service* (Serviço de Alimentação) é uma realidade para estabelecimentos comerciais de pequeno, médio e grande porte, principalmente devido à incorporação de *apps* especializados ao cotidiano das pessoas, que geram oportunidades e desafios. Essa realidade decorre da valorização de compras via *marketplace*, que agilizou os processos comerciais, facilitando o controle de custos e contribuindo para melhorar a experiência dos clientes (Bujalance-López *et al.,* 2025).

O mercado está mais exigente, e estabelecimentos que utilizam ferramentas digitais para estruturar o seu processo de gestão comercial apresentam o conhecimento necessário para

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

enfrentar crises de saúde e econômicas, o que lhes permite reagir mais rápido do que empresas menos digitalizadas (Grijalba *et al.,* 2024; Campos; Farina; Florian, 2022). A transformação digital, impulsionada pela busca por soluções baseadas em Inteligência Artificial (IA), mudou a forma como os negócios operam, sendo inviável para empresas consolidadas gerir seus setores sem controle de estoque, painéis de vendas e canais digitais (Rabhi; Beheshti; Gill, 2025).

Além disso, a IA tem sido aplicada para auxiliar lojas e equipes de *marketing* na identificação dos fatores que afetam as vendas e, principalmente, na elaboração de estratégias de *marketing* mais efetivas. Sob esse aspecto, foram encontradas na literatura pesquisas voltadas para a análise de transações em lojas de varejo, franquias e supermercados, o que indica um claro interesse na compreensão de padrões de consumo.

Os autores Pradana *et al.,* (2022) realizaram um estudo de caso para identificar as possíveis causas que levaram à queda nas vendas da franquia de vestuário *Hoyjakarta*. A solução propôs o uso do algoritmo FP-Growth como abordagem de apoio à descoberta de padrões relevantes a partir da análise de 571 registros de transações disponibilizados pela franquia. Vale ressaltar que os resultados foram promissores, fornecendo um panorama geral sobre a confiabilidade nas vendas de roupas, tanto em casos de vendas isoladas quanto naquelas que impulsionaram a aquisição de outros produtos, mesmo para vestuários que não despertaram tanto interesse dos clientes.

Josephine e Rajan (2023) investigaram o baixo volume de vendas de uma loja de departamento a fim de identificar as preferências dos clientes e melhorar o direcionamento de promoções. Os autores enfatizaram que a escolha do instrumento de análise levou em consideração três algoritmos de mineração de dados, sendo eles Apriori, FP-Growth e Eclat. Essa escolha favoreceu o algoritmo Apriori por ser amplamente utilizado em pesquisas, enquanto os quesitos observados para análise tomaram como base 16.753 transações de venda. Após o procedimento, foram obtidas 20 regras de associação envolvendo os produtos mais procurados em toda a loja, o que favoreceu a equipe de *marketing* na elaboração de campanhas mais direcionadas.

A pesquisa de Bhagampriyal *et al.,* (2023) propôs a elaboração de um modelo híbrido de recomendações de produtos para supermercados, baseado em dois aspectos essenciais, as avaliações recebidas pelos produtos e as transações de vendas realizadas. Em termos práticos, ambas as informações vieram de uma única base de dados proveniente de um supermercado, que disponibilizou no *Kaggle* 38.765 transações, cada uma contendo o produto vendido associado a uma escala numérica de avaliação variando entre 1 e 5. O resultado e o diagnóstico foram obtidos por meio da análise de três algoritmos, sendo eles o PPR, que ranqueia os produtos mais bem avaliados, e os modelos de mineração de dados Apriori e FP-Growth, usados para identificar os padrões frequentes de compra. Após os testes, os autores concluíram que, devido à eficiência de execução, a utilização conjunta dos algoritmos PPR e FP-Growth é a alternativa mais vantajosa como modelo híbrido.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Embora existam trabalhos com aplicação prática de *software* e IA na área de *food service*, ainda existem desafios técnicos, pois as soluções já desenvolvidas e apresentadas na literatura ainda apresentam lacunas e não são capazes de atender de forma eficaz o volume de clientes e vendas, em relação ao tempo de resposta, qualidade das informações, velocidade de processamento, entre outros. (Liu et al., 2024). Deste modo, sendo necessário investigar e comparar algoritmos baseados em IA que possam servir a esse propósito, respeitando parâmetros de desempenho de *hardware* e *software*, assim como retornando ao usuário informações precisas e no menor tempo e custo possíveis.

Nesse contexto, surgiu o Projeto de Pesquisa e Desenvolvimento (P&D) da plataforma de compra e vendas *online Hambre Delivery* (Gonçalves, 2024), visando à inclusão digital de empreendedores da região do Baixo Tocantins, que enfrentaram a redução parcial ou total de suas vendas durante a pandemia de Covid-19. O app *Hambre Delivery* foi idealizado e implantado em cinco cidades do estado do Pará, sendo elas Cametá, Baião, Mocajuba, Oeiras do Pará e Limoeiro do Ajuru. Vale ressaltar que havia captadores responsáveis por credenciar os estabelecimentos do setor de *food service* em cada cidade.

Diante disso, este estudo apresenta uma das soluções desenvolvidas para o *projeto Hambre Delivery*. A solução proposta consiste em uma técnica de IA híbrida que combina o método de Monte Carlo e algoritmos de Mineração de Associação para analisar o desempenho computacional e oferecer resultados mais consistentes, validando a escolha do modelo mais adequado para a tomada de decisão na gestão de vendas do *Hambre Delivery*.

Especificamente, buscou-se implementar os algoritmos de associação Apriori, FP-Growth, FP-Max e Eclat voltados para o ambiente computacional de pesquisa. Definir os cenários e configurações para os testes de Monte Carlo, considerando simulações preliminares e de sensibilidade. Avaliar o desempenho computacional quanto à escalabilidade, tempo de execução e uso de memória. Determinar se o uso do método de Monte Carlo proporciona maior estabilidade no desempenho computacional, para apoiar a identificação do algoritmo mais eficiente. Por fim, analisar as principais tendências de venda do *Hambre Delivery*.

A solução destaca-se pela importância teórica especialmente para as áreas de computação aplicada, pois a combinação do Monte Carlo com algoritmos de mineração de associação representa uma abordagem inovadora para aumentar a precisão da avaliação de desempenho computacional, além de contribuir para elaboração de pesquisas similares.

As demais seções do trabalho estão organizadas da seguinte forma. A Seção 1 contextualiza as principais abordagens de mineração de dados. A Seção 2 descreve a metodologia proposta. Na Seção 3, detalha o caso de estudo de aplicação dos modelos de mineração. Na Seção 4, são discutidos os experimentos e resultados da análise de desempenho computacional. Por fim, as conclusões são apresentadas na Seção 5.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

1. ABORDAGENS DE MINERAÇÃO DE DADOS

Esta Seção contextualiza os principais aportes teóricos encontrados na literatura e adotados na implementação da solução em mineração de dados proposta nesta pesquisa.

1.1. Métricas Aplicadas na Mineração de Dados

A validação dos padrões extraídos por técnicas de mineração de associação requer o uso de métricas para determinar sua frequência e força. Por isso, em uma análise contínua, são aplicadas métricas de avaliação de suporte, confiança e *lift* (Lin; Tseng; Su, 2002). Quando bem configuradas, conseguem avaliar itens frequentes e regras de associação, contribuindo de forma efetiva para a análise e interpretação de resultados.

Seguindo essa abordagem, uma regra de associação é candidata a item frequente quando satisfaz a um suporte mínimo (sup_{min}). Por isso, a escolha de um suporte é responsável por definir a porcentagem de itens que serão incluídos no conjunto de itens frequentes. Logo, a condição $A \Rightarrow B$ está relacionada à probabilidade de $P(A \cap B)$, que representa a frequência de ocorrência dos itens em um conjunto. A equação (1) expressa o cálculo de suporte:

$$suporte(A \Rightarrow B) = P(A \cap B) = \frac{|A \cap B|}{N} \quad (1)$$

O suporte é calculado com base na quantidade de transações que incluem A e B, sendo o módulo da interseção $|A \cap B|$ desses itens dividido pelo total de transações N que formam o conjunto de dados.

Por outro lado, a métrica de confiança mede a precisão de uma regra de associação, considerando a probabilidade de B ocorrer em uma transação, dado que A ocorreu. Matematicamente, isso é expresso pela probabilidade condicional $P(B|A)$, conforme ilustra a equação (2).

$$confiança(A \Rightarrow B) = P(B|A) = \frac{suporte(A \cup B)}{suporte(A)} \quad (2)$$

Dada a equação, a confiança é calculada por meio da frequência da união $A \cup B$, sendo o número de vezes que esses itens são encontrados juntos, dividido pelo número de vezes que A aparece nas transações. A maneira como essa métrica atua permite determinar a precisão de uma regra de associação, porém sua forma de abordagem é limitada, deixando uma lacuna quanto à real força de uma regra de associação.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Nesse sentido, a métrica de *lift* surge como uma medida simples de correlação (Han; Kamber; Pei, 2012), cujo foco é avaliar a força de uma regra de associação, considerando o quanto os eventos de uma regra são dependentes e correlacionados. A equação (3) ilustra a métrica de *lift*.

$$lift(A,B) = \frac{P(A \cup B)}{P(\mathrm{A}) \times P(\mathrm{B})} \quad (3)$$

Logo, o valor de *lift* para uma associação é calculado como a razão entre a probabilidade do conjunto de itens *A* e *B* e o produto das probabilidades individuais desses itens. Um *lift* maior que 1 indica uma associação positiva, ou seja, a presença de um item eleva a probabilidade de presença do outro. Se o *lift* for igual a 1, sugere que os itens são independentes e não afetam a ocorrência um do outro. Caso o *lift* seja menor que 1, indica uma associação negativa, em que a presença de um item reduz a probabilidade de ocorrência do outro.

1.2. Algoritmos Apriori, FP-Growth e FP-Max

A mineração de dados voltada para a descoberta de associações apresenta modelos computacionais robustos, que se assemelham bastante na finalidade de análise, tanto no formato do conjunto de dados utilizado em procedimentos quanto na configuração inicial dos parâmetros de suporte e confiança (Zaki, 2000).

Sob essa perspectiva, os algoritmos Apriori e FP-Growth são amplamente utilizados para extrair itens frequentes e gerar regras de associação, especificamente em um vetor de transações, onde o processo de análise é orientado pela predefinição dos parâmetros de suporte e confiança (Han; Kamber; Pei, 2012). Por sua vez, o algoritmo FP-Max considera o mesmo formato de dados e um suporte para extrair itens frequentes máximos (Hu *et al.,* 2008).

Contudo, o diferencial entre eles está na forma de realizar o processo de mineração de dados, no qual o algoritmo Apriori realiza um processo iterativo de junção e poda de candidatos, baseado no princípio de que, se um conjunto de itens é frequente, todos os seus subconjuntos também devem ser frequentes. Assim, o conhecimento prévio do vetor de transações é essencial para determinar a relevância das associações (Agrawal e Srikant, 1994). Enquanto os algoritmos FP-Growth e FP-Max utilizam a estrutura de dados *Frequent Pattern Tree* (FP-Tree), projetada para armazenar conjuntos esparsos de forma comprimida, otimizando a mineração de itens frequentes e regras de associação (Grahne; Zhu, 2005).

Deste modo, a inicialização adequada do conjunto de transações e a definição das métricas de mineração são essenciais para a descoberta de associações.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

1.3. Algoritmo Eclat

O algoritmo Eclat consiste em uma técnica de mineração de itens frequentes que utiliza a interseção de listas de transações como base para a análise (Zaki *et al.,* 1997). Nesse contexto, o formato vertical de dados permite uma busca em profundidade mais eficiente, especialmente quando o conjunto de dados apresenta baixa cardinalidade de itens por transação.

Em termos conceituais, o formato de dados vertical pode ser representado como um conjunto de pares (i, T_n), sendo i um item e T_n o conjunto de transações que incluem i. Então, se houver dois itens A e B, por exemplo, suas representações seriam $A = \{T_1, T_2\}$ e $B = \{T_1, T_2, T_3\}$. A partir da definição do conjunto de dados no formato adequado e da escolha de um sup_{min}, o processo de extração de itens frequentes pode ser realizado.

1.4. Método de Monte Carlo

No contexto de simulações, o Método de Monte Carlo (MMC) foi proposto como uma abordagem estatística para a resolução de problemas que envolvem incerteza e variabilidade. Sua consistência é útil em ambientes computacionais complexos, que exigem explorar diversas características para a tomada de decisão. Assim, a realização de múltiplas varreduras combinada a métodos estatísticos robustos pode gerar estimativas confiáveis (Hastings, 1970).

No que se refere à aplicação do método, todo seu funcionamento é baseado na definição de cenários que simulam a condição real do ambiente de investigação. Essa abordagem precisa estar inserida em uma perspectiva que possibilite a aplicação de cenários variados, compostos por diferentes números de iterações. Tais aspectos favorecem uma avaliação mais ampla em termos de comparação, resultando em estimativas mais precisas e consistentes.

Em meio às simulações, são calculadas estatísticas que dão suporte à análise dos resultados, como média, desvio padrão e margem de erro. Cabe à média proporcionar uma estimativa equilibrada do resultado obtido, enquanto o desvio padrão avalia a dispersão dos resultados em relação a ela. Em uma compreensão mais ampla, a margem de erro representa a faixa em torno da estimativa na qual se espera que o valor real esteja.

Assim, a capacidade do MMC de incorporar esses recursos de avaliação determina sua posição como ferramenta poderosa de simulação e análise de sistemas complexos.

2. MÉTODOS

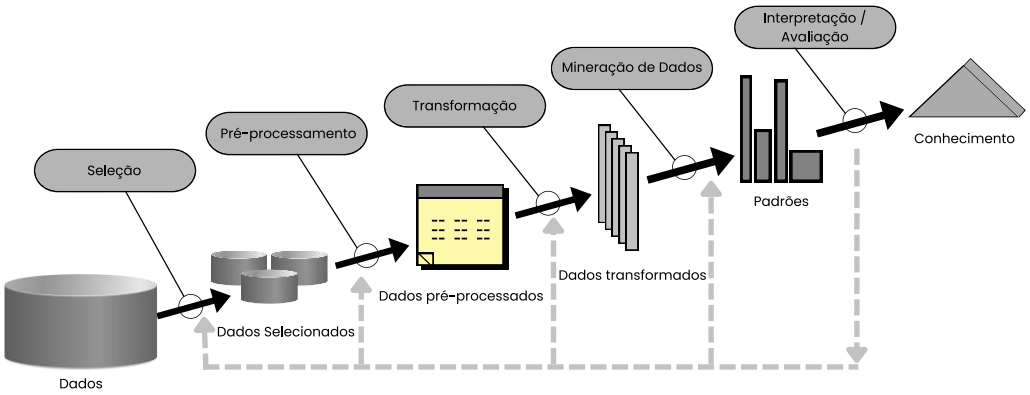
Quanto à sua natureza, este estudo alinha-se ao escopo da pesquisa aplicada de cunho exploratório, por ser dirigido à compreensão de um problema real, em que a investigação centrada na mineração de dados serve de estímulo à descoberta de conhecimento em bases de dados.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Considerando que o foco desta investigação é a análise de desempenho computacional, combinando o MMC e algoritmos de mineração de dados, foram adotadas neste trabalho as etapas metodológicas propostas em KDD (*Knowledge Discovery in Databases*), por Fayyad, Piatetsky-Shapiro e Smyth (1996), que incluem a coleta, pré-processamento, transformação, mineração e interpretação dos dados, que agregam robustez aos métodos de análise e possibilitam a extração de informações relevantes. A Figura 1 ilustra o fluxo de análise em KDD.

Figura 1. Fluxo das etapas no processo de KDD



Fonte: Fayyad, Piatetsky-Shapiro, Smyth (1996)

2.1. Coleta de Dados

A coleta de dados foi realizada em um período condizente com o lançamento do aplicativo *Hambre Delivery* até se estabelecer como *marketplace* de *food service*. Este processo se deu especificamente entre 02/2022 e 03/2024, totalizando 24 meses de amostras transacionais que representam pedidos concluídos pelas lojas.

Assim, foram coletados 2.191 registros de vendas concluídas via aplicativo. Cada registro é composto apenas pelos atributos *id* e *products*, sendo que *products* pode conter um ou mais itens comprados, o que torna esse fator uma variável suscetível à vontade do usuário no momento da compra. A Figura 2 ilustra o formato de armazenamento dos dados.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Figura 2. Amostragem do Formato de Armazenamento de Dados Brutos

id	Products
1	Caldo carne com ovo, Caldo de frango, Cola cola Lata
2	Batata frita com bacon e cheddar
3	Pizza Media (M) Mussarela
4	Marmitex frango empanado, Marmitex bife de carne
5	Açaí de 1 litro, Farinha de Mandioca
...	...
2.191	Filé (F), Paraense (F), Nordestina (F), Refrigerante Coca-cola

Fonte: Dados da Pesquisa

2.2. Pré-Processamento

Após a coleta de dados, notou-se inicialmente uma heterogeneidade na descrição textual de alguns registros. Entretanto, em uma análise exploratória, foi possível notar que a divergência de formatos textuais era mais ampla e que registros ruidosos poderiam direcionar a pesquisa para resultados incertos.

Para fins de esclarecimento, as divergências incluíam emojis, caracteres especiais, espaços extras, números, medidas volumétricas, itens duplicados no mesmo registro e diferentes formas de escrever o mesmo item, a exemplo do item "Refrigerante Coca-Cola", que foi escrito como "Coca-Cola 2 litros", "Coca Original", "Coca cola 2L", "Coca cola Lata", "143 Coca de 1L", "144 Coca-Cola 2L", "Coca Cola", "Cola cola" ou "Refrigerante Coca-Cola 310 ml".

Nesse contexto, foram estabelecidas medidas de limpeza de dados que, por sua vez, possibilitaram padronizar, substituir e remover duplicatas em um mesmo registro. A Figura 3 demonstra o trecho de código em *Python* responsável pela padronização do produto *"Refrigerante Coca-Cola"*.

Figura 3. Código de Padronização da Escrita do Refrigerante Coca-Cola

```
(1)  # Lista de variações identificadas para "Refrigerante Coca-Cola"
(2)  coca_cola_variants = [
(3)      'Coca-Cola 2 litros', 'Coca Original', 'Coca cola 2L',
(4)      'Coca cola 600 mL', 'Coca cola Lata', '143 Coca de 1L',
(5)      '144 Coca-Cola 2L', 'Coca Cola', 'Cola cola', 'Refrigerante Coca-Cola 310 ml'
(6)  ]
(7)
(8)  # Função para substituir ocorrências de variações por um nome padronizado
(9)  def standardize_name(text, variations, standard_name):
(10)     for variant in variations:
(11)         text = text.replace(variant, standard_name)
(12)     return text
(13)
(14) # Aplicação da padronização na coluna "products"
(15) dataset['products'] = dataset['products'].apply(standardize_name, args=(coca_cola_variants, 'Refrigerante Coca-Cola'))
```

Fonte: Elaborado pelos autores

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

2.3. Transformação de Dados

Nesta etapa, foi considerado o formato de dados mais adequado para aplicar os modelos de mineração. A análise exploratória inicial revelou desafios relacionados à dimensionalidade e à necessidade de conversão de formato.

Embora a quantidade de registros na base de dados não seja extensa, observou-se muitos registros com apenas um item, o que pode reduzir a quantidade de itens frequentes e regras de associação geradas. Com base nisso, foi aplicada a técnica de redução de dimensionalidade para filtrar registros com mais de um item comprado. Esse procedimento resultou na obtenção de 622 registros únicos e consistentes. Após essa filtragem, os registros foram submetidos à técnica de conversão de formato, com o intuito de gerar duas estruturas de dados: horizontal e vertical.

Para obter a estrutura horizontal, os registros foram transformados em um vetor de transações, sendo o formato ideal para os algoritmos FP-Growth, FP-Max e Apriori.

Figura 4. Resultado da Conversão para o Formato Horizontal

```
(1)   # formato horizontal de dados
(2)   horizontal_format = [
(3)      {'Refrigerante Coca-Cola', 'Filé Parmegiana'},
(4)      {'Caldo de Frango', 'Caldo Carne com Ovo', 'Refrigerante Coca-Cola'},
(5)      {'Refrigerante Coca-Cola', 'Pizza Filé com Bacon'},
(6)      {'BIFE', 'Frango Empanado'},
(7)      {'Pizza Marguerita', 'Pizza Mista do Pará', 'Pizza Portuguesa'},
(8)      {'Pizza Metade Filé com bacon', 'Pizza Metade Calabresa'},
(9)      {'Pizza Metade Portuguesa', 'Refrigerante Guaraná', 'Pizza Metade Mussarela'},
(10)     {'Pizza Metade 4 queijos', 'Refrigerante Pepsi', 'Pizza Metade Calabresa'}
(11)     ...,
(12)     { 'Hambúrguer X-Tudo', 'Batata Frita Simples', 'Refrigerante Guaraná}
(13)  ]
```

Fonte: Dados da Pesquisa

Por outro lado, no formato vertical, os dados foram organizados como um dicionário de itens ou tabela *hash*, sendo este o formato aceito para a mineração de dados utilizando o algoritmo Eclat. A Figura 5 ilustra a saída da conversão para o formato vertical.

Figura 5. Resultado da Conversão para o Formato Vertical

```
(1)   # formato horizontal de dados
(2)   horizontal_format = [
(3)      {'Refrigerante Coca-Cola', 'Filé Parmegiana'},
(4)      {'Caldo de Frango', 'Caldo Carne com Ovo', 'Refrigerante Coca-Cola'},
(5)      {'Refrigerante Coca-Cola', 'Pizza Filé com Bacon'},
(6)      {'BIFE', 'Frango Empanado'},
(7)      {'Pizza Marguerita', 'Pizza Mista do Pará', 'Pizza Portuguesa'},
(8)      {'Pizza Metade Filé com bacon', 'Pizza Metade Calabresa'},
(9)      {'Pizza Metade Portuguesa', 'Refrigerante Guaraná', 'Pizza Metade Mussarela'},
(10)     {'Pizza Metade 4 queijos', 'Refrigerante Pepsi', 'Pizza Metade Calabresa'}
(11)     ...,
(12)     { 'Hambúrguer X-Tudo', 'Batata Frita Simples', 'Refrigerante Guaraná}
(13)  ]
```

Fonte: Dados da Pesquisa

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

2.4. Mineração de Dados

No que se refere à etapa de mineração de dados, foi adotada uma análise mais direcionada à avaliação do desempenho dos algoritmos, tendo como instrumento de simulação e observação o MMC. Assim, foram definidos critérios centrados em garantir a reprodutibilidade do método.

A análise de desempenho foi realizada por meio de três cenários de simulação MMC, que permitem verificar a confiabilidade e a precisão de critérios como escalabilidade, memória e tempo de execução. A Tabela 1 apresenta os parâmetros técnicos adotados para a aplicação do MMC.

Tabela 1. Parâmetros de Configuração do MMC

Parâmetro	Descrição	Valor
N_ITERATIONS	Número de interações por cenário MMC	100, 1.000, 10.000
BOOTSTRAP_SAMPLING	Particionamento aleatório de dados	Replace=True
RANDOM_STATE	Controle da aleatoriedade (*Seed*)	42
N_SAMPLE	Amostragem gerada por cenário MMC	622, 73.000
CARDINALIDADE	Itens por transação	3 a 7

Fonte: Elaborado pelos autores

Os cenários de simulação do MMC tomaram por base critérios adaptados de Heijungs (2020) e Tanaka *et al.,* (2021), que sugerem a realização de simulações entre 100 e 10.000 iterações. Considerando o foco desta investigação, o cenário de 100 iterações foi definido para confirmar o comportamento correto dos modelos e possíveis erros, sem gastar muitos recursos, visando desde testes preliminares até testes de sensibilidade. Enquanto isso, os cenários com 1.000 e 10.000 iterações foram definidos a fim de obter mais consistência e precisão em análises cujo conjunto de dados seja pequeno ou mais robusto.

Nessa perspectiva, a mineração de dados foi dividida em duas etapas: para a primeira, foi proposta a aplicação de simulações preliminares em cada cenário MMC, visando observar como os algoritmos se comportariam diante das configurações estabelecidas, principalmente para favorecer ajustes. Na segunda, foi realizada uma análise mais profunda dos cenários MMC, por meio da utilização de mais critérios, para sustentar a investigação.

2.4.1. Simulações Preliminares

Com base nos parâmetros expostos, as simulações preliminares consistem na observação de dois critérios de comparação, sendo eles: tempo de execução e escalabilidade. A expectativa é que a investigação forneça um diagnóstico inicial para indicar a necessidade de revisitar as etapas

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

de KDD e as configurações de aplicação, principalmente em casos de erros ou ajustes que exijam mais atenção ao longo da pesquisa.

Em paralelo aos cenários MMC estabelecidos, foi adotada a técnica de amostragem *Bootstrap* (Chernick, 2011) para particionar os dados e repor as amostras de forma aleatória. Outro aspecto relevante é que essas amostras foram estabelecidas de acordo com a quantidade de registros da base de dados, ou seja, 622 registros de transações normalizados.

Além disso, para garantir a consistência dos dados entre as execuções, o processo de aplicação do *Bootstrap* foi configurado com o parâmetro *Seed*, tomando como referência o valor 42, conforme adotado na pesquisa de Madhyastha e Batra (2019), principalmente para controlar a aleatoriedade e manter as mesmas condições de aplicação nos três cenários de simulação. Nesse contexto, o processo de execução dos algoritmos em cada cenário de simulação permite mensurar e avaliar a escalabilidade e o tempo de execução por meio de variações estatísticas, como tempo absoluto, média absoluta e desvio padrão.

2.4.2. Simulações de Sensibilidade

Conforme apontado no estudo de Farias *et al.,* (2016), para embasar as simulações de sensibilidade, foram adicionados novos critérios de observação, sendo eles: escalabilidade, tempo de execução e memória, mantendo os mesmos parâmetros definidos na Tabela 1.

A partir disso, os cenários MMC atuaram em conjunto com a técnica de amostragem *Bootstrap* para reposição de amostras aleatórias, e o parâmetro *Seed* para controlar a aleatoriedade. No entanto, a quantidade de registros sintéticos gerados seguiu uma abordagem diferente.

Esse processo envolveu levantar e avaliar os principais aplicativos de *delivery* de alimentos atuando no Brasil, com base no número de *downloads* e na avaliação dos usuários (entre quatro e cinco estrelas), pois esses aspectos representam o que o público realmente está buscando e utilizando. Assim, foram selecionados três aplicativos: o *iFood*, seguido do Zé Delivery de Bebidas e Mais Delivery, dentre os quais se destacou o *iFood.*

Para gerar os registros sintéticos necessários, levou-se em consideração o número de pedidos/mês registrados no iFood durante seu segundo ano de atuação no mercado de *delivery food service*. De acordo com Ozemela (2024), o *iFood* em 2012 alcançou o marco de 73 mil pedidos mensais, evidenciando um crescimento seis vezes maior em relação ao primeiro ano. Além de utilizar esse valor como referência, o conjunto sintético tem como diferencial o controle da cardinalidade transacional, com aumento aleatório da quantidade de itens por transação, variando entre 3 e 7 itens.

Mesmo com esses ajustes, foram mantidas as características da base de dados original do *Hambre Delivery*, considerando as 8 categorias de vendas indicadas por Gonçalves (2024), entre

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

as quais estão refeição, bebida regional, bebida alcoólica, hambúrguer, pizza, salgado, sobremesa e vitamina. Quanto à aplicação dos algoritmos e à especificação das métricas de execução, serão detalhadas na Seção Caso de Estudo.

2.5. Interpretação de Dados

Com base nos processos aplicados nas etapas de KDD, foram obtidos os resultados das simulações do MMC, considerando o desempenho computacional dos algoritmos em experimentos de extração de itens frequentes e geração de regras de associação.

As análises foram conduzidas por meio da interpretação de gráficos e tabelas comparativas que expõem os padrões obtidos. Esses resultados serão apresentados na Seção Resultados e Discussão.

3. CASO DE ESTUDO

Esta Seção descreve os parâmetros adotados para a aplicação dos algoritmos de mineração de dados, bem como os experimentos comparativos de desempenho realizados nas simulações preliminares e de sensibilidade. Os experimentos foram executados em um *MacBook Air* com sistema macOS Sequoia, equipado com processador M1 e 16 GB de memória unificada entre a CPU de 8 núcleos e a GPU de 7 núcleos. Todos os scripts de simulação foram desenvolvidos na linguagem Python.

3.1. Parâmetros Adotados para Configuração dos Algoritmos

Para avaliar o desempenho dos algoritmos, foram definidos experimentos específicos voltados à realização das simulações preliminares e de sensibilidade, utilizando os parâmetros de mineração de dados apresentados na Tabela 2.

Tabela 2. Parâmetros Adotados para Configuração dos Algoritmos

Parâmetro	Valor (%)			Referência
Suporte	0.6	0.8	1	Agrawal, Imieliński e Swami (1993)
Confiança	50	60	-	Suma e Sagar (2006) e Siswanto *et al.* (2024)

Fonte: Elaborado pelos autores

As configurações de suporte foram definidas dentro das variações de 0.1% a 5% indicadas por Agrawal, Imieliński e Swami (1993) para identificar itens frequentes e regras de associação, o que permite adaptações ao tipo de abordagem avaliada. Por isso, os valores de suporte adotados foram 0.6%, 0.8% e 1%, mantendo uma sequência nas simulações.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Os valores de confiança foram definidos em 50% para equilibrar a geração de regras de associação fortes e fracas (Suma e Sagar, 2006), enquanto 60% foram adotados para uma abordagem mais focada na obtenção de regras que ocorrem com mais frequência (Siswanto *et al.,* 2024).

3.2. Aplicação das Simulações Preliminares e de Sensibilidade

A aplicação das simulações se deu em duas etapas: preliminares e de sensibilidade. Para compreender o desempenho inicial dos algoritmos, as simulações preliminares testaram diferentes configurações, envolvendo a composição dos cenários MMC, a escolha dos parâmetros de suporte e confiança, e o desempenho quanto ao tempo de execução e à escalabilidade.

Em uma perspectiva mais ampla, a análise de sensibilidade submeteu os algoritmos a condições de extrema carga de uso, favorecendo a identificação de gargalos e limites operacionais, além de mensurar tempo de execução, uso de memória e escalabilidade.

Mesmo com essas diferenças, foram mantidos em ambas as etapas os experimentos de extração de itens frequentes e geração de regras de associação.

3.2.1. Experimento de Extração de Itens Frequentes

Este experimento propõe avaliar o desempenho na extração de itens frequentes, considerando as variações de suporte de 0.6%, 0.8% e 1%. Como instrumentos desse processo, foram selecionados os algoritmos Apriori, FP-Growth, FP-Max e Eclat.

O procedimento de comparação ocorreu mediante a aplicação de três cenários MMC, definidos para executar os algoritmos por 100, 1.000 e 10.000 iterações, o que possibilitou monitorar os recursos utilizados, a fim de mensurar a média absoluta e o desvio padrão durante as variações de suporte.

3.2.2. Experimento de Geração de Regras de Associação

Este experimento consiste em uma abordagem comparativa centrada na geração de regras de associação. A análise foi aplicada entre os algoritmos FP-Growth e Apriori, utilizando os parâmetros de suporte em 1% e as variações de confiança em 50% e 60%.

Em relação às simulações do MMC, mantiveram-se os mesmos cenários, definidos com 100, 1.000 e 10.000 iterações, garantindo continuidade e coerência em relação ao experimento de extração de itens frequentes.

3.2.3. Extração das Tendências de Venda

Este experimento propõe extrair as tendências de vendas contidas no conjunto de dados normalizado, composto por 622 registros de transações efetuadas via *app Hambre Delivery*.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Como instrumento para obtenção de tendências, foi utilizado o algoritmo de melhor desempenho na geração de regras de associação, configurado com suporte em 1% e confiança de 60%. Essa definição é experimental, baseada especificamente nos resultados de precisão alcançados nos experimentos anteriores.

4. RESULTADOS E DISCUSSÃO

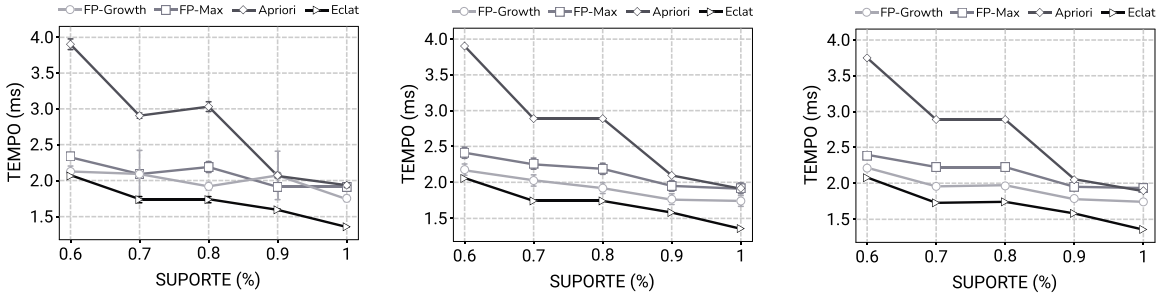
4.1. Resultados das Simulações Preliminares

Esta seção detalha os resultados das simulações preliminares referentes ao desempenho dos algoritmos de mineração de associação, sob critérios de escalabilidade e tempo de execução. Outro aspecto relevante foi a aplicação dessas simulações no conjunto de dados com 622 registros de transações, visando à escolha da ferramenta mais adequada para identificar os padrões de venda do *Hambre Delivery.*

4.1.1. Experimento de Extração de Itens Frequentes

Com base nos requisitos estabelecidos, foram obtidos os resultados de desempenho dos algoritmos FP-Growth, FP-Max, Apriori e Eclat. A Figura 6 apresenta o experimento de extração de itens frequentes, englobando os três cenários MMC. A Figura 6 (a) representa o cenário MMC inicial, com 100 iterações. A Figura 6 (b) ilustra um cenário MMC mais amplo, definido com 1.000 iterações. Por fim, a Figura 6 (c) detalha o cenário com 10.000 iterações.

Figura 6. Resultados das Simulações Preliminares Para a Extração de Itens Frequentes



Fonte: Resultados da pesquisa

Os resultados da extração de itens frequentes nos cenários de 100 e 1.000 iterações revelaram um desvio padrão significativo no tempo de execução dos algoritmos FP-Growth, FP-Max, Apriori e Eclat, que reduzem com o aumento das iterações. Todavia, o cenário de 10.000 iterações evidencia o comportamento mais estável dos modelos.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

A Tabela 3 sintetiza o tempo acumulado em milissegundos pelos algoritmos em cada cenário MMC, enquanto a Tabela 4 informa a quantidade de itens frequentes extraídos por eles nas variações de suporte 0.6%, 0.8% e 1%.

Tabela 3. Tempo Acumulado pelos Algoritmos na Extração de Itens Frequentes

MMC	Suportes (%)	Tempo (ms)			
		FP-Growth	FP-Max	Apriori	Eclat
100	0.6 - 1	987	1.028	1.374	846
1.000	0.6 - 1	9.745	10.544	13.560	8.469
10.000	0.6 - 1	97.350	105.750	133.620	85.210
Total (ms)		108.082	117.322	148.554	94.525
Total (min)		1.80	1.95	2.47	1.57

Fonte: Resultados da pesquisa

Tabela 4. Quantidade Itens Frequentes Obtidos pelos Algoritmos

Suporte (%)	Algoritmos			
	FP-Growth	FP-Max	Apriori	Eclat
0.6	168	72	168	238
0.8	126	63	126	168
1	83	49	83	91

Fonte: Resultados da pesquisa

Quanto ao desempenho dos modelos, os algoritmos FP-Growth e FP-Max alcançaram tempos de execução satisfatórios, nos quais concluíram as simulações em 1.80 (min) e 1.95 (min) nos três cenários do MMC, respectivamente. Por outro lado, o algoritmo Apriori levou cerca de 2.47 (min) para concluir as simulações.

Todavia, o algoritmo Eclat obteve o melhor desempenho entre os modelos, concluindo as simulações em 1.57 (min). Além disso, foi responsável por extrair a maior quantidade de itens. Essa capacidade de processamento está relacionada a dois fatores: à mineração vertical e ao pequeno volume de transações com baixa cardinalidade de itens por transação.

No contexto da escalabilidade, os cenários MMC demonstraram resultados mais precisos com o aumento no número de iterações, mostrando que o algoritmo Eclat se adequa melhor à tarefa de extrair itens frequentes em conjuntos de transações pequenos.

ISSN: 2675-6218 - RECIMA21

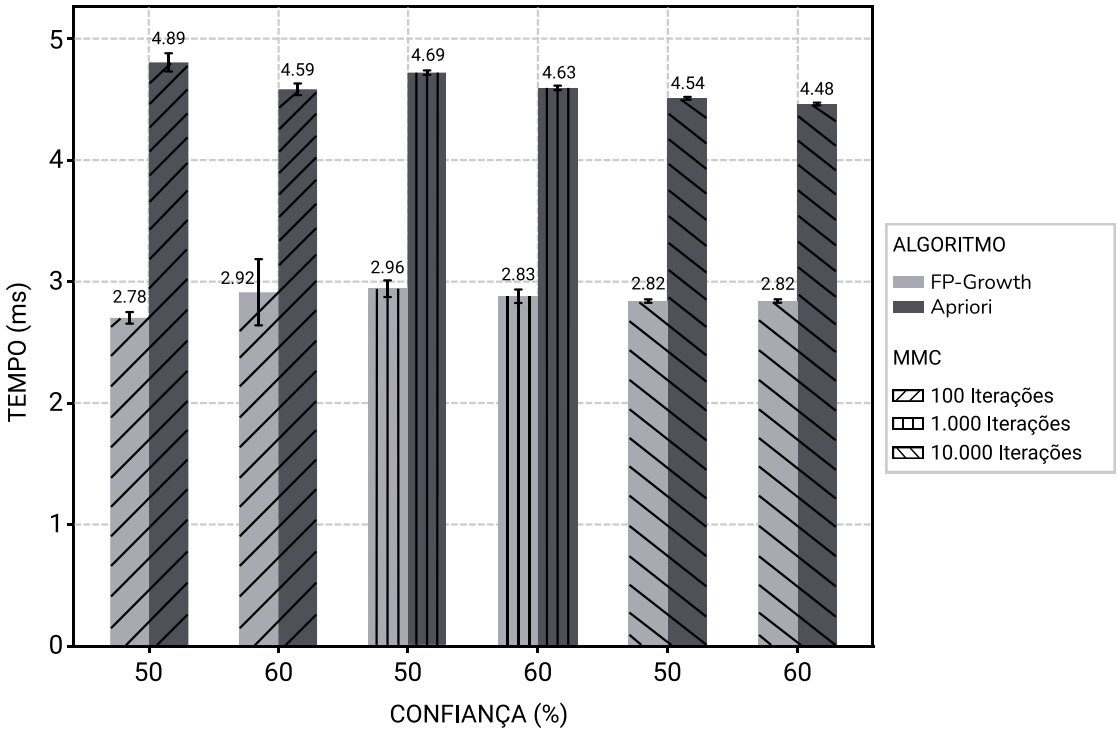
REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

4.1.2. Experimento de Geração de Regras de Associação

Conforme a proposta deste experimento, foram obtidos os resultados de desempenho dos algoritmos FP-Growth e Apriori. A Figura 7 apresenta a análise de geração de regras de associação para os três cenários MMC. Cada simulação MMC é representada por linhas dispostas sobre as barras dos algoritmos. As linhas inclinadas à direita ilustram o cenário de 100 iterações. As linhas verticais representam o cenário de 1.000 iterações. Por fim, as linhas inclinadas à esquerda representam o cenário de 10.000 iterações.

A Tabela 5 descreve a quantidade de regras de associação gerada pelos algoritmos nas variações de confiança de 50% e 60%. Ambos os algoritmos atuaram sob as mesmas condições e geraram as mesmas regras.

Figura 7. Resultados das Simulações Preliminares Para a Geração de Regras de Associação



Fonte: Resultados da pesquisa

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Tabela 5. Quantidade de Regras de Associação Obtidas pelos Algoritmos

Confiança (%)	Algoritmos	
	FP-Growth	Apriori
50	119	119
60	101	101

Fonte: Resultados da pesquisa

Os resultados da Figura 7 indicam uma redução no tempo de execução e no desvio padrão conforme o aumento no número de iterações. Essa estabilidade estatística é mais evidente no cenário MMC de 10.000 iterações.

Ainda neste cenário, o algoritmo FP-Growth alcançou tempos muito baixos, levando cerca de 2.82 (ms) para gerar regras de associação nas variações de confiança de 50 e 60%. Em contrapartida, o algoritmo Apriori precisou de mais tempo para concluir a tarefa, registrando 4.54 e 4.48 (ms) nas mesmas variações. Em termos de versatilidade, o algoritmo FP-Growth representa a escolha mais adequada para gerar regras de associação.

4.2. Resultados das Simulações de Sensibilidade

Esta seção apresenta os resultados das simulações de sensibilidade conduzidas em um conjunto sintético de 73 mil registros, gerado a partir dos dados originais do *Hambre Delivery*. A análise avalia a escalabilidade, o tempo de execução e o uso de memória dos algoritmos, a fim de identificar o melhor desempenho e possíveis limites operacionais.

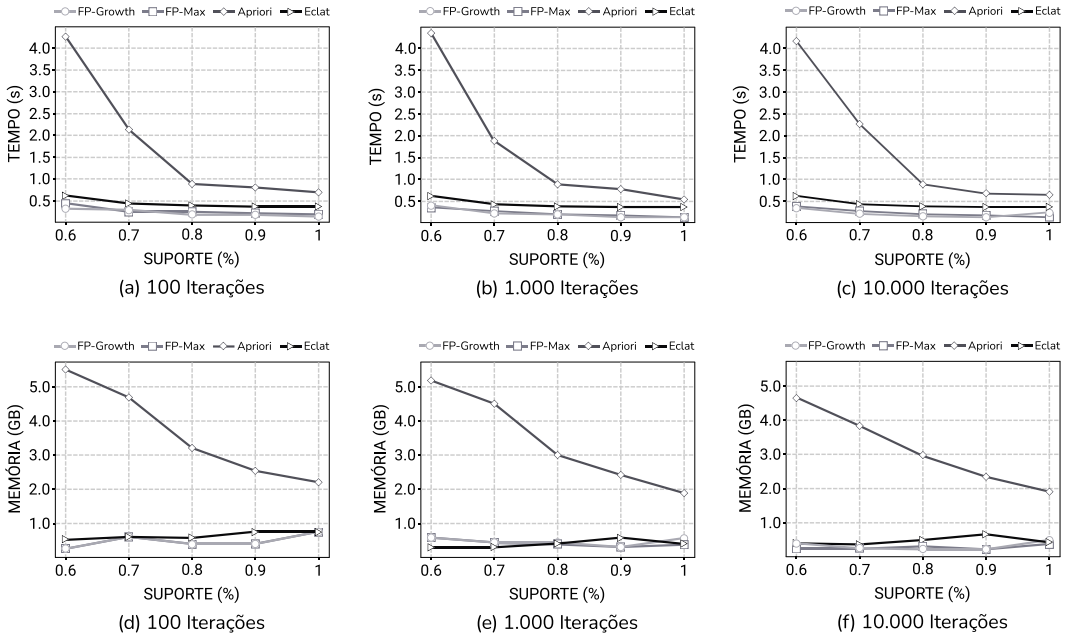
4.2.1. Experimento de Extração de Itens Frequentes

Conforme a proposta deste experimento, foi comparado o desempenho dos algoritmos FP-Growth, FP-Max, Apriori e Eclat. A Figura 8 apresenta os resultados da extração de itens frequentes, relativos ao tempo de execução e ao uso de memória, em três cenários MMC definidos com 100, 1.000 e 10.000 iterações. A Figura 8 (a), (b) e (c) ilustram as simulações de tempo de execução. A Figura 8 (d), (e) e (f) ilustram as simulações de uso de memória.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Figura 8. Resultados das Simulações de Sensibilidade para Extração de Itens Frequentes Relativos ao Tempo de Execução e Uso de Memória



Fonte: Resultados da pesquisa

Os resultados revelam coerência entre os recursos computacionais utilizados, indicando que a oscilação no tempo de execução foi proporcional ao uso de memória. Além disso, o desvio padrão foi praticamente nulo nas variações do MMC, resultando em padrões semelhantes.

Complementando a análise, a Tabela 6 sintetiza os resultados das simulações de itens frequentes, detalhando o tempo acumulado pelos algoritmos FP-Growth, FP-Max, Apriori e Eclat em cada cenário MMC, enquanto a Tabela 7 informa a memória média utilizada.

Tabela 6. Tempo Acumulado pelos Algoritmos nas Simulações de Extração de Itens Frequentes

MMC	Suportes (%)	Tempo (s)			
		FP-Growth	FP-Max	Apriori	Eclat
100	0.6 - 1	143.10	148.03	853.18	222.65
1.000	0.6 - 1	1.478,16	1.489,02	8.366,04	2.123,99
10.000	0.6 - 1	15.206,10	15.157,68	86.905,42	21.459,86
Total (s)		16.827,36	16.794,73	96.124,64	23.806,50
Total (h)		4.68	4.67	26.62	6.62

Fonte: Resultados da pesquisa

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Tabela 7. Memória Média Utilizada pelos Algoritmos na Extração de Itens Frequentes

MMC	Suportes (%)	Memória (GB)			
		FP-Growth	FP-Max	Apriori	Eclat
100	0.6 - 1	0.403	0.404	3.5	0.471
1.000	0.6 - 1	0.459	0.431	3.7	0.422
10.000	0.6 - 1	0.187	0.150	3.1	0.285
Média Total (GB)		0.350	0.329	3.43	0.393

Fonte: Resultados da pesquisa

Quanto ao desempenho dos modelos, pode-se perceber que os algoritmos FP-Growth e FP-Max obtiveram os melhores resultados nos critérios tempo de execução e uso de memória. Os valores registrados para o tempo acumulado em horas indicam que as simulações foram concluídas em 4.68 e 4.67 (h), com uso médio de memória de 0.350 e 0.329 (GB), respectivamente.

Nas mesmas condições, o algoritmo Apriori revelou o pior resultado entre os modelos analisados, levando cerca de 26.62 (h) para realizar as simulações. Esse desempenho mais lento aumentou o uso de memória, requisitando uma média de 3.43 (GB). Por outro lado, o algoritmo Eclat demonstrou desempenho melhor que o Apriori e mais próximo dos modelos FP-Growth e FP-Max, sendo necessárias 6.62 (h) para concluir as simulações. Esse aspecto é validado pela memória média registrada, com uso de 0.393 (GB).

A Tabela 8 informa a quantidade de itens frequentes extraídos pelos algoritmos nas variações de suporte 0,6%, 0,8% e 1%. Os resultados destacam que FP-Growth, Apriori e Eclat obtiveram a mesma quantidade de itens, enquanto o FP-Max extraiu menos, coerente com sua lógica de identificar itens frequentes máximos.

Tabela 8. Quantidade Itens Frequentes Obtidos pelos Algoritmos

Suporte (%)	Algoritmos			
	FP-Growth	FP-Max	Apriori	Eclat
0.6	737	397	737	737
0.8	517	307	517	517
1	410	230	410	410

Fonte: Resultados da pesquisa

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

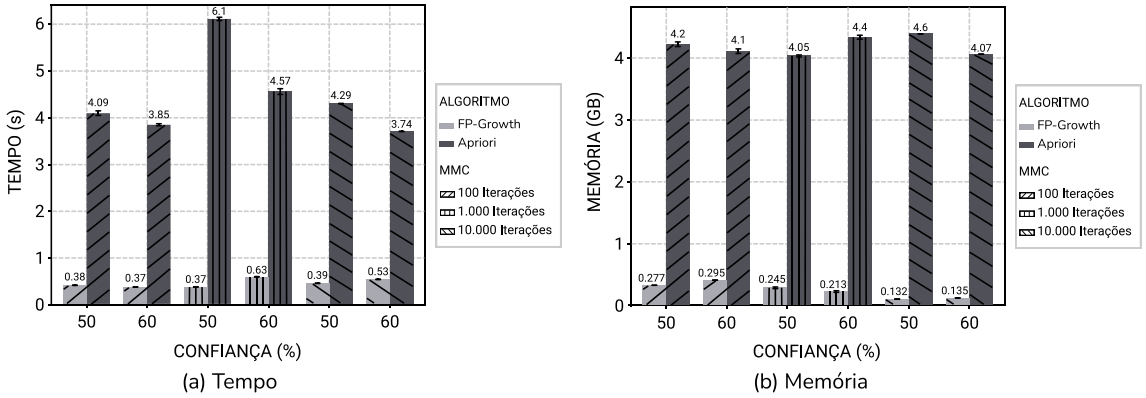
A partir dos resultados discutidos, é possível concluir que os algoritmos FP-Growth e FP-Max são mais escaláveis para lidar com grandes volumes de transações e alta cardinalidade por transação. Contudo, a escolha entre FP-Growth e FP-Max depende do tipo de análise, da complexidade da base de dados e da quantidade de itens frequentes que se pretende extrair.

4.2.2. Experimento de Geração de Regras de Associação

Considerando o enfoque deste experimento foram comparados os algoritmos FP-Growth e Apriori. A Figura 9 ilustra o desempenho dos algoritmos FP-Growth e Apriori na geração de regras de associação, sob os aspectos de tempo de execução e uso de memória.

Cada simulação MMC é identificada por linhas de variação dispostas sobre as barras dos algoritmos. As linhas inclinadas à direita representam o cenário MMC com 100 iterações. As linhas verticais ilustram o cenário MMC de 1.000 iterações. Por fim, as linhas inclinadas à esquerda representam a simulação MMC de 10.000 iterações.

Figura 9. Resultados das Simulações para a Geração de Regras de Associação



Fonte: Resultados da pesquisa

Os resultados da geração de regras de associação demonstram consistência entre os recursos computacionais analisados, revelando que as oscilações no tempo de execução são proporcionais ao uso de memória, além de apresentarem redução no desvio padrão conforme o aumento das iterações nos cenários MMC.

A Tabela 9 sintetiza os resultados das simulações de regras de associação, detalhando o tempo acumulado pelos algoritmos FP-Growth e Apriori em cada cenário MMC, enquanto a Tabela 10 detalha a memória média utilizada.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Tabela 9. Tempo Acumulado pelos Algoritmos nas Simulações Para a Geração de Regras de Associação

MMC	Confiança (%)	Tempos (s)	
		FP-Growth	Apriori
100	50 - 60	76.12	795.67
1.000	50 - 60	1.011,09	10.675,16
10.000	50 - 60	9.313,27	80.407,33
Total (s)		10.400,48	91.878,16
Total (h)		2,89	28,41

Fonte: Resultados da pesquisa

Tabela 10. Memória Média Utilizada pelos Algoritmos na Geração de Regras de Associação

MMC	Confiança (%)	Memória (GB)	
		FP-Growth	Apriori
100	50 - 60	0.286	4.08
1.000	50 - 60	0.229	4.12
10.000	50 - 60	0.134	4.23
Média Total (GB)		0.216	4.14

Fonte: Resultados da pesquisa

Em relação à análise de desempenho dos modelos, pode-se observar que o algoritmo FP-Growth apresentou os melhores resultados, levando cerca de 2.89 (h) para concluir as simulações. Esse desempenho refletiu no baixo consumo de memória apresentado, com uso médio de 0.216 (GB). Em condições semelhantes, o algoritmo Apriori apresentou os piores resultados, sendo necessário 28.41 (h) para realizar as simulações. Esse desempenho mais lento contribuiu para o uso elevado de memória, que demandou uma média de 4.14 (GB).

A Tabela 11 complementa a análise, detalhando a quantidade de regras de associação geradas pelos algoritmos Apriori e FP-Growth nas variações de confiança de 50 e 60%. Como ambos atuam sob as mesmas condições de configuração, as regras de associação obtidas foram as mesmas.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Tabela 11. Quantidade de Regras de Associação Obtidas pelos Algoritmos

Confiança (%)	Algoritmos	
	FP-Growth	Apriori
50	795	795
60	269	269

Fonte: Resultados da pesquisa

Diante dos resultados obtidos, é possível concluir que o algoritmo FP-Growth demonstrou ser mais escalável na mineração de grandes volumes de dados, especialmente no tempo de execução e uso de memória.

4.3. Análise das Tendências de Venda

Os resultados da simulação de tendências de venda são provenientes da aplicação do algoritmo FP-Growth, no qual foram obtidas 27 regras de associação. Por se tratar de uma análise em escopo reduzido, serão apresentadas as 8 regras mais relevantes com base nas orientações da *startup* responsável pelo *Hambre Delivery*, conforme detalha a Tabela 12.

Tabela 12. Regras de Associação Geradas pelo Algoritmo de Melhor Desempenho

ID	Antecedente	Consequente	Confiança (%)	Lift
A1	Pão Careca	Pão Hambúrguer	100	51.83
A2	Pão Hambúrguer	Pão Careca	83.33	51.83
A3	Lanche de Empadas, Canudinho com Creme de Frango	Salgados	80	9.21
A4	Lanche de Empadas, Salgados	Canudinho com Creme de Frango	66.67	14.81
A5	Pizza Metade Bacon	Pizza Metade Portuguesa	100.0	31.1
A6	Pizza Metade Portuguesa	Pizza Metade Bacon	60.0	31.1
A7	Farinha de Mandioca	Açaí de 1 Litro	84.38	14.58
A8	Açaí de 1 Litro	Farinha de Mandioca	75.0	14.58

Fonte: Resultados da pesquisa

Quanto às tendências registradas na Tabela 12, referentes aos setores de salgado, pizzaria e bebida regional, pode-se perceber que as regras A1 e A2 são complementares na montagem de lanches. Isso indica que os itens Pão Careca e Pão Hambúrguer são frequentemente adquiridos

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

juntos, tendo 100% de confiança na relação direta e 83,33% na inversa, sugerindo uma associação extremamente forte.

Também se destacam as regras A3 e A4, que indicam a preferência dos clientes pela combinação de Lanches de Empadas, Canudinho com creme de frango e Salgados. A regra A3 revela que essa combinação ocorre em 80% dos casos, sendo uma tendência recorrente de lanches mistos para mais pessoas.

As regras A5 e A6 revelam uma associação bidirecional entre os sabores bacon e portuguesa. Essa combinação representa uma tendência mais clássica e recorrente entre grupos de amigos e familiares, na qual os clientes que escolhem a Pizza Metade Bacon tendem a combiná-la com a Pizza Metade Portuguesa, com 100% de confiança na escolha direta e 60% na inversa.

Por fim, as regras A7 e A8 indicam uma prática cultural consolidada no consumo *de* Farinha de Mandioca e Açaí de 1 Litro. Por serem complementares a qualquer refeição, os consumidores que compram Farinha de Mandioca frequentemente levam o Açaí de 1 Litro, com 84.38% de confiança na relação direta e 75% na inversa.

5. CONSIDERAÇÕES

Esta pesquisa alcançou o objetivo de avaliar o desempenho computacional de algoritmos de mineração de associação, combinando o MMC para obter resultados mais precisos e identificar o algoritmo mais eficiente para suporte à decisão no *app Hambre Delivery.*

O alcance dessa finalidade requereu a realização de metas específicas, como a seleção do conjunto de transações do *Hambre Delivery*, a implementação dos algoritmos Apriori, FP-Growth, FP-Max e Eclat, a definição de configurações e cenários do MMC com base em experimentos preliminares e de sensibilidade, além da criação de *scripts* para coletar e avaliar o desempenho quanto à escalabilidade, tempo de execução e uso de memória.

A aplicação de métodos combinados, tendo em vista o desempenho computacional, revelou duas realidades distintas, conforme os experimentos. Na primeira, foi observado que a aplicação de simulações preliminares em um conjunto de transações pequeno indicou que o algoritmo Eclat cumpriu melhor a tarefa de extrair itens frequentes, em virtude da baixa cardinalidade por transação, enquanto o algoritmo FP-Growth se mostrou mais eficiente para gerar regras de associação, principalmente pelo uso da FP-Tree. Além de ambos serem escaláveis quanto ao tempo de execução, a proposta dos algoritmos pode ser generalizada para ambientes em que as vendas estão em fase inicial e não requerem modelos extremamente complexos e robustos.

Na segunda, as simulações de sensibilidade aplicadas em um conjunto de dados sintético com mais transações e maior cardinalidade de itens por transação revelaram que ambos os algoritmos FP-Growth e FP-Max foram eficientes na extração de itens frequentes, mas o algoritmo FP-Growth também foi eficaz na geração de regras de associação, uma vez que ambos utilizam a

ISSN: 2675-6218 - RECIMA21

FP-Tree para otimizar essa tarefa de mineração de dados. Por esse motivo, foram escaláveis em relação ao tempo de execução e ao uso de memória, comprovando que essa abordagem pode ser ampliada para ambientes de vendas consolidados no mercado de *food service*, que apresentam conjuntos de dados mais complexos e necessitam lidar com tendências de vendas mais precisas.

Por mais que esta pesquisa utilize métodos de análise conhecidos, adaptando critérios e métricas, nenhum trabalho apresenta uma abordagem que combine o MMC e algoritmos de mineração de dados por associação para avaliar o desempenho computacional.

Diante deste trabalho, é possível concluir que a abordagem proposta é viável para apoiar a escolha do modelo mais estável como suporte à decisão de tendências de venda. Conclui-se que a aplicação de simulações MMC a partir de 1.000 iterações resultará em resultados mais precisos, demonstrando que essa sistematização pode ser generalizada para outros contextos de mineração de dados, sejam em conjuntos de dados mais simples ou mais robustos.

Para as próximas etapas desta pesquisa, pretende-se integrar os *scripts* de pré-processamento e transformação de dados ao servidor do projeto *Hambre Delivery*. Além disso, pretende-se expandir o código-fonte para incorporar o algoritmo de mineração de associação mais eficiente à versão final do *app*.

REFERÊNCIAS

AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **ACM SIGMOD Record**, v. 22, n. 2, p. 207–216, 1993.

BHAGAMPRIYAL, M. *et al.* Recommendation Systems for Supermarket. 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM). **Anais [**...] Uttar Pradesh, India: IEEE, 2023.

BUJALANCE-LÓPEZ, L. *et al.* Restaurant revenue management: a systematic literature review and future challenges. **British Food Journal**, v. 127, n. 6, p. 2169–2196, 1 abr. 2025.

CAMPOS, W. P.; FARINA, R. M.; FLORIAN, F. Inteligência Artificial: Machine Learning na Gestão Empresarial. **RECIMA21 - Revista Científica Multidisciplinar**, v. 3, n. 6, p. e361617, 24 jun. 2022.

CHERNICK, M. R. et al. Bootstrap Methods. *In:* LOVRIC, M. (Ed.). **International Encyclopedia of Statistical Science**. Berlin: Springer, 2011. p. 169–174.

FARIAS, F. *et al.* Cost- and energy-efficient backhaul options for heterogeneous mobile network deployments. **Photonic Network Communications**, v. 32, n. 3, p. 422–437, 22 nov. 2016.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37–37, 15 mar. 1996.

GONÇALVES, K. R. G. **Desenvolvimento e Avaliação do Aplicativo Hambre Delivery:** Apresentação do Produto e Análise de Desempenho Usando Inteligência Artificial. Trabalho de

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Conclusão de Curso (Bacharelado em Sistemas de Informação) – Faculdade de Sistemas de Informação – FASI, Cametá, PA, 2024.

GRAHNE, G.; ZHU, J. Fast algorithms for frequent itemset mining using FP-trees. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, n. 10, p. 1347–1362, out. 2005.

GRIJALBA, M. A. *et al.* Does the use of digital tools improve a firm's performance? **Review of Managerial Science**, v. 19, p. 2193–2210, 2 mar. 2024.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining:** Concepts and Techniques. 3. ed. Burlington: Elsevier, 2012.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, n. 1, p. 97–109, 1 abr. 1970.

HEIJUNGS, R. On the number of Monte Carlo runs in comparative probabilistic LCA. **The International Journal of Life Cycle Assessment**, v. 25, n. 2, p. 394–402, 2020.

HU, T. *et al*. Discovery of maximum length frequent itemsets. **Information Sciences**, v. 178, n. 1, p. 69–87, Jan. 2008.

JOSEPHINE, H.; RAJAN, D. Enhancing Customer Experience and Sales Performance in a Retail Store Using Association Rule Mining and Market Basket Analysis. 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). **Anais** [...] Delhi, India: IEEE, 2023.

LIN, W.-Y.; TSENG, M.-C.; SU, J.-H. A Confidence-Lift Support Specification for Interesting Associations Mining. **Advances in Knowledge Discovery and Data Mining**, v. 2336, p. 148–158, 2002.

LIU, W. *et al.* Mobile platform expansion: How does it affect the incumbent food delivery app and other sales channels? **Journal of Retailing**, v. 100, n. 3, p. 422–438, 1 jun. 2024.

MADHYASTHA, P.; BATRA, D. On Model Stability as a Function of Random Seed. Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). **Anais** [...] Hong Kong: Association for Computational Linguistics, 1 Jan. 2019.

OZEMELA, L. **iFood alcança marco histórico de 100 milhões de pedidos em um só mês**. [S. l.]: iFood, 2024. Disponível em: https://institucional.ifood.com.br/noticias/ifood-alcanca-marco-historico-de-100-milhoes-de-pedidos-em-um-so-mes/?utm_source=chatgpt.com. Acesso em: 11 abr. 2025

PRADANA, M. R. *et al.* Market Basket Analysis Using FP-Growth Algorithm on Retail Sales Data. 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). **Anais** [...] Jakarta, Indonesia: IEEE, 2022.

RABHI, F.; BEHESHTI, A.; GILL, A. Editorial: Business transformation through AI-enabled technologies. **Frontiers in Artificial Intelligence**, v. 8, 11 mar. 2025.

SISWANTO, B. *et al.* SDFP-growth Algorithm as a Novelty of Association Rule Mining Optimization. **IEEE access**, v. 12, p. 21491–21502, 1 Jan. 2024.

SUMA, D.; SAGAR, L. Data Mining Techniques. **Technometrics**, v. 48, n. 1, p. 159–160, fev. 2006.

ISSN: 2675-6218 - RECIMA21

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

TANAKA, T. *et al.* Optimality Between Time of Estimation and Reliability of Model Results in the Monte Carlo Method: A Case for a CGE Model. **Computational Economics**, v. 59, n. 1, p. 151–176, 3 jan. 2021.

ZAKI, M. J. *et al.* New algorithms for fast discovery of association rules. **Knowledge Discovery and Data Mining**, p. 283–286, 14 ago. 1997.

ZAKI, M. J. Scalable algorithms for association mining. **IEEE Transactions on Knowledge and Data Engineering**, v. 12, n. 3, p. 372–390, 2000.

ISSN: 2675-6218 - RECIMA21