

UTILIZAÇÃO DE MODELOS DE REGRA DE ASSOCIAÇÃO PARA GESTÃO DE VENDAS DE SUPERMERCADOS

CARLA CRISTINA CAMARGO MARTINS¹; LUAN ANGELO ROCHA¹; MARCEL SANTOS SILVA²; ELOÍZA MARTINS PRIMO CAPELOCI²

¹Discente em Big Data No Agronegócio na FATEC Pompeia “Shunji Nishimura”, Pompeia-SP. carla.martins2@fatec.sp.gov.br; luan.rocha4@fatec.sp.gov.br

²Docentes do curso de Big Data no Agronegócio, FATEC Pompeia “Shunji Nishimura”, Pompeia-SP. marcel.silva9@fatec.sp.gov.br; eloiza.capeloci@fatec.sp.gov.br

RESUMO

O comércio varejista possui um papel de extrema importância na economia do país representando 23% do Produto Interno Bruto (PIB) brasileiro, gerando receitas significativas todos os anos. Com a evolução constante do setor, uma gestão eficiente das vendas é essencial para administração dos negócios. Este trabalho detalha uma análise do padrão de compra dos clientes em um supermercado localizado na cidade de Queiroz - SP, no período de 2022 e 2023, em que foi utilizado o algoritmo de *Machine Learning* Apriori. A análise foi capaz de revelar uma variedade de associações de produtos, desde combinações frequentes, até associações menos óbvias, indicando a diversidade das preferências dos clientes. Ferramentas avançadas de *Business Intelligence* (BI) e a plataforma Microsoft Power BI foram utilizadas para transformar dados brutos em gráficos interativos e visualmente acessíveis. Esses gráficos proporcionaram uma representação visual dos padrões de compra e outras informações referentes a base de dados, facilitando a interpretação do comportamento de consumo da clientela. Além disso, um aplicativo foi construído na plataforma Power Apps proporcionando aos usuários uma interface fácil de usar para explorar os *insights* descobertos, impulsionando a estratégia de marketing e aumentando os lucros do estabelecimento. O desenvolvimento deste trabalho mostrou uma visão detalhada do padrão de compra dos clientes e estabeleceu uma base sólida para investigações futuras a fim de expandir para outros setores. Com a aplicação da tecnologia de *Machine Learning* no setor de varejo, foram criadas estratégias de negócios eficazes baseadas em dados concretos e interpretações precisas.

Palavras - chaves: Apriori, *Machine Learning*, Gestão de vendas, *Business Intelligence*

INTRODUÇÃO

O comércio varejista contribui consideravelmente para a economia brasileira sendo responsável pela geração de receitas tanto para os varejistas quanto para os fornecedores e produtores. O setor de vendas no Brasil está em constante evolução e vem se tornando cada vez mais competitivo, de acordo com a Associação Brasileira de Supermercados (ABRAS), o setor de vendas de varejo faturou cerca de R\$695,7

bilhões no ano de 2022, valor referente ao aumento de 13,8% em relação ao ano anterior. (Abrás, 2022)

A globalização impulsionou o desenvolvimento do mercado, exigindo dele a gestão das demandas em busca do balanceamento e alinhamento estratégico desses conceitos ao longo da gestão de vendas originadas das necessidades dos supermercados. Com a gestão das demandas, busca-se o equilíbrio entre capacidades da empresa e necessidades dos clientes, tendo a redução de incertezas e o fornecimento de *insights* para administração dos produtos. Considerado um comportamento importante para o sucesso das vendas, com bons resultados dessa implementação obtém-se benefícios diversos. (MELO, 2010)

Uma das alternativas para melhorar a gestão de vendas de um estabelecimento é a utilização da tecnologia, conhecida como *Machine Learning* (Aprendizado de Máquinas), que é um mecanismo capaz de identificar padrões, tomar decisões e estabelecer previsões baseadas nos padrões analisados por meio de um algoritmo. O processo do Aprendizado de Máquinas envolve sete etapas: coletar dados, pré-processar dados e escolher um algoritmo, treinar, avaliar, otimizar e implantar. (BOENTE; OLIVEIRA; ROSA, 2020)

Um dos algoritmos utilizado no *Machine Learning* para regras de associações é o Apriori. De acordo com Vasconcelos e Carvalho (2004) a tarefa da associação tem o intuito de identificar associações entre registros de dados que, de alguma maneira, estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam na presença de outros em uma mesma transação, ou seja, esse algoritmo é capaz de encontrar padrões entre os produtos de uma base de dados e com isso prever comportamentos relevantes entre dados específicos.

Atualmente, nos lugares onde a tecnologia desempenha um papel importante nas operações comerciais, a gestão é mais do que uma necessidade, mas sim uma estratégia para o melhor desempenho. Neste caso, gerir os dados de vendas em um período determinado, torna-se possível alcançar redução de custos, conseqüentemente, aumentando a geração de receita, com esse estudo também é possível identificar padrões entre os produtos, gerando conhecimentos sobre o público-alvo, capacitando a criação de estratégias de marketing que farão dos planejamentos futuros um sucesso.

Ao aderir o *Knowledge Discovery in Database* (KDD), traduzido como Descoberta de Conhecimento de Base de Dados, as empresas podem extrair informações valiosas

por meio da aprendizagem organizacional que busca gestões estratégicas do conhecimento (BRANQUINHO, 2015). Por outro lado, o processo de *Business Intelligence* (BI) visa novas percepções sobre negócios e mercado, com o objetivo de alcançar eficiência operacional e promover um relacionamento eficaz com os clientes (CHEN G; CHEN H; LIM, 2013). Ao gerar gráficos a partir de grandes volumes de dados, o BI consegue apresentar informações relevantes, transformando dados brutos em conhecimento e oferecendo auxílio direto na tomada de decisão empresarial.

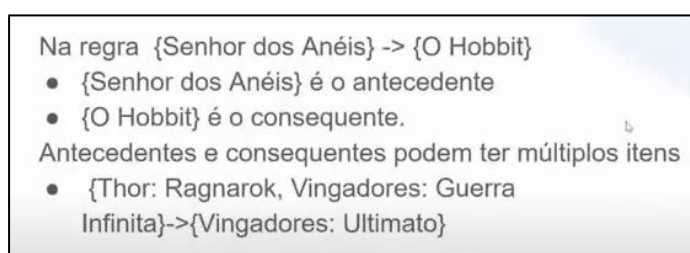
O presente estudo desenvolver um modelo de Aprendizado de Máquina capaz de realizar previsões de vendas, para auxiliar a gestão das demandas do mercado varejista utilizando a regra de associação Apriori, desenvolvida na linguagem de programação *Python*, após a mineração e análise da base obtida de um supermercado. Com o intuito de auxiliar a tomada de decisão, gerando predição dos produtos mais demandados, tendo em vista aumentar o lucro, reduzir perdas do estoque e gerar *insights* valiosos para marketing.

REGRA DE ASSOCIAÇÃO

A Regra de Associação se caracteriza pela presença de um aglomerado de atributos nos dados de uma base de dados que influencia em outro aglomerado de atributos distintos dentro da mesma base (MELANDA; REZENDE, 2003). As regras de associação possuem dois elementos principais: o antecedente e o conseqüente. Antecedente é o conjunto de itens que apresentam uma causa ou condição para ocasionar o conseqüente. Representado pelo lado “se” da regra o antecedente descreve as características e/ou eventos de um comportamento.

O conseqüente representa o conjunto de itens que são os resultados ou conseqüência do ocorrido no antecedente. Na regra é representado como o lado “então”, e apresenta eventos e ou ações que irão ocorrer após o antecedente.

Figura 1: Exemplo de antecedente e conseqüente



Fonte: ALMADA (2020)

A Figura 1 exemplifica a funcionalidade do antecedente e consequente, em que os filmes estão representando os itens, ou seja, “Senhor dos Anéis” é o antecedente para aqueles que assistirão “O Hobbit”.

Quando se possui um conjunto de itens como “Thor: Ragnarok” e “Vingadores: Guerra Infinita”, então como consequência pode-se presumir que o indivíduo assistirá “Vingadores: Ultimato”.

O algoritmo Apriori aplica a regra de associação e procura por relações entre os dados enquanto eles são separados, calculando o valor correspondente a confiança e ao suporte. Tal algoritmo define dois conceitos: *itemsets* frequentes e *itemsets* candidatos. Os *itemsets* frequentes são aqueles que o valor de suporte e confiança superam os limites definidos, e os *itemsets* candidatos, consistem nos itens que podem ser frequentes e devem ser explorados para confirmar essa frequência. (TANURE, 2010). O Apriori faz o uso de três medidas para geração da associação, sendo elas: *Support*, *Confidence* e *Lift*.

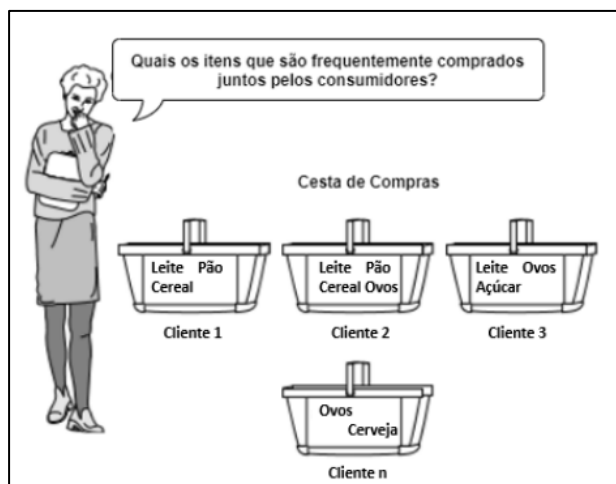
A medida *Support* (Suporte) é a frequência de um aglomerado de itens de acordo com o total de transações, pode ser calculado como uma porção das transações que contêm esse conjunto de itens.

Já a medida *Confidence* (Confiança) indica a probabilidade de que o consequente ocorra, medindo a força da influência entre o antecedente e consequente, sendo uma confiança alta o indicativo de uma forte dependência entre os itens.

A medida *Lift* compara a frequência da ocorrência de antecedente e consequente com a frequência esperada se eles fossem independentes. O *Lift* indica uma boa associação entre os *itemsets*, ou seja, a ocorrência do antecedente aumenta a probabilidade de que ocorra o consequente. Senso assim, o *Lift* menor que 1 indica uma associação indireta entre os *itemsets*.

As medidas supramencionadas auxiliam na identificação de padrões interessantes e úteis em um conjunto de dados, sendo usadas para avaliar a relevância, a força e a significância das regras de associação descobertas e permitindo *insights* valiosos para análises, tomadas de decisões e estratégias de negócios.

Figura 2: Cesta de compras



Fonte: LINOFF E BERRY (2011). Adaptado

Levando em considerações as transações realizadas dentro de um varejo, assim como demonstrado na Figura 2, a regra de associação atua ao receber como entrada diferentes compras em uma base de dados, em que cada cesta é formada por itens obtidos na compra individual, ignorando a quantidade e preço de cada item. Após a aplicação das regras de associações, *insights* sobre a mercadoria são gerados, trazendo informações de quais produtos saem juntos nas transações e quais podem ser colocados em oferta. (FROEHLICH, 2020)

MATERIAIS E MÉTODOS

Para o desenvolvimento do modelo de Aprendizado de Máquinas foi usado o ambiente de desenvolvimento (IDE) Visual Studio Code (VSCODE), que é uma plataforma que possibilita a execução de códigos em diversas linguagens de programação. (HANASHIRO, 2021)

A biblioteca Pandas é utilizada para linguagem *Python* em que tem funcionalidade de realizar a limpeza e tratamento da base, a fim de fazer a análise exploratória dos dados (SANTOS; MORENO, 2022). Nesse projeto, a biblioteca Pandas foi importante para realizar o *upload* da base e auxiliou em todo o pré-processamento. *Numpy*, é a biblioteca bastante usada para *arrays* multidimensionais e funções categóricas de alto desempenho, utilizada neste trabalho para realizar as transformações dos dados numéricos em categóricos. (MULINARI, 2023)

Para a aplicação do algoritmo Apriori, fez-se necessário a importação da biblioteca *apriori*, que oferece suporte de execução ao algoritmo usado para relacionar os produtos da base de dados (MOCHIZUKI, 2019).

O método “OneHotEncoder()” efetua a criação de todas as colunas que foram criadas durante o treinamento de uma base de dados. Atribuindo valores zero (0) e um (1), relacionados a ausência ou presença das características, respectivamente. Para utilizar esse método se faz necessário a criação de uma instância do objeto e, em seguida, ajustá-la aos dados por meio do método “fit()” que retém as informações para realizar a transformação dos dados. (MIRANDA, 2023)

O Microsoft Power BI realiza a geração de gráficos interativos e *insights* (MICROSOFT, 2023), nesse trabalho foi utilizado para melhor visualização das informações da base.

A fim de desenvolver uma interface *Front-End* de exibição e interação com o usuário, utilizou-se o software Microsoft Power Apps. Este é capaz de criar aplicativos que usam *low-code*, modernizando e facilitando processos complexos. (MICROSOFT, 2023)

O Microsoft Sharepoint é uma ferramenta utilizada para o armazenamento de dados online (MICROSOFT, 2023). Neste trabalho se fez necessário um meio capaz de armazenar de forma conjunta os dados obtidos no *Front-End*.

A base de dados utilizada no presente trabalho foi disponibilizada por um supermercado, localizado na cidade de Queiroz-SP, que apresenta os dados das vendas referentes ao período de agosto de 2022 a agosto de 2023.

DISCUSSÃO E RESULTADOS

A base de dados original, possui um total de 13 atributos: “codigo_caixa”, “codigo_venda”, “código_cliente”, “Nome_Cliente”, “Id_Produto”, “Descricao_Produto”, “unidade”, “unitário”, “quantidade”, “total”, “Data”, “hora”, “Id_Loja”, totalizando 1.475.681 linhas, em que estão disponibilizados os dados referentes as transações de compras realizadas no período supramencionado, a Figura 3 representa a base de dados completa.

Figura 3: Base de dados Original

codigo_caixa	codigo_venda	codigo_cliente	Nome_Cliente	ID_PRODUTO	DESCRICAO_PRODUTO	unidade	unitario	quantidade	total	DATA	hora	ID_LOJA	
0	1	877033	0	CLIENTE NÃO IDENTIFICADO	33419	CAFE 500G MACALI EXT.FORTE	UN	14.99	1.0	14.99	02/08/2022	08:05:55	1
1	1	877033	0	CLIENTE NÃO IDENTIFICADO	19085	MANTEIGA 200G AVIACAO POTE	UN	15.40	1.0	15.40	02/08/2022	08:05:55	1
2	1	877033	0	CLIENTE NÃO IDENTIFICADO	3723	ESMALTE RISQUE CARMIM	UN	4.50	1.0	4.50	02/08/2022	08:05:55	1
3	1	877034	0	CLIENTE NÃO IDENTIFICADO	38753	SAB.85G JOHNSONS DAILY CARE ROSAS	UN	1.99	1.0	1.99	02/08/2022	08:07:16	1
4	1	877034	0	CLIENTE NÃO IDENTIFICADO	38753	SAB.85G JOHNSONS DAILY CARE ROSAS	UN	1.99	1.0	1.99	02/08/2022	08:07:16	1
...
1475676	1	918948	1022	JOAO VITOR DE MOURA SILVA	11529	SALGADINHO KELECK 60G PURURUCA	UN	2.65	1.0	2.65	28/08/2023	19:36:21	1
1475677	3	895964	1022	JOAO VITOR DE MOURA SILVA	34299	TIRINHAS DE FRANGO 300G AURORA	UN	5.99	1.0	5.99	29/08/2023	13:07:45	1
1475678	3	895964	1022	JOAO VITOR DE MOURA SILVA	34299	TIRINHAS DE FRANGO 300G AURORA	UN	5.99	1.0	5.99	29/08/2023	13:07:45	1
1475679	4	744666	1023	THAISA APARECIDA BONACINA	1797	BANANA NANICA KG	KG	6.99	1.0	6.99	26/08/2023	17:24:08	1
1475680	2	1014049	1023	THAISA APARECIDA BONACINA	52969	JOGO GARRAFA C/3	UN	39.90	1.0	39.90	28/08/2023	12:11:53	1

1475681 rows x 13 columns

Fonte: Elaborado pelos autores (2023)

Para iniciar a análise exploratória da base de dados, foi verificado que não há valores nulos, por essa razão não foi necessário modificar linhas faltantes. Para realização desta operação foi implementada a função “.isna” da biblioteca Pandas.

O mercado em questão, realiza vendas à vista e a prazo, na base original os clientes que realizavam compras à vista foram representados como “CLIENTE NÃO IDENTIFICADO” e os que realizaram compras a prazo eram identificados pelos seus respectivos nomes, a Figura 4 apresenta a separação desses dados, criando duas variáveis, “qtd_a_vista” e “qtd_a_prazo”, para guardar esses valores separadamente.

Figura 4: Variável qtd_a_vista

```
# quantidade de compras feitas a vista
filtro = base['Nome_Cliente'] == 'CLIENTE NÃO IDENTIFICADO'
qtd_a_vista = base[filtro].groupby('Nome_Cliente')['codigo_venda'].nunique()

print(qtd_a_vista)

✓ 0.9s

Nome_Cliente
CLIENTE NÃO IDENTIFICADO    167090
Name: codigo_venda, dtype: int64

# quantidade de compras a prazo
qtd_a_prazo = qtd_compras - qtd_a_vista
int(qtd_a_prazo)

✓ 0.0s

17426
```

Fonte: Elaborado pelos autores (2023)

Para a transformação da data, pode ser visto na Figura 5, foi utilizado a função “datetime” para modificar os valores em data e hora, desta forma tornou-se possível

realizar a separação da coluna de data e hora que era uma data completa (dia-mês-ano-hora) em suas colunas respectivas.

Figura 5: Criação das Colunas de Data

	ano	dia	mes	hora_inteira
	2022.0	8.0	2.0	8
	2022.0	8.0	2.0	8
	2022.0	8.0	2.0	8
	2022.0	8.0	2.0	8
	2022.0	8.0	2.0	8

Fonte: Elaborado pelos autores (2023)

As colunas “código_caixa”, “unidade”, “unitário”, “Id_Loja”, “Data” e “hora”, foram excluídas no processo de análise da base por não serem consideradas importantes para a futura execução do algoritmo de associação. Para isso, foi utilizado a função `drop()` do Pandas que foi apresentado na Figura 6.

Figura 6: Exclusão das colunas

```
# excluindo colunas
base = base.drop(columns=['codigo_caixa', 'unidade', 'unitario', 'ID_LOJA', 'DATA', 'hora'])
base.head()
```

	codigo_venda	codigo_cliente	Nome_Cliente	ID_PRODUTO	DESCRICAO_PRODUTO	quantidade	total	ano	dia	mes	hora_inteira
0	877033	0	CLIENTE NÃO IDENTIFICADO	33419	CAFE 500G MACALI EXT.FORTE	1.0	14.99	2022.0	8.0	2.0	8
1	877033	0	CLIENTE NÃO IDENTIFICADO	19085	MANTEIGA 200G AVIACAO POTE	1.0	15.40	2022.0	8.0	2.0	8
2	877033	0	CLIENTE NÃO IDENTIFICADO	3723	ESMALTE RISQUE CARMIM	1.0	4.50	2022.0	8.0	2.0	8
3	877034	0	CLIENTE NÃO IDENTIFICADO	38753	SAB.85G JOHNSONS DAILY CARE ROSAS	1.0	1.99	2022.0	8.0	2.0	8
4	877034	0	CLIENTE NÃO IDENTIFICADO	38753	SAB.85G JOHNSONS DAILY CARE ROSAS	1.0	1.99	2022.0	8.0	2.0	8

Fonte: Elaborado pelos autores (2023)

Antes da implementação do algoritmo Apriori os dados da base foram separados por meses, na figura 7 pode ser visto como foi realizada a criação das tabelas relacionadas aos seus respectivos meses do ano dentro do período de agosto de 2022 a dezembro de 2022. Dessa forma foi possível aplicar o algoritmo e obter as regras de associação geradas.

Figura 7: Criação de tabelas 2022

```

#filtrando 2022
Ag22 = base[(base['Ano'] == 2022) & (base['Mes'] == 8)]
Agosto22 = Ag22[['codigo_venda', 'DESCRICAO_PRODUTO']]

Se22 = base[(base['Ano'] == 2022) & (base['Mes'] == 9)]
Setembro22 = Se22[['codigo_venda', 'DESCRICAO_PRODUTO']]

Ou22 = base[(base['Ano'] == 2022) & (base['Mes'] == 10)]
Outubro22 = Ou22[['codigo_venda', 'DESCRICAO_PRODUTO']]

No22 = base[(base['Ano'] == 2022) & (base['Mes'] == 11)]
Novembro22 = No22[['codigo_venda', 'DESCRICAO_PRODUTO']]

De22 = base[(base['Ano'] == 2022) & (base['Mes'] == 12)]
Dezembro22 = De22[['codigo_venda', 'DESCRICAO_PRODUTO']]

```

Fonte: Elaborado pelos autores (2023)

Para continuação do desenvolvimento do código, os dados do ano de 2023 também foram separados para que as tabelas de cada mês fossem criadas, janeiro de 2023 a agosto de 2023, conforme demonstrado pela figura 8.

Figura 8: Criação de tabelas 2023

```

# 2023
Ja23 = base[(base['Ano'] == 2023) & (base['Mes'] == 1)]
Janeiro23 = Ja23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Fe23 = base[(base['Ano'] == 2023) & (base['Mes'] == 2)]
Fevereiro23 = Fe23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Ma23 = base[(base['Ano'] == 2023) & (base['Mes'] == 3)]
Marco23 = Ma23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Ab23 = base[(base['Ano'] == 2023) & (base['Mes'] == 4)]
Abril23 = Ab23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Ma23 = base[(base['Ano'] == 2023) & (base['Mes'] == 5)]
Maio23 = Ma23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Ju23 = base[(base['Ano'] == 2023) & (base['Mes'] == 6)]
Junho23 = Ju23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Jl23 = base[(base['Ano'] == 2023) & (base['Mes'] == 7)]
Julho23 = Jl23[['codigo_venda', 'DESCRICAO_PRODUTO']]

Ag23 = base[(base['Ano'] == 2023) & (base['Mes'] == 8)]
Agosto23 = Ag23[['codigo_venda', 'DESCRICAO_PRODUTO']]

```

Fonte: Elaborado pelos autores (2023)

Ao modificar a estrutura da base de dados, para seguir o padrão de execução desse aprendizado de máquina, foram selecionadas as colunas “codigo_venda” e “Descricao_Produto” criando, assim, um *Data Frame* relacionado ao mês em questão. Os próximos passos são iguais para as 13 tabelas geradas para separação dos meses, no exemplo exibido pela figura 9, o mês utilizado foi “Agosto23”.

Por meio da técnica de pré-processamento *One-Hot Encoding* foi feita reorganização das colunas para colocar “codigo_venda” como a primeira coluna e os

produtos como colunas seguintes, logo após, foi feita a substituição dos valores para “1” quando o produto está presente na transação e “0” quando não está presente. Os valores da base foram padronizados como booleanos para prosseguir com o processamento.

Figura 9: Gerando *transactions*

```
df = pd.DataFrame(Agosto23)

transactions = pd.get_dummies(df['DESCRICAO_PRODUTO'])
transactions['codigo_venda'] = df['codigo_venda']
transactions = transactions[['codigo_venda'] + list(transactions.columns[:-1])]
transactions = transactions.groupby('codigo_venda').sum()
transactions = transactions.applymap(lambda x: 1 if x > 1 else x)
```

Fonte: Elaborado pelos autores (2023)

Dessa forma a variável “*transactions*” retornou a tabela com todas as transações de compras realizadas, em que o “*código_venda*” se situa nas linhas, enquanto os produtos estão localizados nas colunas.

Figura 10: tabela 0 e 1

	00107 - PINO TRES SAIDAS 2P + T 10A/250V	00539 - ALICATE CORTE DIAGONAL 5 POLEG. FOXLUX	A. C MOIDA	A. COST	A. CT FILE PRIMEIRA	A. CT LAO	A. CX DURO	A. FGO COX PRO	A. SUINO PICADAO	A. C PEDACO	...	WAFER 80G PASSION LIMAO C/CHOC.BRANCO	WRAP TA PRONTO WICKBOLD 270G	WRAP TA PRONTO WICKBOLD 270G INTEGRAL	YPI 500M MULTI USC ANTIBAC
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
...
22596	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
22597	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
22598	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

Fonte: Elaborado pelos autores (2023)

Dando início ao processo de associação dos dados, fez-se necessário a definição do suporte mínimo limiar com o valor de 0.07, significando que o algoritmo Apriori só considerará conjuntos de itens que aparecem em pelo menos 7% das transações. Com a variável “*dados_relacao*” foi feita a conversão dos valores para *True*, se o valor for maior ou igual ao suporte mínimo e *False* caso contrário.

Para encontrar conjuntos frequentes a variável “*frequencia_itens*” atuou junto ao Apriori recebendo o agrupamento de dados booleanos para obter os conjuntos de dados frequentes conforme apresentado na Figura 11.

Figura 11: Aplicando Apriori

```
min_support_threshold = 0.07

dados_relacao = dados_relacao >= min_support_threshold

frequencia_itens = apriori(dados_relacao, min_support=0.07, use_colnames=True)

✓ 0.7s
```

Fonte: Elaborado pelos autores (2023)

Após identificar conjuntos frequentes de itens no conjunto de dados, “dados_relacao”, foram derivadas regras de associação. A função “association_rules” foi utilizada para gerar essas regras a partir do conjunto de itens frequentes “frequencia_itens”. Na métrica *Lift* foi selecionado para definir a probabilidade de os itens em uma regra de associação serem adquiridos em conjunto, em comparação com a probabilidade de serem adquiridos individualmente. Um valor de *Lift* superior a 1 indica uma associação acima do esperado. Além disso, estabeleceu-se um “min_threshold” de 1, o qual representa o valor mínimo para o *Lift*, com isso, somente as regras com um valor superior a 1 foram incluídas nos resultados obtidos. Esse procedimento ocorreu conforme a Figura 12.

Figura 12: Frequência

```
#gerando itens frequentes
frequencia_itens = apriori(dados_relacao, min_support=0.01, use_colnames=True)

✓ 0.3s

#gerando as regras
regras = association_rules(frequencia_itens, metric="lift", min_threshold=1)

✓ 0.0s
```

Fonte: Elaborado pelos autores (2023)

Após a execução dos códigos, a variável “regras” retornou um *Data Frame* com regras de associação geradas entre os produtos do supermercado em questão para o mês de agosto de 2023.

Figura 13: Geração das regras

regras										
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(ALHO KG)	(CEBOLA CRUA SOLTA KG)	0.030383	0.062294	0.015528	0.511066	8.204145	0.013635	1.917860	0.905626
1	(CEBOLA CRUA SOLTA KG)	(ALHO KG)	0.062294	0.030383	0.015528	0.249264	8.204145	0.013635	1.291556	0.936445
2	(ALHO KG)	(TOMATE KG)	0.030383	0.083079	0.012593	0.414487	4.989093	0.010069	1.566013	0.824617
3	(TOMATE KG)	(ALHO KG)	0.083079	0.030383	0.012593	0.151582	4.989093	0.010069	1.142853	0.872008
4	(BANANA NANICA KG)	(BATATA CRUA SOLTA KG)	0.082834	0.042854	0.011065	0.133579	3.117105	0.007515	1.104713	0.740531
5	(BATATA CRUA SOLTA KG)	(BANANA NANICA KG)	0.042854	0.082834	0.011065	0.258203	3.117105	0.007515	1.236410	0.709598
6	(CEBOLA CRUA SOLTA KG)	(BANANA NANICA KG)	0.062294	0.082834	0.014855	0.238469	2.878876	0.009695	1.204371	0.695999
7	(BANANA NANICA KG)	(CEBOLA CRUA SOLTA KG)	0.082834	0.062294	0.014855	0.179336	2.878876	0.009695	1.142619	0.711586
8	(LARANJA KG)	(BANANA NANICA KG)	0.029282	0.082834	0.013266	0.453027	5.469091	0.010840	1.676803	0.841804
9	(BANANA NANICA KG)	(LARANJA KG)	0.082834	0.029282	0.013266	0.160148	5.469091	0.010840	1.155819	0.890956
10	(BANANA NANICA KG)	(MACA KG)	0.082834	0.037719	0.016934	0.204428	5.419828	0.013809	1.209547	0.889144
11	(MACA KG)	(BANANA NANICA KG)	0.037719	0.082834	0.016934	0.148947	5.419828	0.013809	1.554286	0.847457

Fonte: Elaborado pelos autores (2023)

Após observar as regras geradas nos dados, notou-se a presença de padrões coerentes entre os produtos antecedentes e consequentes desse período que obtiveram um valor médio de 65% de confiança. A análise mensal das regras resultou em *insights* significativos, ao observar as cinco regras de associações que obtiveram maior confiança de cada mês, notou-se a combinação recorrente de batata, cebola e tomate que ocorreu em dez transações ao longo do ano. Além disso, a combinação entre mortadela e pão francês foi registrada em sete transações, assim como o presunto e a muçarela.

Entretanto, algumas combinações diferentes foram identificadas com confianças maiores que a média registrada, como maçã com banana e suco Tang uva com suco Tang laranja que apareceram poucas vezes.

Essa descoberta foi capaz de oferecer oportunidades de estratégias de marketing para influenciar no lucro do estabelecimento, ao pensar na implementação de promoções concentradas nos produtos que surgiram como padrões ao longo do período estudado. Notou-se que alguns produtos possuíam uma demanda significativamente alta, principalmente nos momentos de pico, que normalmente ocorrem entre os dias 5 e 7 de cada mês. Por essa razão, posicioná-los estrategicamente próximos uns aos outros dentro do estabelecimento acarretará um melhor desempenho de transações.

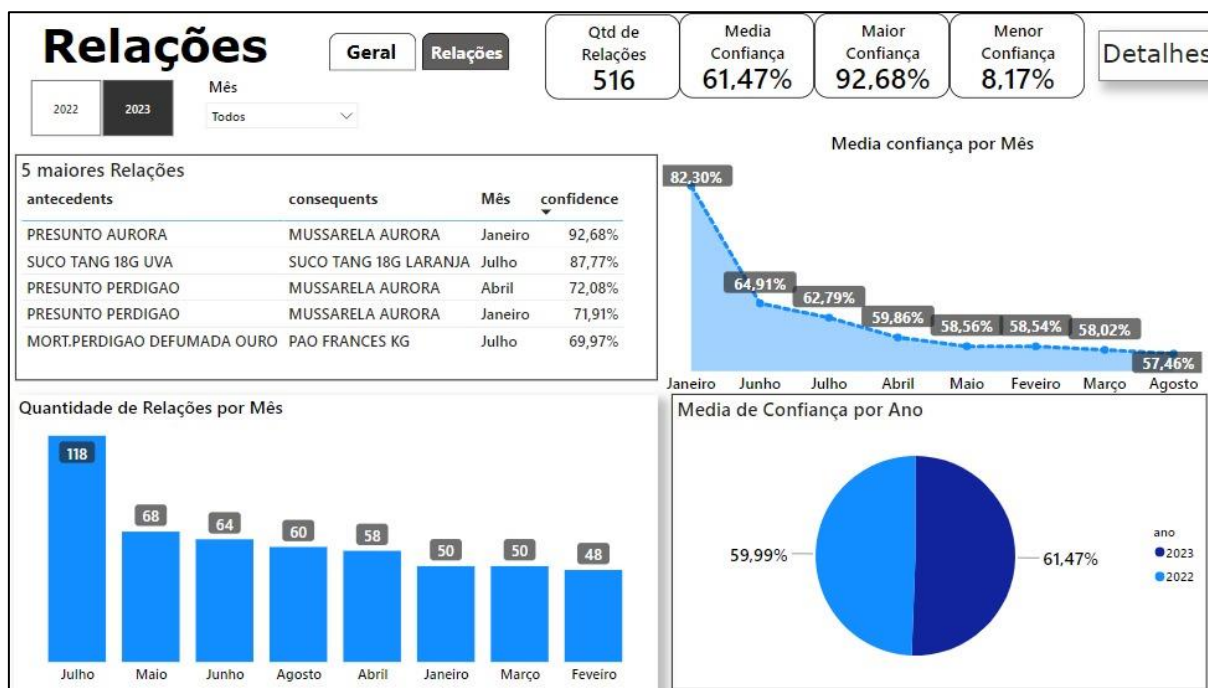
Para a visualização mais detalhada da base de dados, o Microsoft Power Bi auxiliou na geração de gráficos e tabelas interativas para a interpretação das informações que não seriam perceptíveis na tabela em sua forma bruta. No processo de construção do *dashboard*, iniciou-se com o desenvolvimento das medidas que são essenciais para o processo de geração dos indicadores

As medidas são fórmulas na linguagem DAX, esta é a linguagem padrão de cálculos da plataforma. Foram criadas as seguintes medidas: “maximo” e “menor”, que retorna o maior e menor valor de confiança da base de dados, “qtd de relações”, é a quantidade total de relações entre os produtos vendidos, “total de Confiança”, se caracteriza pela soma total dos valores de confiança.

A interface dinâmica apresenta inicialmente uma tabela e três gráficos na tela. Por meio da seleção do ano 2022 ou 2023 a tabela exhibe as cinco maiores relações entre os produtos antecedentes e consequentes, quando relacionados a confiança da regra, caso o cliente deseje visualizar as confianças de determinado mês basta selecionar a opção pretendida. O gráfico de linha “Média confiança por Mês” demonstra de forma contínua as médias das confianças, enquanto o gráfico intitulado “Quantidade de Relações por Mês” numera quantas regras de associações foram geradas no mês em questão por meio de um gráfico de barras. Além disso, o gráfico em formato de pizza denominado “Média de Confiança por Ano” apresenta a média das confianças ao longo do ano selecionado, podendo ser neste caso, 2022 e 2023.

O botão “Detalhes” direciona o cliente a outra página que apresenta os dados da base em formato de tabela, já o botão “Geral” leva o cliente à tela onde as demais informações estão localizadas, conforme demonstram as Figuras 14 e 15.

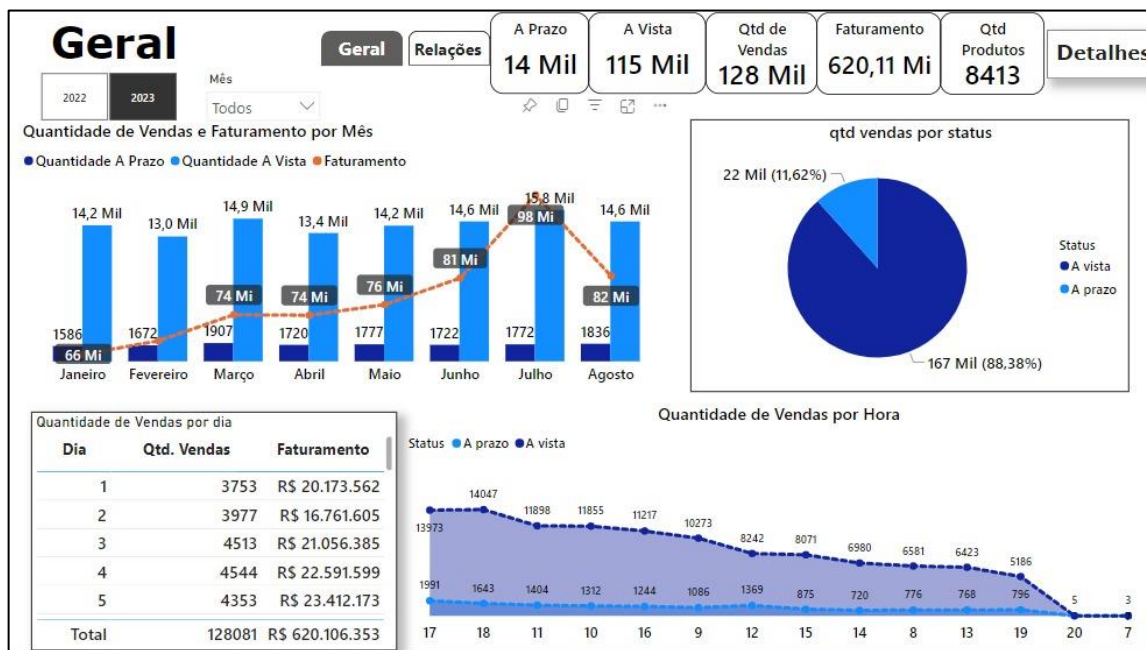
Figura 14: Relações das regras de associações



Fonte: Elaborado pelos autores (2023)

A tela “Geral” disponibiliza informações para os clientes ao selecionarem os anos de 2022 ou 2023, relacionados aos valores adquiridos e às quantidades de vendas. As informações apresentadas podem ser separadas por mês, dias e horas. Por meio dessa interface é possível visualizar os momentos de pico nas vendas do supermercado, proporcionando uma análise detalhada do desempenho comercial.

Figura 15: Informações Gerais



Fonte: Elaborado pelos autores (2023)

Pensando na facilidade de interpretação do cliente, foi desenvolvida uma interface *Front-end*, utilizando o *software* Microsoft Power Apps, permitindo que o usuário possa acessar todo o conjunto de informações geradas a partir dos dados de seu estabelecimento, essas informações apresentam *insights* e indicadores que irão impulsionar o varejo. Nessa aplicação as telas adicionadas ilustram outras possíveis funções para novos usuários.

Para a visualização da aplicação apresentada o usuário deverá ter acesso as plataformas SharePoint, Power Apps e Power Bi, caso contrário, deverá solicitar acesso aos responsáveis. O aplicativo se encontra no link <https://apps.powerapps.com/play/e/default-cf72e2bd-7a2b-4783-bdeb-39d57b07f76f/a/857fe5b0-18b8-4062-9baa-631d06c0dbe0?tenantId=cf72e2bd-7a2b-4783-bdeb-39d57b07f76f&hint=51d7c12b-ab84-40c4-8cd1-ae1bc6764f2b&sourcetime=1698106680022&source=portal>.

A Figura 15 representa a interface inicial da aplicação, que serve como a página inicial para os usuários. Nesta tela, é claramente exibido o objetivo central do trabalho realizado, ao oferecer para os usuários uma visão imediata do propósito e funcionalidade do aplicativo.

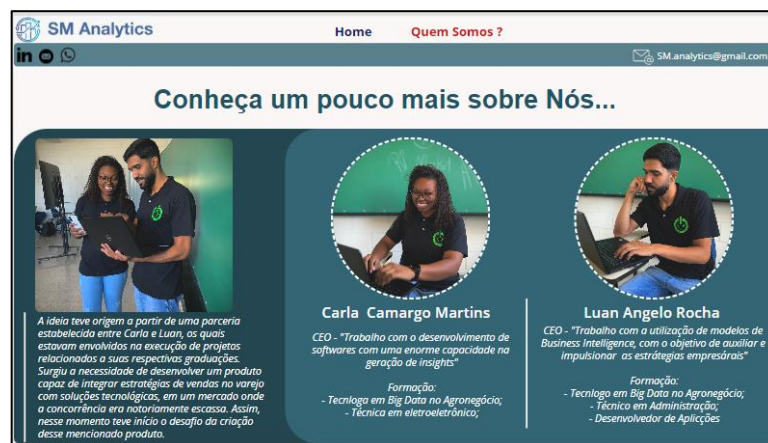
Figura 15: Home



Fonte: Elaborado pelos autores (2023)

A seção “Quem Somos” demonstrada na Figura 16, oferece aos usuários uma visão pessoal dos desenvolvedores por trás do projeto. Na aplicação, essa tela exibe as imagens da equipe, com informações relacionadas as suas certificações e qualificações, junto com o processo que levou a criação do projeto.

Figura 16: Quem Somos



Fonte: Elaborado pelos autores (2023)

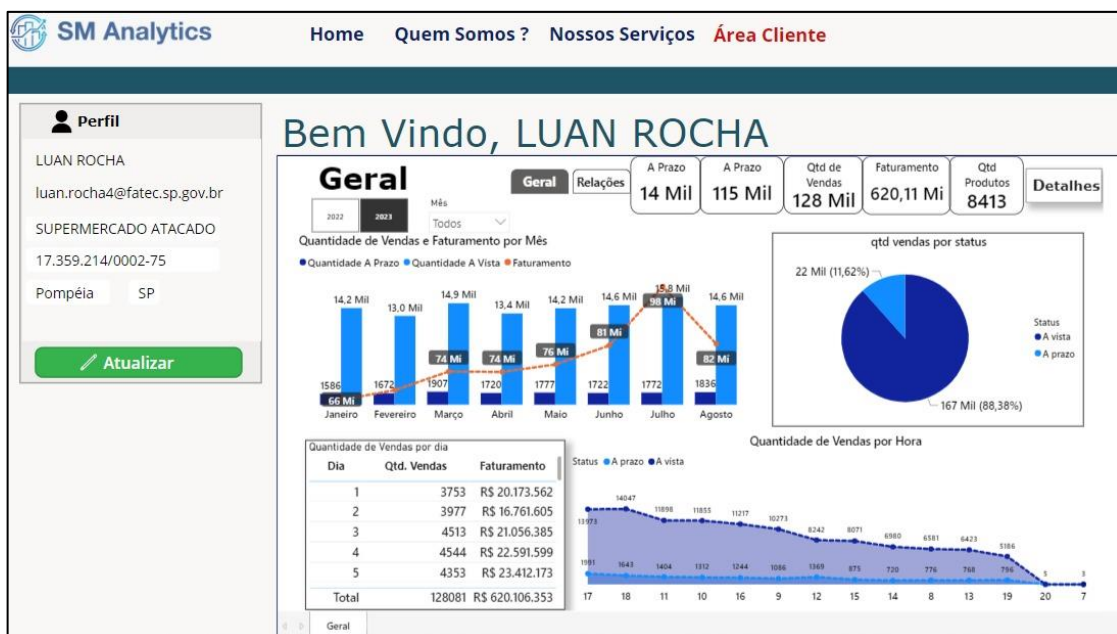
A imagem abaixo demonstra a área “Saiba Mais”, permite que o cliente obtenha informações de como entrar em contato com a empresa e, a partir dos campos disponíveis, insira os dados pessoais e os dados do estabelecimento para que sejam analisados.

Figura 17: Saiba Mais

Fonte: Elaborado pelos autores (2023)

A Figura 18 representada como “Área Cliente” permite que o usuário realize o *login* no seu perfil pessoal e tenha acesso a todos os dados e materiais gerados para ter conhecimento e auxiliar nas tomadas de decisões do estabelecimento.

Figura 18: Área Cliente



Fonte: Elaborado pelos autores (2023)

As telas criadas para a aplicação desempenham um papel importante ao proporcionar para os clientes uma experiência dinâmica e fácil de acessar as tabelas. Por meio das interfaces intuitivas, os usuários podem absorver *insights* valiosos de

forma eficaz, permitindo uma análise aprofundada e uma compreensão abrangente dos dados apresentados como resultado do algoritmo aplicado.

CONCLUSÃO

Diante do trabalho realizado, a aplicação do algoritmo Apriori permitiu identificar padrões de compras no supermercado na cidade de Queiroz. Por meio do Suporte, Confiança e *Lift*, combinações recorrentes entre os grupos de produtos cebola, batata e tomate, mortadela e pão francês, presunto e muçarela. Entretanto, notou-se que algumas combinações diferentes foram identificadas comprovando a variedade do supermercado.

A análise da base de dados foi realizada utilizando o BI em conjunto com o Microsoft Power BI, permitindo a geração de gráficos e tabelas dinâmicas para compreensão dos resultados. Ademais, foi desenvolvido um aplicativo com a plataforma Power Apps, concedendo aos usuários acesso aos *insights* obtidos, capazes de incentivar estratégias de marketing e impulsionar os lucros do supermercado.

Com o aprimoramento do projeto podem ser realizadas alterações na aplicação do algoritmo Apriori, a fim de aumentar a precisão na confiança das associações entre produtos. Visando ir além dos supermercados, pretende-se explorar diferentes setores e contextos, para uma compreensão mais abrangente dos padrões de comportamento do consumidor.

REFERÊNCIAS

BOENTE, ALFREDO NAZARENO PEREIRA; OLIVEIRA, FABIANO SALDANHA GOMES DE; ROSA, JOSÉ LUIZ DOS ANJOS; **Utilização de Ferramentas de KDD para Integração de Aprendizagem e Tecnologia em Busca da Gestão Estratégica do Conhecimento na Empresa.** Disponível em: <https://www.aedb.br/seget/arquivos/artigos07/1219_Artigo%20SEGET%202007.pdf> Acesso em: 10. ago. 2023.

BRANQUINHO, LUCÉLIA PINTO; **Modelo Para Suporte À Descoberta De Conhecimento Em Base De Dados (Kdd): Aplicação Em Estratégias No Mercado De Medicina Diagnóstica.** 2015. Disponível em: <https://repositorio.ufmg.br/bitstream/1843/BUBD-A8SJWL/1/dissertacao_lucelia_pinto_branquinho.pdf> Acesso em: 05. set. 2023.

CHEN, GUOQING; CHEN, HSINCHUN; LIM, EE-PENG **Business Intelligence and Analytics: Research Directions.** 2013. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/2407740.2407741>> Acesso em: 15. out. 2023.

FROEHLICH, ELIAS SIEDEKUM; **APLICAÇÃO DO DATA MINING EM UMA LOJA DE CONVENIÊNCIA, COMO SUPORTE À TOMADA DE DECISÃO**. 2020. Disponível em:

<<http://repositorio.jesuita.org.br/bitstream/handle/UNISINOS/11620/Elias%20Siedekum%20Froehlich.pdf?sequence=1&isAllowed=y>> Acesso em: 12. out. 2023.

HANASHIRO, AKIRA; **VS Code - O que é e por que você deve usar?**. 2023.

Disponível em: <<https://www.treinaweb.com.br/blog/vs-code-o-que-e-e-por-que-voce-deve-usar>> Acesso em: 02. out. 2023.

MELANDA, EDSON AUGUSTO; REZENDE, SOLANGE OLIVEIRA; **Sintaxe Padrão para Representar Regras de Associação**. 2003. Disponível em: <

https://web.icmc.usp.br/SCATUSU/RT/Relatorios_Tecnicos/BIBLIOTECA_113_RT_2_06.pdf> Acesso em: 20. ago. 2023.

MICROSOFT POWER APPS. **Documentação do Power Apps**. Disponível em:

<<https://learn.microsoft.com/pt-br/power-apps/>> Acesso em: 15. out. 2023.

MICROSOFT POWER BI. **Documentação do Power BI**. Disponível em:

<https://learn.microsoft.com/pt-br/power-bi/>> Acesso em: 15. out. 2023.

MIRANDA, JOÃO VITOR; **get_dummies vs OneHotEncoder: qual método escolher?**. 2023. Disponível em:<

<https://www.alura.com.br/artigos/get-dummies-vs-onehotencoder-qual-metodo-escolher>> Acesso em: 19. ago. 2023.

MOCHIZUKI, YU; **Apyori**. 2019. Disponível em: <https://pypi.org/project/apyori/> >

Acesso em: 02. out. 2023.

MULINARI, Bruna; **Numpy Python: O que é, vantagens e tutorial inicial**. 2023.

Disponível em: <https://harve.com.br/blog/programacao-python-blog/numpy-python-o-que-e-vantagens-e-tutorial-inicial/>> Acesso em: 15. out. 2023.

SANTOS, WELLINGTON PINHEIRO DOS; MORENO, GISELLE MACHADO MAGALHÃES; YARA, RICARDO; **Anais do VI Simposio de Inovacao em Engenharia Biomedica - SABIO 2022**. 2022. Disponível em:

https://www.researchgate.net/profile/Wellington-Dos-Santos/publication/369591716_Anais_do_VI_Simposio_de_Inovacao_em_Engenharia_Biomedica_-_SABIO_2022/links/64239e2b92cfd54f84388297/Anais-do-VI-Simposio-de-Inovacao-em-Engenharia-Biomedica-SABIO-2022.pdf#page=23 >

Acesso em: 01. out. 2023.

SUPERMERCADOS, ASSOCIAÇÃO BRASILEIRA DE; **ABRAS**. 2021. Disponível em

:< <https://www.abras.com.br/> > Acesso em: 02. out. 2023.

TANURE, CARLA MOREIRA; **Uma Arquitetura de Software para Descoberta de Regras de Associação Multidimensional, Multinível e de Outliers em Cubos OLAP: Um Estudo de Caso com os Algoritmos Apriori e FP - Growth**. 2010. Disponível em:

<https://repositorio.ufpe.br/bitstream/123456789/2207/1/arquivo2236_1.pdf> Acesso em: 20. ago. 2023.

VASCONCELOS, LÍVIA MARIA ROCHA; CARVALHO, CEDRIC LUIZ DE; **Aplicação de Regras de Associação para Mineração de Dados na Web**. 2004. Disponível em:< <https://revista.telematicafactal.com.br/index.php/telfract/article/view/8/5> >
Acesso em: 10. ago. 2023.