

Maria Vitória Reiter Grandi
mgrandi@alunos.utfpr.edu.br
<http://orcid.org/0000-0002-6383-535>
Universidade Tecnológica Federal do
Paraná, Medianeira, Paraná

Mineração de dados para comparação dos padrões de compras em uma rede de supermercados do oeste do Paraná antes e durante a pandemia de COVID-19

RESUMO

No Brasil, além das consequências trágicas que a pandemia de COVID-19 trouxe para a área da saúde, surgiram outros problemas graves, como o aumento do desemprego, aumentos sucessivos nos preços de diversos produtos e a diminuição da renda domiciliar per capita. Algumas pesquisas recentes indicam que esses problemas provocaram mudanças no perfil de consumo das famílias, porém são pesquisas bastante específicas e que empregaram de forma bastante restrita técnicas de mineração de dados. Neste sentido, este trabalho teve como objetivo aplicar a técnica de mineração de regras de associação para: a) identificar os padrões de compras em uma rede de supermercados da região oeste do Paraná durante o período anterior e durante a pandemia de COVID-19; e b) comparar os padrões obtidos para os dois períodos, a fim de possibilitar que os especialistas da empresa pudessem analisar se as diferenças foram inesperadas e relevantes sob o ponto de vista da administração das lojas da rede. Como resultados, o algoritmo Apriori foi aplicado utilizando diferentes valores de suporte e confiança para cada um dos municípios analisados a fim de obter as 30 melhores regras de associação em cada um deles. Por meio destes resultados foi possível identificar que durante o período da pandemia houve uma alteração no padrão de compras dos clientes em relação aos anos de 2018 e 2019, o que para os especialistas da rede de supermercados já era esperado. Porém, ainda segundo os especialistas, houve descobertas importantes para a empresa com a realização deste trabalho, como as alterações dos padrões de compras específicas de cada município durante os anos de 2020 e 2021.

PALAVRAS-CHAVE: Regras de associação, algoritmo Apriori, big data, análise de cestas de compras.

INTRODUÇÃO

O ano de 2020 foi marcado pelo surgimento da pandemia da Síndrome Respiratória Aguda Grave (SARS-CoV-2), denominada como COVID-19 (Coronavirus Disease 2019) pela Organização Mundial da Saúde. Uma das restrições impostas para o controle do avanço da doença foi o isolamento social, o que levou a mudanças drásticas na rotina da população mundial. Uma dessas mudanças no Brasil, que foi amplamente destacada pela imprensa, foi relacionada à frequência de ida das pessoas aos supermercados.

Durante o período pandêmico, também houve em nosso país um aumento expressivo da inflação, em comparação ao período pré-pandêmico. Mensalmente, o Instituto Brasileiro de Geografia e Estatística (IBGE) é responsável por fazer o cálculo do IPCA (Índice Nacional de Preços ao Consumidor Amplo), sendo ele o principal índice de inflação do país. O IPCA acumulado de 12 meses no período de outubro de 2020 a dezembro de 2021 foi de 10,06% (IBGE, 2021), enquanto no mesmo período dos anos de 2018 e 2019 o valor obtido foi de 4,31% (IBGE, 2019).

As tecnologias da informação continuam a se desenvolver rapidamente nos dias de hoje e as empresas podem obter, armazenar, analisar e interpretar os dados de seu interesse com muito mais facilidade e a um custo menor (ÜNVAN, 2021). As técnicas de mineração de dados têm sido aplicadas em diferentes áreas do conhecimento, como na análise de cestas de compras de supermercados (FACELLI et al., 2011). Esse tipo de análise possibilita determinar o comportamento de compras dos clientes através da utilização de regras de associação nos conjuntos de dados obtidos. Dentre várias possibilidades, um tipo de análise que poderia ser feita sobre regras de associação mineradas nesse domínio é em relação à mudança significativa de padrões de compras, sob influência de um ou mais fatores relevantes.

Já haviam sido publicados trabalhos correlatos que avaliaram o comportamento alimentar das pessoas durante o período da pandemia de COVID-19, como Rodrigues et al. (2021) e Tribst et al. (2021). Porém, nenhum destes trabalhos utilizou a mineração de regras de associação para a descoberta de conhecimento e a comparação com um período anterior à pandemia.

METODOLOGIA

Definiu-se a utilização de dados de vendas de produtos de supermercado de um período de 4 anos, com intuito de incluir transações de anos anteriores e durante a pandemia de COVID-19. Para isso, estabeleceu-se uma parceria com uma rede de supermercados da região oeste do Paraná, que possui 16 lojas físicas em 11 municípios e está sediada no município de Medianeira – Paraná. Na Tabela 1, estão disponíveis as informações sobre a população estimada (IBGE, 2021d) e o número de lojas dos municípios que possuem as unidades da rede de supermercados.

Tabela 1 – Número de lojas e população estimada de cada município analisado

Municípios	População Estimada (2021)	Nº de Lojas
Céu Azul	11.872	1
Diamante d'Oeste	5.279	1
Itaipulândia	11.588	1
Matelândia	18.266	2
Medianeira	46.940	3
Missal	10.706	1
Ramilândia	4.500	1
Santa Helena	27.036	2
Santa Terezinha de Itaipu	23.927	1
São Miguel do Iguaçu	27.696	2
Serranópolis do Iguaçu	4.460	1

Fonte: Autoria própria

Os dados armazenados na planilha referem-se aos atributos *id* (identificador único de cada venda realizada), *data da venda*, *horário da venda*, *filial* que realizou a venda, *cidade*, *categoria do produto*, *subcategoria do produto*, *nome do produto*, *código de barras*, *preço unitário*, *quantidade vendida*, *desconto*, *valor total da venda do produto*, *valor líquido* (valor referente a subtração do valor total do produto com o desconto), *sexo*, *tipo da pessoa* (física ou jurídica) e *tipo da venda*.

O conjunto de dados utilizado possui 138.367.792 transações de vendas, com um tamanho total de 29,8 GB, correspondendo a um histórico de dados do período de janeiro de 2018 até dezembro de 2021. Estes dados, que na empresa estão armazenados em bancos de dados mantidos com o uso do sistema gerenciador Microsoft SQL Server, foram disponibilizados para a utilização neste trabalho em uma planilha, em um arquivo com extensão csv (*Comma Separated Values*).

Para que os dados ficassem no formato adequado para a mineração de dados foi necessário dividir o conjunto de dados em cada ano de venda, assim como também separar as vendas em cada município analisado para avaliar, se houver, as mudanças no padrão de compra dos clientes ao longo dos anos. Esse processo foi realizado por meio de um algoritmo desenvolvido em Python que recebe como entrada os dados e filtra as colunas *data* e *cidade* através dos valores selecionados.

Também houve a necessidade de aplicar o mapeamento de valores, pois alguns valores de atributos não foram utilizados para a mineração. Primeiramente houve a eliminação das colunas *hora*, *CodigoBarra*, *Filial*, *Desconto*, *ValorTotal*, *valor_liquido*, *Sexo*, *PessoaTipo* e *TipoVenda*, pois as mesmas não possuíam informações relevantes para a aplicação do algoritmo de filtro descrito anteriormente. Em seguida, outro algoritmo em Python foi desenvolvido para agrupar os cupons que possuíam os mesmos ID para formar todos os itens que cada compra possuía. Como neste processo apenas as colunas *ID* e *Produto* eram necessárias, as colunas *data*, *Cidade*, *Categoria*, *SubCategoria*, *Preço* e *Quantidade* foram eliminadas.

A próxima etapa foi a aplicação das técnicas de mineração de regras de associação sobre o conjunto completo de transações de vendas, para descobrir quais conjuntos de produtos foram vendidos juntos, com maior frequência, com

interesse principalmente nas regras que não fossem consideradas óbvias pelos especialistas da empresa. A aplicação desta técnica foi realizada por meio de implementação usando a linguagem de programação Python, sendo empregadas as bibliotecas mlxtend, Numpy e Pandas, que auxiliaram na aplicação do algoritmo Apriori na ferramenta Google Colaboratory Pro.

Por fim, houve a análise e a discussão dos resultados que foram obtidos por meio da aplicação da técnica de mineração de regras de associação, contando com a participação de especialistas da rede de supermercados. Houve necessidade de realização de uma reunião com tais especialistas, inicialmente para lhes mostrar os resultados obtidos, e em seguida para que eles fossem capazes de identificar quais resultados eram considerados novos e relevantes, sob o ponto de vista da administração das lojas da rede.

RESULTADOS

O perfil de compras dos clientes de todas as unidades da rede de supermercados é muito parecido, destacando o consumo de produtos de conserva e enlatados, itens da cesta básica, como arroz, feijão, macarrão, farinha de trigo, café, molho de tomate, e óleo de soja, percebe-se também que é comum a compra da lentilha, outro produto não perecível.

É possível analisar que há um consumo de erva mate e linguiça nas compras dos clientes da região oeste, sendo comum esses tipos de compras, pois é uma tradição o consumo de erva mate para chimarrão ou tereré, e linguiça para se fazer churrasco. Por fim, os clientes da rede de supermercado consomem uma variedade de tipos de carnes, assim como também produtos embutidos, como por exemplo presunto, mortadela, salsicha, hambúrguer e apresuntado.

Através da reunião realizada no dia 12 de maio de 2022 com os especialistas da rede de supermercados foi possível que os mesmos apresentassem seus comentários sobre os resultados obtidos a partir deste trabalho. Foi destacado que já era esperado que houvesse uma alteração nos padrões de compras dos seus clientes das 16 unidades, como o aumento nas compras de itens da cesta básica, produtos de limpeza e bebidas alcoólicas, assim como a diminuição no consumo de carne bovina e aumento na procura de carnes mais acessíveis, como carnes de frango e suínas durante a pandemia.

Entretanto, não se sabia como cada unidade se comportava diante dessas alterações. Este trabalho possibilitou analisar que cada município se comportou de forma distinta dos outros, tendo alguns que apresentaram alterações mais evidentes do que outros ao longo dos 4 anos analisados.

Nesta reunião também foi evidenciada a presença da regra de associação entre Cachepôs e Orquídeas em alguns municípios, como em Medianeira e Santa Helena. Nas unidades desses dois municípios, a localização destes produtos é mais visível para os clientes, assim que eles entram nas unidades. Portanto, essa localização possivelmente chama mais a atenção dos clientes para tais produtos. Nas análises ficou evidente que nestas unidades há uma venda maior desses produtos do que nas outras que não possuem essa mesma distribuição; logo, é interessante para a rede de supermercados adotar essa mesma estratégia nas outras unidades para que a venda conjunta desses produtos tenha possibilidade de aumentar também. Além disso, essa regra de associação permitiu recomendar

aos especialistas da empresa que observem outras situações similares nas unidades, ou seja, em que a disposição de certos produtos na loja possivelmente tem influenciado na venda conjunta deles.

CONCLUSÕES

Por meio deste trabalho foi possível identificar as alterações nos padrões de compras dos clientes da rede de supermercados durante o período da pandemia de COVID-19, utilizando o algoritmo Apriori para a mineração de regras de associação. Assim, identificou-se o aumento nas compras de itens da cesta básica, produtos de limpeza, bebidas alcoólicas e carnes mais acessíveis, como de frango e suína.

Para os especialistas da rede de supermercados, essas alterações encontradas já eram esperadas no panorama geral das 16 unidades, entretanto não eram de seu conhecimento as alterações particulares que cada município apresentou durante os anos de 2020 e 2021.

Data mining for comparisons of shopping patterns in a supermarket chain in western Paraná before and during the COVID-19 pandemic

ABSTRACT

In Brazil, in addition to the tragic consequences that the COVID-19 pandemic brought to the health area, other serious problems emerged, such as the increase in unemployment, successive increases in the prices of various products and the decrease in per capita household income. Some recent researches indicate that these problems provoked changes in the consumption profile of the families, however they are very specific researches and that they used in a very restricted way data mining techniques. In this sense, this work aimed to apply the association rule mining technique to: a) identify shopping patterns in a supermarket chain in the western region of Paraná during the previous period and during the COVID-19 pandemic; and b) compare the patterns obtained for the two periods, in order to enable the company's specialists to analyze whether the differences were unexpected and relevant from the point of view of the management of the chain's stores. As a result, the Apriori algorithm was applied using different support and confidence values for each of the analyzed municipalities in order to obtain the 30 best association rules in each of them. Through these results, it was possible to identify that during the period of the pandemic there was a change in the pattern of customer purchases in relation to the years 2018 and 2019, which for the specialists of the supermarket chain was already expected. However, according to the experts, there were important discoveries for the company with the completion of this work, such as changes in the specific purchasing patterns of each municipality during the years 2020 and 2021.

KEYWORDS: Association rules, Apriori algorithm, big data, shopping basket analysis.

REFERÊNCIAS

FACELLI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. de Leon Ferreira de. "Inteligência artificial: uma abordagem de aprendizado de máquina". Rio de Janeiro: LTC - Livros Técnicos e Científicos, 2011.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. "IPCA: Índice Nacional de Preços ao Consumidor Amplo". Rio de Janeiro: IBGE, 2019. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=downloads>. Acesso em: 25 nov. 2021.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. "IPCA: Índice Nacional de Preços ao Consumidor Amplo". Rio de Janeiro: IBGE, 2021. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=downloads>. Acesso em: 18 mai. 2022.

RODRIGUES, Jessica Ferreira; FILHO, Marcus Túlio Cunha dos Santos; OLIVEIRA, Lorena Eduarda Aparecida de; SIMAN, Ingrid Brandenburg; BARCELOS, Alessandrade Fátima; RAMOS, Gustavo Luis de Paiva Anciens; ESMERINO, Erick Almeida; CRUZ, Adriano Gomes da; ARRIEL, Rhaí André. Effect of the covid-19 pandemic on food habits and perceptions: A study with brazilians. "Trends in Food Science Technology", v. 116, p. 992–1001, 2021. ISSN 0924-2244. Disponível em: <https://www.sciencedirect.com/science/article/pii/S092422442100529X>.

TRIBST, Alline Artigiani Lima; TRAMONTT, Cláudia Raulino; BARALDI, Larissa Galastri. Factors associated with diet changes during the covid-19 pandemic period in brazilian adults: Time, skills, habits, feelings and beliefs. "Appetite", v. 163, p. 105220, 2021. ISSN 0195-6663. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0195666321001276>.

ÜNVAN, Yüksel Akay. Market basket analysis with association rules. "Communications in Statistics - Theory and Methods", Taylor Francis, v. 50, n. 7, p. 1615–1628, 2021. Disponível em: <https://doi.org/10.1080/03610926.2020.1716255>.

Recebido: 24/05/2022

Aprovado: 30/12/2023.

DOI: 10.3895/recit.v14i35.15532

Como citar: GRANDI, M. V. Mineração de dados para comparação dos padrões de compras em uma rede de supermercados do oeste do Paraná antes e durante a pandemia de covid-19. R. Eletr. Cient. Inov. Tecnol, Medianeira, v. 14. n. 35, p. 23- 31, set/dez 2023 Disponível em: <<https://periodicos.utfpr.edu.br/recit>>. Acesso em: XXX.

Correspondência:

Maria Vitória Reiter Grandi

Universidade Tecnológica Federal do Paraná, Medianeira, Paraná.]

Direito autoral: Este artigo está licenciado sob os termos da Licença creativecommons.org/licenses/by-nc/4.0/ Internacional.

