

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE GRADUAÇÃO
CURSO DE ENGENHARIA DE PRODUÇÃO**

ELIAS SIEDEKUM FROEHLICH

**APLICAÇÃO DO DATA MINING EM UMA LOJA DE CONVENIÊNCIA, COMO
SUPORTE À TOMADA DE DECISÃO**

São Leopoldo

2020

ELIAS SIEDEKUM FROEHLICH

**APLICAÇÃO DO DATA MINING EM UMA LOJA DE CONVENIÊNCIA, COMO
SUPORTE À TOMADA DE DECISÃO**

Trabalho de Conclusão de Curso
apresentado como requisito parcial para
obtenção do título de Bacharel em
Engenharia de Produção, pelo Curso de
Engenharia de Produção da Universidade
do Vale do Rio dos Sinos - UNISINOS

Orientador: Prof. Ms. Marcos Leandro Hoffmann Souza

São Leopoldo

2020

Dedico esse trabalho à minha família.

AGRADECIMENTOS

Esta é a última parte ao qual escrevo, e ao olhar para trás, passa um breve filme e relembro aqueles que tiveram um papel fundamental ao longo desta caminhada, e aos quais gostaria de agradecer.

Primeiramente aos meus pais, pelo amor, apoio e dedicação de sempre. Obrigado por me ensinarem o valor dos estudos, por acreditarem na minha capacidade e pelo incentivo na busca constante de evolução.

À minha irmã Milena, pelo apoio sempre prestado e por sempre me incentivar na busca dos sonhos, “tamo junto”.

Ao grande professor e meu orientador neste trabalho, Marcos Hoffmann. Muito obrigado pela confiança depositada, pelo apoio e por ter aceitado, 2 anos atrás, me orientar no desenvolvimento deste trabalho. Cada troca de ideias contigo foi um novo aprendizado neste meu processo de evolução e transição. Saibas que a partir das tuas aulas, consegui encontrar e me orientar àquilo que desejo profissionalmente. Gratidão aos ensinamentos.

Aos meus amigos, que souberam compreender minha ausência ao longo deste ano, mas que estão sempre de braços abertos para me receber.

Aos colegas de curso, Eric Mezzomo, Tomas Luft e Rodrigo Jacob, grato pela amizade e trocas de experiências. Roberto de Mattos e Franciele Laura, vocês foram fundamentais ao longo desta graduação, obrigado pelo companheirismo e apoio na busca constante da evolução acadêmica e profissional.

Aos professores Jayme Peixoto e Leandro Gauss, pelas conversas e dicas repassadas. Ficam aqui minhas desculpas pela importunidade naqueles minutos pós aula, mesmo sendo breves debates, sempre me proporcionaram maior conhecimento e incentivaram ao aprofundamento das questões.

Deixo o meu “muito obrigado” a todos, que de alguma forma contribuíram na minha formação.

“Não é a espécie mais forte que sobrevive, nem a mais inteligente, mas aquela que é mais adaptável à mudança”.

Charles Darwin

RESUMO

As empresas vêm gerando e acumulando muitos dados, e a atual transformação da sociedade exige que estas organizações aprendam a lidar constantemente com os dados. Dados existem sem informações, mas informações só existem por meio dos dados, e o tipo de relação que uma organização estabelece com estas informações é fundamental para compreender o comportamento de seus clientes. Por meio do *Data Mining* é possível extrair informações a partir da modelagem e análise dos dados, gerando *insights* do dia a dia do negócio para uma melhor decisão dos gestores de uma organização. Baseando-se nisso, este trabalho busca encontrar na literatura um algoritmo capaz de realizar a tarefa de *Market Basket Analysis*, objetivando a geração de regras de associação a partir de uma base de dados de transações de venda de uma loja de conveniência. Posteriormente, estas regras geradas são relacionadas aos dias da semana, originando informações úteis do comportamento diário dos clientes no ponto de venda. Para que as etapas da mineração de dados sejam atingidas, o trabalho segue a metodologia CRISP-DM, expondo desde a compreensão do negócio até a avaliação dos resultados. Com a aplicação das etapas da metodologia e do modelo, foi possível descobrir quais os produtos que tendem a ser vendidos conjuntamente em determinados períodos da semana, nos quais estes resultados trazem insights para suportar as tomadas de decisões dos responsáveis do negócio na busca da diferenciação frente à concorrência. O estudo comprova que a metodologia e a ferramenta utilizada podem ser importantes meios para, no atual mercado competitivo, gerar informações de grande importância às organizações.

Palavras-chave: *Data Mining*. Regras de Associação. *Market Basket Analysis*. Loja de Conveniência.

ABSTRACT

Companies have been generating and accumulating a lot of data, and the current transformation of society requires that these organizations learn to constantly deal with data. Data exists without information, but information only exists through data, and the type of relationship that an organization establishes with this information is fundamental to understanding the behavior of its customers. Through Data Mining it is possible to extract information from the modeling and analysis of data, generating insights from the day-to-day business for a better decision by the managers of an organization. Based on this, this work seeks to find in the literature an algorithm capable of performing the task of Market Basket Analysis, aiming at the generation of association rules from a database of sales transactions in a convenience store. Subsequently, these generated rules are related to the days of the week, providing useful information on the daily behavior of customers at the point of sale. In order for the data mining stages to be achieved, the work follows the CRISP-DM methodology, exposing everything from understanding the business to evaluating the results. With the application of the steps of the methodology and the model, it was possible to discover which products tend to be sold together in certain periods of the week, in which these results bring insights to support the decision-making of those responsible for the business in the search for differentiation in front competition. The study proves that the methodology and the tool used can be important means to, in the current competitive market, generate information of great importance to organizations.

Key-Words: Data Mining. Association rules. Market Basket Analysis. Convenience store.

LISTA DE FIGURAS

Figura 1 – Busca de Referências	21
Figura 2 - Mapa conceitual do escopo do referencial teórico	33
Figura 3 - Etapas do KDD	43
Figura 4 - Ciclo Virtuoso da Mineração de Dados	44
Figura 5 - Árvore de Categoria de Produto	47
Figura 6 - Análise de Cesta de Compras	50
Figura 7 - Etapas de Busca por Algoritmos de KDD	55
Figura 8 - Métodos Científicos.....	59
Figura 9 - Classificações de pesquisa.....	60
Figura 10 - Método CRISP-DM	63
Figura 11 - Dados Originais.....	71
Figura 12 - Níveis de Categorias de Produto	74
Figura 13 - Dados Tratados.....	75
Figura 14 - Construção do Modelo	76
Figura 15 - Validação do Modelo.....	79
Figura 16 - Comando <i>summary(transactions)</i>	80
Figura 17 - Comando <i>summary(rules)</i>	80
Figura 18 - Comandos iniciais.....	82
Figura 19 - Comando Finais	84
Figura 20 - Resultado do comando <i>summary(rules)</i>	84
Figura 21 - Comando <i>inspect(rules_high_lift)</i>	85

LISTA DE GRÁFICOS

Gráfico 1 - Evolução da quantidade de lojas.....	26
Gráfico 2 - Faturamento do mercado total (R\$ Bilhões/Ano).....	27
Gráfico 3 - Algoritmos encontrados na pesquisa.....	56
Gráfico 4 - Cinco itens mais frequentes	83
Gráfico 5 – Associações geradas.....	85
Gráfico 6 - Volume de vendas em decorrência dos dias	90

LISTA DE QUADROS

Quadro 1 - Protocolo para a Revisão Sistemática da Literatura.....	20
Quadro 2 – Trabalhos incluídos	24
Quadro 3 – Características de Lojas de Conveniência	37
Quadro 4 - Transações de uma loja de conveniência	53
Quadro 5 - Classificação dos objetivos da pesquisa	61
Quadro 6 - Métodos de Pesquisa.....	62
Quadro 7 - Amostra das Regras de Associação Geradas em Teste.....	81
Quadro 8 - Resultados das principais regras	83
Quadro 9 - Regras de Associação	86
Quadro 10 - Períodos	91
Quadro 11 - Regra de Associação comparada aos períodos.....	92

LISTA DE TABELAS

Tabela 1 - Regras de Associação x Dia da Semana	89
Tabela 2 - Associação mais vendida em cada dia	90
Tabela 3 - Combos x Turnos	95

LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BI	<i>Business Intelligence</i>
CASC	<i>Computers & Applied Sciences Complete</i>
DM	<i>Data Mining</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extraction, transformation and loading</i>
ISTA	<i>Information Science & Technology Abstracts</i>
KDD	<i>Knowledge Discovery in Databases</i>
MBA	<i>Market Basket Analysis</i>
MD	Mineração de Dados
PDV	Ponto de Venda
RSL	Revisão Sistemática da Leitura
SKU	<i>Stock Keeping Unit</i>
TI	Tecnologia da Informação
TID	<i>Transaction ID</i>
RSL	Revisão Sistemática da Leitura

SUMÁRIO

1 INTRODUÇÃO	14
1.1 OBJETO E PROBLEMA DE PESQUISA	16
1.2 OBJETIVOS	18
1.2.1 Objetivo Geral	18
1.2.2 Objetivos Específicos	18
1.3 JUSTIFICATIVA	19
1.3.1 Justificativa acadêmica	19
1.3.2 Justificativa gerencial	26
1.3.3 Justificativa para o acadêmico	30
1.4 DELIMITAÇÕES DO TRABALHO	30
1.5 ESTRUTURA DO TRABALHO	31
2 REFERENCIAL TEÓRICO	33
2.1 VAREJO	34
2.1.1 Lojas de Conveniência	36
2.1.2 Marketing de Varejo	38
2.2 ARMAZÉM DE DADOS	40
2.2.1 Descoberta de Conhecimento em Banco de Dados	41
2.2.2 Mineração de Dados	43
2.2.2.1 Métodos de <i>Data Mining</i>	45
2.3 TAXONOMIAS DE PRODUTOS	46
2.4 REGRAS DE ASSOCIAÇÃO	48
2.4.1 Análise de Cestas de Compras	49
2.4.1.1 Conjunto de Itens Frequentes	51
2.5 ALGORITMOS DE KDD	54
3 PROCEDIMENTOS METODOLÓGICOS	58
3.1 MÉTODO CIENTÍFICO	58
3.2 MÉTODO DE PESQUISA	59
3.3 MÉTODO DE TRABALHO	64
3.4 COLETA DE DADOS	66
3.5 ANÁLISE DOS DADOS	67
4 ANÁLISE DOS RESULTADOS	69
4.1 ORGANIZAÇÃO ANALISADA	69

4.2 COMPREENSÃO DOS DADOS	70
4.3 PREPARAÇÃO DOS DADOS	71
4.3.1 Limpeza dos dados	71
4.3.2 Integração de dados.....	72
4.3.3 Transformação dos dados.....	72
4.4 MODELAGEM.....	75
4.4.1 Desenvolvimento do Modelo	76
4.4.2 Teste de Validação do Modelo Construído.....	78
4.4.3 Utilização do Modelo	82
4.5 AVALIAÇÃO	86
4.5.1 Análise das Regras de Associação	86
4.5.2 Análise das Regras de Associação x Dia da Semana.....	88
5 DISCUSSÃO DOS RESULTADOS.....	96
6 CONSIDERAÇÕES FINAIS	98
6.1 LIMITAÇÕES DO TRABALHO	100
6.2 SUGESTÕES PARA TRABALHOS FUTUROS	101
REFERÊNCIAS.....	103
APÊNDICE A – ASSOCIAÇÕES GERADAS EM COMPARAÇÃO AOS PERÍODOS DO DIA.....	113

1 INTRODUÇÃO

Frente a uma necessidade primária das empresas em alcançar o destaque no mercado ao qual competem, presenciado um ambiente empresarial gradativamente mais competitivo, estas organizações buscam melhorar os seus produtos e serviços fornecidos, a fim de atender de forma efetiva às exigências do consumidor. (SILVA, 2009). Os consumidores, cada vez mais exigentes, visam soluções de melhor qualidade e com preços acessíveis. O dinamismo do mercado atual exige uma resposta rápida as circunstâncias do consumidor, que está cada vez mais seletivo em suas compras. (PEZZOTTI, 2020). Um fator crítico de sucesso a um negócio, é a possibilidade de coletar informações, com respostas efetivas às turbulências e necessidades do mercado. (PERIZZOLO, 2005).

Os avanços da tecnologia, as quais têm acesso facilitado oriundo dos custos que estão sendo reduzidos por meio da ampla oferta de sistemas e ferramentas digitais existentes no mercado, torna a tecnologia mais acessível a pequenos e médios empreendedores. Empresários têm sido motivados a investir em tecnologia para obter um melhor desempenho frente à concorrência (CALDAS; RIBEIRO, 2017) buscando extrair o máximo de informações em bases de dados. Conforme afirmação de Costa e Junior (2018), o volume de informações disponíveis e produzidas pelas empresas acarreta necessidade de busca de conhecimentos sobre os registros.

Em virtude da amplitude de canais existentes decorrentes da transformação digital, que facilita o acesso de pessoas à internet, e com o aumento no uso de dispositivos que compõem a Internet das Coisas (IoT – *Internet of Things*), há a geração de um conjunto de dados extremamente amplos. (MORAIS, 2018). Estes dados necessitam de ferramentas preparadas para lidar com esta volumetria, de forma que toda e qualquer informação nesse meio possa ser encontrada, analisada e aproveitada em tempo hábil dado a elevada frequência de modificação e valor que possuem às empresas. (MORAIS, 2018).

Empresas capazes de explorar e analisar melhor os diferentes tipos de dados. (WERNECK, 2019), provendo reconhecimento a partir dos dados e reagindo por meio da inteligência empresarial tendem a possuir destaque frente aos seus concorrentes por meio do uso desta tática. A inteligência empresarial que é caracterizada pela busca por entender os fatores e processos humanos e organizacionais envolvidos na busca de informações. (MATHEUS; PARREIRAS, 2004).

De acordo com Assad (2011), a inteligência empresarial é dada como uma atividade estratégica visando a articulação junto ao processo de coleta, análise e disseminação de informações relevantes para uma organização, de modo a favorecer o processo de tomada de decisões. De acordo com Davenport e Prusak (1998), a inteligência empresarial está focada nos sistemas de informações empresariais, onde é apoiada em tecnologias adequadas na intensa interação entre as fontes e usuários de informação, na construção de cenários para o negócio e na disposição para inovar. (ASSAD, 2011).

O intenso uso da tecnologia para comunicação e realização de atividades corporativas, leva a um crescimento exponencial da geração de dados. (ARMAZENAMENTO, 2017). Conseqüentemente o número de informações relacionadas as transações comerciais geradas vão se acumulando com o decorrer do tempo. Todos esses aspectos relacionados ao crescimento da geração de dados criam oportunidades para extrair informações úteis que podem ser exploradas. (VILLARS; OLOFSON; EASTWOOD, 2011). Podendo também ser, a transformação de dados imperfeitos, complexos e frequentemente desestruturados em informação útil. Isso implica revelar tendências e correlações entre grandes conjuntos de dados que, de outra forma, permaneceriam desconhecidos. (MICHAELI, SIMON, 2008).

Gamarra, Guerrero e Montero (2016) definem que o processo de descoberta de conhecimento útil a partir de uma coleção de dados pode ser definido como KDD (*Knowledge Discovery Databases*). Bernstein, Provost e Hill (2005) afirmam que a descoberta de conhecimento em banco de dados é o resultado de um processo exploratório que envolve a aplicação de vários procedimentos algorítmicos para manipular dados, criar modelos e a partir destes dados, manipular os modelos. Uma das etapas do processo é denominada de mineração de dados (*Data Mining*), onde uma de suas práticas é a exploração de dados não muito bem conhecidos, objetivando extrair padrões sobre estes, prática realizada por meio de algoritmos específicos.

Um mercado em grande expansão no país e que movimentava bilhões de reais é o setor de lojas de conveniência. O novo perfil do consumidor, que busca tecnologia, vantagens financeiras na aquisição dos produtos, e vêm priorizando as facilidades de ganho de tempo em seus trajetos diários, influencia a expansão do mercado de lojas de conveniência. (VITRINE, 2020). De acordo com Kotler e Keller (2013), o diferencial deste tipo de loja é a agregação da quantidade de serviços aos bens tangíveis oferecidos no ponto de venda, potencializando o consumo. Considerado um

estabelecimento de venda de insumos e que possui um alto fluxo de pessoas, é caracterizado pela ausência de vendas em grandes quantidades de produtos.

A extração de informações a partir da utilização dos dados gerados pode possibilitar aos gestores da empresa maior poder de decisão frente às alternativas diárias para elevar à competitividade deste negócio. Na seção seguinte, serão apresentados o objeto e problema de pesquisa.

1.1 OBJETO E PROBLEMA DE PESQUISA

A tentativa de entender os movimentos do mercado é uma das grandes habilidades requeridas nas empresas frente aos ambientes incertos e turbulentos (MORITZ, 2012). Com o advento da concorrência globalizada, empresas buscam aumentar sua competitividade. Os investimentos em Tecnologia da Informação (TI) são considerados por pesquisadores e gestores empresariais como uma alavanca para melhorar esta competitividade das organizações. (FERREIRA; CHEROBIM, 2012).

O desenvolvimento da TI proporciona uma relação mais próxima entre a organização e seus clientes, visto que permite a identificação de suas necessidades por meio da coleta e tratamento dos dados. (BAHRAMI; GHORBANI; ARABZAD, 2012). Há alguns anos, a TI esteve focada no aprimoramento de processos, atualmente, as aplicações do setor de TI ultrapassam esse objetivo, visando também, o estreitamento das relações entre a empresa e seus parceiros, concorrentes, fornecedores, clientes e governo. (FERREIRA; CHEROBIM, 2012). Assuntos como o modo como as informações do cliente são obtidas, processadas e usadas para agregar estratégias de alto valor aos consumidores-alvo, tornaram-se importantes para os varejistas. (LIAO; CHEN; WU, 2008). Um dos desafios para as corporações modernas é como extrair dados dos padrões de consumo de seus clientes através de banco de dados transacionais. Este dados podem auxiliar no desenvolvimento de uma maior vantagem competitiva, sendo um bom exemplo para dados de suporte à ações de marketing. (KHOLOD, 2018).

Conforme Figueira (1998), a cada ano, companhias acumulam mais e mais informações em seus bancos de dados. Como consequência, os bancos de dados passam a conter verdadeiros “tesouros” de informação a respeito de diversos procedimentos dessas companhias. Toda esta informação pode ser usada para melhorar seus procedimentos, permitindo que a empresa detecte tendências e

características disfarçadas, e reaja rapidamente a um evento que ainda pode estar por ocorrer. No entanto, apesar do enorme valor desses dados, a maioria das organizações é incapaz de aproveitar totalmente o que está armazenando em seus arquivos. (GONÇALVES, 1999).

A aplicação progressiva das tecnologias da informação aumentou a quantidade de dados coletados pelas empresas varejistas. Há muitos dados disponíveis, mas analisar e compreender as informações disponíveis para tomar as melhores decisões é um desafio. A análise correta dos dados pode ser o diferencial para continuidade de uma empresa no mercado. Muitas vezes existem padrões, difíceis de serem percebidos, que são muito úteis para os negócios, mas que não são trivialmente observados sem alguma ferramenta específica de classificação ou agrupamento. (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). É natural a busca de técnicas computacionais para a descoberta de padrões e estruturas significativas dos enormes volumes de dados. O conhecimento sobre padrões de comportamentos pode nos trazer informações a fim de serem transformadas em negócios rentáveis e assim, a serem explorados. As empresas se utilizam de dados para ganho de vantagem competitiva e para prover mais valores aos clientes.

Os empresários têm ciência da necessidade de entender aquilo que seus clientes estão comprando e quando estão comprando. Ganhos operacionais provenientes da gestão eficiente de um estabelecimento comercial poderão ser convertidos em vantagens para os clientes, gerando diferenciais competitivos frente aos concorrentes. (PERIZZOLO, 2005). O uso do conhecimento tem cada vez mais um papel central na capacidade de conquistar novos mercados. (FERREIRA; CAMPOS; MACEDO, 2006).

Entender a relação entre venda conjunta de itens possibilita ações e promoções pontuais e mais eficientes, aumentando margens de lucros em produtos, conseqüentemente levando a loja a destacar-se na comparação aos resultados com parceiros e concorrentes. Obter destaque frente aos atuantes do mesmo mercado também é reflexo do desenvolvimento de ofertas alinhadas às preferências do consumidor do século 21, que busca praticidade e qualidade aliadas à tecnologia. Em lojas de conveniência não há a expectativa de grandes volumes de vendas, estas lojas vendem, pois, atendem à necessidade de compra de momento do cliente. (HAZOFF; SAUAIA, 2010), indiferente o dia da semana. Por este motivo, reforça-se a necessidade do estímulo de vendas por meio da associação de produtos nas

transações. Identificar por meio dos dados, situações onde, em diferentes transações dos dias das semanas, há a semelhante compra de produtos, sem que estes itens estejam sendo ofertados juntos, é um fator importante para a gestão de negócios.

Conforme Fayyad, Piatetsky-Shapiro e Smyth (1996), com o volume de dados crescendo, uma análise de dados de forma manual se torna impraticável. Este aumento se dá e muito pela grande quantidade de registros, e pelos diferentes atributos relacionados a um objeto. O trabalho de análise precisa ser automatizado, pelo menos parcialmente. (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Com o intuito de compreender os padrões de compras em uma loja de conveniência, nesta monografia será analisado um método de mineração de dados que busque compreender as relações de vendas entre os produtos, procurando entender as composições junto aos dias da semana. Com isso, esta pesquisa busca utilizar uma técnica de *Data Mining* como meio para obtenção de melhores resultados em termos de performance, a fim de responder à questão:

Como identificar padrões de demandas dependentes e associá-las a questões temporais?

1.2 OBJETIVOS

Nesta seção são apresentados o objetivo geral e os objetivos específicos que norteiam a pesquisa.

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é, dentre os vários algoritmos existentes de *Data Mining*, selecionar um capaz de, a partir das transações de compras dos clientes de uma loja de conveniência, estabelecer as associações entre os itens mais frequentes para então relacioná-los aos dias da semana.

1.2.2 Objetivos Específicos

Os objetivos específicos do trabalho são os seguintes:

- a) desenvolver uma abordagem de ETL (*Extraction, Transformation and Loading*) sobre os dados não estruturados;
- b) identificar um algoritmo que possa ser empregado na base de dados da loja de conveniência;
- c) identificar as associações de compras geradas a partir do algoritmo aplicado;
- d) analisar as associações frente aos dias da semana para gerar insights do negócio no auxílio às tomadas de decisões.

1.3 JUSTIFICATIVA

Nesta seção são expostos os motivos que justificam a realização desta pesquisa. Buscando apresentá-los sob três perspectivas distintas: justificativa da pesquisa para o meio acadêmico, justificativa da pesquisa para a organização, e justificativa da pesquisa para o pesquisador.

1.3.1 Justificativa acadêmica

A justificativa acadêmica deste estudo consiste em oferecer e complementar os estudos existentes em torno do tema abordado, visando contribuir na evolução dos conhecimentos em torno do assunto. Assim, com base em Morandi e Camargo (2015), foi realizado o procedimento de revisão sistemática da literatura a partir dos termos de interesse deste trabalho que atendam os critérios definidos. No Quadro 1 encontra-se detalhado o protocolo seguido para a RSL.

Para a pesquisa dos trabalhos, buscou-se por meio do processo de ajuste da *string*. Tarefa executada para verificação dos melhores termos relativos ao tema de estudo, com intuito de filtrar os resultados de trabalhos que possam servir como base referencial, ou seja, obter o máximo de estudos primários relacionadas a pergunta de pesquisa utilizando uma estratégia de busca imparcial.

Para o processo de ajuste da *string*, primeiramente realiza-se uma revisão prévia na literatura para a avaliação da volumetria de estudos potencialmente relevantes. (SALAZAR, 2015). A partir dessa avaliação, utiliza-se o Google Acadêmico para testar as várias combinações dos termos possíveis de busca a partir do tema de pesquisa. As *strings* são então construídas utilizando operações *booleanas AND*, que

é uma condição de agregação entre os termos em uma pesquisa, e o operador *OR*, utilizado para representar a equivalência entre os termos.

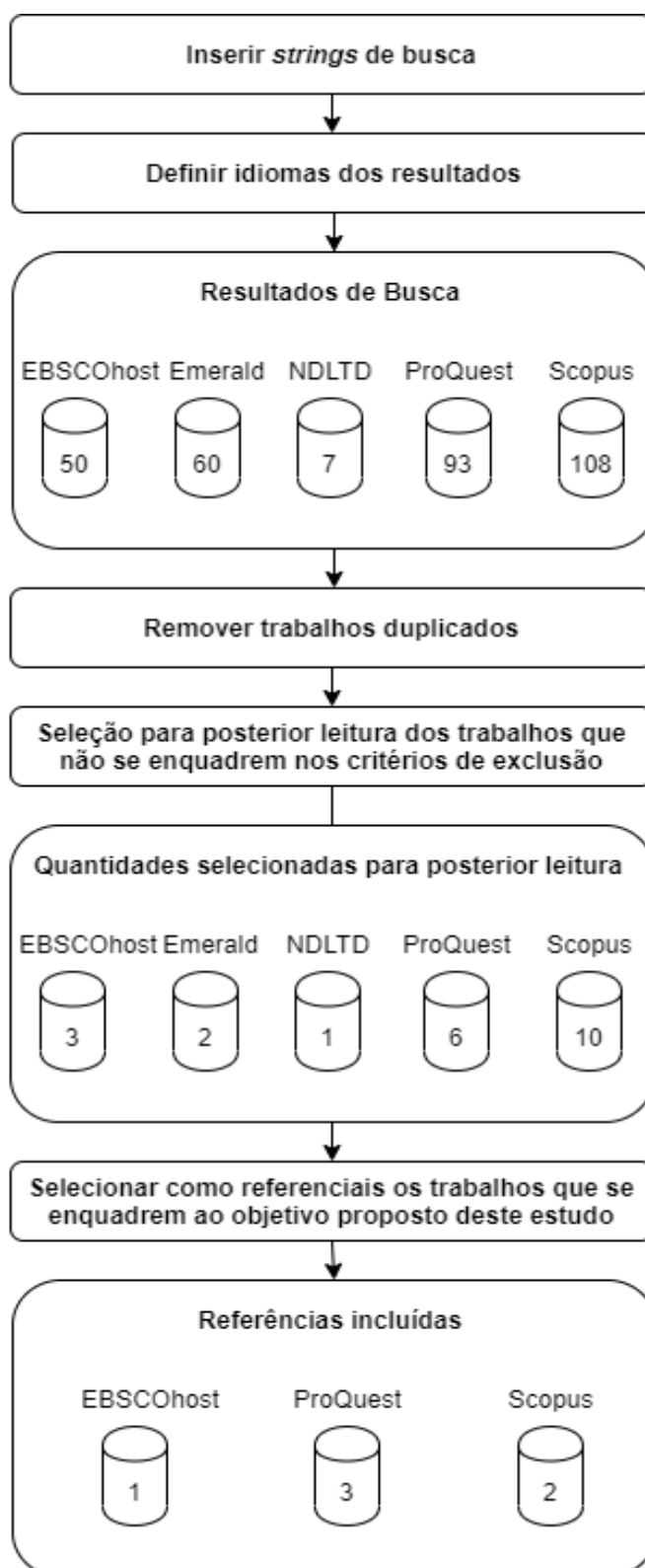
Quadro 1 - Protocolo para a Revisão Sistemática da Literatura

<i>Framework</i> Conceitual	Revisão sistemática da literatura com o objetivo de identificar estudos realizados sobre a aplicação de <i>Data Mining</i> em lojas de conveniência.
Contexto	Análise de combos de venda em lojas de conveniência relacionados aos dias da semana utilizando a ferramenta <i>Data Mining</i>
Horizonte	Sem restrição temporal
Idiomas	Português e Inglês
Questão de revisão	Identificação de um algoritmo de Data Mining que relacione a venda de produtos com os dias da semana.
Crítérios de inclusão	Aplicação do processo em algum dos ramos do mercado varejista
Crítérios de exclusão	Ausência da aplicação de algum método de data mining e sua referência
<i>Strings</i> de Busca	"Mineração de Dados" AND "Regras de Associação" AND "Análise de Cestas de Compras"
	"Mineração de Dados" AND "Regras de Associação" AND "Loja de Conveniência"
	"Mineração de Dados" AND "Regras de Associação" AND "Varejo"
	"Data Mining" AND ("Association Rules" OR "Rules Association") AND "Market Basket Analysis"
	"Data Mining" AND ("Association Rules" OR "Rules Association") AND "Convenience Store"
	"Data Mining" AND ("Association Rules" OR "Rules Association") AND "Retail Store"
Bases de Busca	EBSCOhost
	Emerald
	NDLTD
	ProQuest
	Scopus Elsevier

Fonte: Elaborado pelo autor.

Na Figura 1 é possível avaliar as etapas realizadas, bem como os resultados. As buscas foram realizadas nas bases EBSCO HOST, EMERALD, NDLTD, PROQUEST e SCOPUS, sendo os termos inseridos em português e inglês. Os termos da busca foram inseridos entre aspas, no qual o índice da busca foi o texto completo.

Figura 1 – Busca de Referências



Fonte: Elaborado pelo autor.

A base EBSCO HOST foi acessada através do site da UNISINOS¹, pela opção de acesso via Portal da CAPES. Evidenciado que esta base reúne diferentes bases de dados, a pesquisa realizou-se naquelas que, através de sua descrição mais provavelmente retornariam trabalhos do interesse. As bases selecionadas foram as seguintes²:

- i) *Academic Search Complete*: base de dados de texto completo multidisciplinar, mais valiosa do mundo e mais abrangente academicamente, com mais de 8.500 periódicos de texto completo, incluindo mais de 7.300 periódicos revisados por especialistas. Além de texto completo, essa base de dados oferece resumos para mais de 12.500 periódicos e um total de mais de 13.200 publicações, incluindo monografias, relatórios, documentos de conferências, etc;
- ii) *Business Source Complete*: mais completa base de dados acadêmica na área de negócios do mundo, que oferece a melhor coleção de conteúdo bibliográfico e em texto completo. Como parte da cobertura abrangente oferecida por esta base de dados, também estão incluídos os índices e resumos dos periódicos científicos acadêmicos mais importantes desde 1886, além das referências pesquisáveis citadas fornecidas de mais de 1.300 periódicos científicos;
- iii) *Academic Search Premier*: Esta base de dados multidisciplinar fornece o texto completo de mais de 4.600 periódicos, incluindo o texto completo de praticamente 3.900 títulos analisados por especialistas;
- iv) *Information Science & Technology Abstracts*: É a principal base de dados das áreas de ciência da informação. O ISTA reúne artigos de revistas especializadas de mais de 450 publicações, além de livros, relatórios de pesquisa e anais de conferências e patentes, com cobertura abrangente e contínua dos periódicos mais importantes nessa área.
- v) *Computers & Applied Sciences Complete*: cobre o espectro de pesquisa e desenvolvimento da computação e disciplinas de ciências aplicadas. O CASC fornece indexação e resumos para praticamente 2.200

¹ Acesso via <http://unisininos.br/biblioteca/pesquisa#bases-de-dados>

² Descrições das bases disponíveis em <https://web.a.ebscohost.com/ehost/search/selectdb?vid=0&sid=bd8ef590-b18b-4eb5-a15f-b3876467e7e3%40sdc-v-sessmgr01>

periódicos acadêmicos, publicações profissionais e outras fontes de referência a partir de uma coleção diversificada.

As demais bases também foram acessadas via portal da CAPES, onde com o protocolo definido, as buscas através das *strings* retornaram 318 resultados. Destes, 28 foram removidos por serem identificados como duplicados. Dos restantes 290 títulos, todos os títulos foram lidos, e excluiu-se aqueles que não se enquadravam ao estudo. Através destes critérios, 96 resumos/*abstracts* foram lidos para verificar quais se enquadrariam nos processos de inclusão. Sendo assim, foram selecionados para posterior leitura na íntegra, 22 trabalhos que possuíam relevância com o estudo proposto, ou seja, não se enquadravam em critérios de exclusão.

Além destes 22 trabalhos selecionados, outros 7 foram selecionados e lidos por meio do procedimento conhecido como *backward* ou retrospectivo, no qual consiste em consultar as referências do estudo. (MORANDI; CAMARGO, 2015) com intuito de descobrir diferentes estudos não encontrados nas bases pesquisadas.

Com as avaliações sobre as leituras realizadas, identificou-se uma lacuna de ausência em torno de trabalhos relacionados ao tema, considerando a utilização de DM em lojas de conveniência. Este fator de ausência, ressalta a importância deste estudo para o meio acadêmico. Sendo assim, para a inserção junto a esta pesquisa foram selecionados 7 trabalhos que relacionam o tema *Data Mining* por meio da aplicação de regras de associação em alguma segmentação do mercado varejista. Estes trabalhos encontram-se no Quadro 2. Reforça-se que o trabalho demarcado com um asterisco na coluna base, representa que este é oriundo do processo de *backward*.

Quando os termos pesquisados em português, e após leitura de alguns resumos, pode ser observado uma limitada qualidade associada ao assunto do tema de pesquisa. Devido a esta ausência de relevantes estudos no âmbito nacional, nenhum estudo nesta língua foi adicionado à pesquisa.

Quadro 2 – Trabalhos incluídos

Título/Resumo	Ano	Autores	País de registro	Base
Discovering Iranians Shopping Culture by Considering Virtual Items Using Data Mining Techniques	2009	J. Shahrabi e R.S. Neyestani	Iran	ProQuest
Market Basket Analysis of Convenience Store POS Data	2010	Marina Kholod	Japão	Scopus
Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data	2018	Fachrul Kurniawan, Binti Umayah. Jihad Hammad, Supeno Mardi Susiki Nugroho e Mochammad Hariadi	Indonésia	*
Product Selection for promotion planning	2010	Yiunghui Yang e Chunhui Hao	Estados Unidos da América	ProQuest
Evaluating time variations to identify valuable association rules in market basket analysis	2013	Vasilios Papavasileiou e Athanasios Tsadiras	Grécia	EBSCOhost
Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions	2011	Pramond Prasad e Latesh Malik	Índia	ProQuest
Retail business analytics: Customer visit segmentation using market basket data	2018	Anastasia Griva, Cleopatra Bardaki, Katerina Pramadari, Dimitris Papanikolaou	Grécia	Scopus

*Selecionado por meio do processo de backward

Fonte: Elaborado pelo autor.

Pesquisado na literatura de âmbito internacional, observou-se a utilização de diferentes métodos de DM para avaliações de negócios varejistas, onde, muitos estudos visam o estreitamento da relação com os clientes. Para esta aproximação aos clientes, em muitos casos buscou-se compreender os hábitos dos consumidores com base em seu histórico de compras. Shahrabi e Neyestani (2009) realizaram a análise de cluster para classificar os clientes de uma loja varejista Iraniana, e paralelamente realizaram a aplicação de regras de associação sobre os produtos para descobrir produtos tendenciosos a venderem juntos. Posteriormente, relacionou-se a

clusterização dos clientes com a associação dos produtos para identificar padrões de compras dos grupos de clientes.

Os autores Kholod (2010) e Kurniawan et al (2018) reforçam a utilização do MBA para compreender quais itens são comprados pelos clientes simultaneamente, a fim de ter uma oportunidade de promoção e prover ações por meio de marketing para obter vantagens competitivas. Análogo a estes trabalhos, Yang e Hao (2011) utilizam o método *Top Product Selection* (TP) para a descoberta de conjunto de itens que estão associados, com o objetivo de descobrir como selecionar o produto certo para promover, na busca por maximizar o benefício promocional. O pressuposto do embasamento é de que, com um conjunto de itens em promoções, tendem a surgir um efeito promocional em outros produtos associados, aumentando potencialmente o lucro do estabelecimento.

No estudo realizado sobre mineração de dados, Papavasileiou e Tsadiras (2013) buscam gerar as relações entre produtos, propondo um método considerando as variações de tempo das regras de associação, visando obter as associações de produtos que são mais estáveis com o decorrer do tempo. Os autores reforçam que a compreensão do comportamento do consumidor influencia em estratégias de marketing assertivas. Prasad e Malik (2011), abordou o conhecimento sobre as associações de vendas dos produtos para maior embasamento em decisões estratégicas de uma companhia varejista.

Em semelhança ao objetivo proposto nesta dissertação, no trabalho selecionado de Griva et al (2018), foi inserida a análise de cesta de compras para compreender cada compra dos clientes, associando o período do dia com os itens atribuídos às compras. O intuito dos autores é encontrar que tipo de produto um cliente tende a comprar em determinado momento do dia. Assim, poderá junto ao marketing da empresa propor promoções de itens específicos, alegando que com a submissão de descontos em determinados produtos, outros, que não estão interligados de forma direta a estes, tenham considerável aumento escalar.

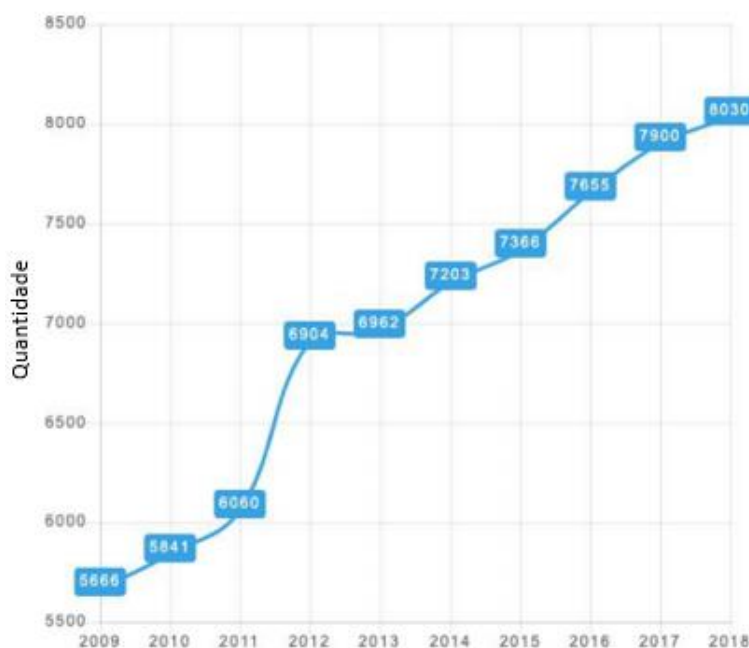
Por meio da RSL foi possível detectar a presença de inúmeros estudos com aplicações de DM. Porém, observou-se a inexistência de trabalhos relevantes centralizados na aplicação no mercado varejista de lojas de conveniências e que considerem algum período temporal, podendo ser turno, dia, mês ou range de meses. Assim sendo, esta pesquisa avança no sentido de fornecer à academia referências

bibliográficas que relacionam as regras de associações aos períodos de suas ocorrências.

1.3.2 Justificativa gerencial

Ao final do ano de 2018, de acordo com o anuário da SINDICOM³, o Brasil, conforme ilustrado no Gráfico 1, possuía 8.030 lojas de conveniência localizadas dentro de postos de combustíveis, sendo o número de postos em torno de 41.000, ou seja, reforça que ainda há um grande espaço para crescimento deste tipo de negócio em estudo. O canal de lojas de conveniência em postos de combustíveis, alcançou em 2018, o faturamento de R\$ 7,54 bilhões, conforme apresentado no Gráfico 2, tendo um ticket médio de R\$12,82 por cliente.

Gráfico 1 - Evolução da quantidade de lojas.



Fonte: Adaptado de SINDICOM (2019).

³ Anuário disponível para acesso em <https://sindicom.com.br/index.php/anuarios/anuario-sindicom-2019/>

Gráfico 2 - Faturamento do mercado total (R\$ Bilhões/Ano)



Fonte: Adaptado de SINDICOM (2019).

A ampla capilaridade e o hábito de ir até o posto de combustível complementam as fortalezas desse canal aqui no Brasil. Um posto que tem uma loja de conveniência junto do seu negócio fatura em média de 20% a 30% a mais em combustível. (BEECK, 2018).

No primeiro semestre de 2020, embora o mundo esteja presenciando uma terrível pandemia, no Brasil o segmento de lojas de conveniência teve um aumento de faturamento de 11,6% em comparação ao mesmo período do ano anterior. (AS LOJAS, 2020). Nesse sentido, as vendas em lojas de conveniência no mundo em 2020 deverão registrar o 3º maior crescimento entre as categorias de varejo, ficando atrás somente para as vendas por *e-commerce* e para as lojas de descontos. Fazer da visita do cliente uma experiência marcante, ou até inesquecível, passou a ser o objetivo de muitos estabelecimentos para competir com o *e-commerce*.

Portanto, a busca por dados para entender como e o que seus clientes consomem, o que buscam, como pensam, como se comunicam, seus ideais, seus hábitos, e a relação com a tecnologia, representam um caminho importante a percorrer quando o objetivo do varejista é conquistá-los e fidelizá-los. Dados estão sendo gerados em velocidade e quantidade antes inimagináveis, e por vezes, de difícil análise. (MORGADO, 2017). O desafio atual é administrá-los e transformá-los em informação útil, ou seja, a maior dificuldade é a análise destes dados e a capacidade

de prover conhecimento sobre estas informações. Entende-se que o conhecimento é um ativo de uma organização e proporciona vantagem competitiva em relação aos concorrentes, visto que é passível de ser gerado e replicado. (VENKITACHALAM; AMBROSINI, 2017).

A tecnologia de comunicação fez com que os varejistas cada vez mais coletassem facilmente os dados diários das transações a um custo muito baixo. Por meio do sistema de ponto de venda e da tecnologia de código de barras, uma loja de varejo pode coletar um grande volume de dados de transações de clientes. A partir do *Big Data*, grandes quantidades de informações úteis podem ser extraídas para apoiar as estratégias de negócios de gerenciamento de varejo.

Informações como o tipo de público atendido, a quantidade de produtos vendidos, tempo médio de atendimento, recomendação de produtos e de fornecedores podem ser cruzadas para extrair não apenas informações sobre um período passado, mas também para gerar *insights* para o futuro. Uma das tarefas mais conhecidas e utilizadas da mineração de dados é a regra de associação. Como salientado por Lamboia e Pereira (2005, p. 7), “consiste em identificar fatos que possam ser direta ou indiretamente associados, sendo esta estratégia geralmente usada em aplicações onde se busca identificar itens que possam ser colocados juntos em uma transação”, ou seja, encontrar relações ou padrões frequentes entre conjuntos de dados. De acordo com Da Silva (2016), indicar itens consultados em uma mesma operação designa-se como transação. Por meio da análise de cesto de compras, poderá ser possível encontrar o mix ideal de produtos para ofertar em uma transação.

As condições de vida moderna nos grandes centros urbanos levam as pessoas a buscar resolver suas necessidades em locais próximos da residência ou do local de trabalho. (MERCADO, 2018). Com as mudanças corriqueiras dos cenários econômicos, proprietários de lojas de conveniência querem participar e garantir um retorno sobre estas mudanças, quebrando paradigmas e ofertando diferentes categorias de produtos, tornando esta oferta bem variável. De acordo com estudo realizado pela consultoria Nielsen⁴, cerca de 70% das decisões de compras no Brasil são tomadas no ponto de venda. A capacidade de fazer previsões de tendências de consumo tem se apresentado como um grande avanço para os negócios. Afinal,

⁴ Estudo realizado pela consultoria Nielsen, disponível em <https://www.nielsen.com/br/pt/>.

utilizar estratégias mais assertivas para garantir as próximas vendas é o grande objetivo desse setor.

O *Data Mining* vem sendo utilizado por empresas que mantenham seu foco direcionado para o consumidor, principalmente organizações de varejo e marketing, que buscam aprimorar o conhecimento através dos dados tomando decisões rápidas e assertivas para o negócio. (KERCKHOFF, 2018). O fato de identificar a venda de produtos relacionados em um determinado dia da semana, pode ser viável para adoção de uma estratégia de marketing para propor promoções diárias. A existência da promoção de determinados itens, a cada dia, possibilita a tendência de o cliente comprar algum outro produto associado. Além de que, a existência de um preço menor em determinado dia desperta no cliente o chamado senso de urgência de compra, que é um interessante potencializador de promoções. De acordo com Koiffmann (2018), o senso de urgência é conhecido por estimular na mente do consumidor a necessidade de compra momentânea.

Com o intuito de compreender os padrões de compras em uma loja de conveniência, nesta monografia será utilizado a regra de associação por meio da aplicação de mineração de dados, para entender as composições junto aos dias da semana.

Diante de uma era de desejos conflitantes, onde encontra-se a conveniência de realizar pedidos e recebê-los em casa, ou na busca por uma experiência prática de loja, a tecnologia digital capacita o mercado com inúmeras ferramentas. Um conceito de loja do futuro vem sendo introduzido no mercado por grandes varejistas, através de lojas menores. Estas lojas apresentam valor agregado, oferecendo práticas mais convenientes e adaptadas ao novo estilo de vida dos clientes, facilidade no acesso e disposição da seleção de produtos, possibilitando experiências inovadoras, flexíveis, eficientes e acessíveis.

A flexibilização em ofertas que englobam necessidades diárias dos consumidores pode gerar uma maior frequência de visitas dos clientes, uma vez que, sendo prazerosa sua visita ao estabelecimento, no qual consiga unir bom preço com uma compra ágil, o consumidor atual não encara como problema uma visita diária a uma loja deste padrão. E visitas diárias, conseqüentemente influenciam a compra de novos produtos.

1.3.3 Justificativa para o acadêmico

É essencial para qualquer Engenheiro que atue no âmbito empresarial, estar atento às tecnologias emergentes e suas utilidades. A possibilidade de enriquecimento do conhecimento frente a uma tarefa como o *Data Mining*, que tem aumentado substancialmente sua importância no dia a dia dos negócios, capaz de analisar, especificar e solucionar possíveis problemas nas cadeias empresariais por meio de dados, é fator primordial para o acadêmico. Ao mesmo tempo em que o encontro com informações do mercado varejista, que é um dos maiores influenciadores na economia mundial, eleva a possibilidade de compreensão do tipo de mercado e suas inúmeras futuras possibilidades de aplicações e oportunidades para novos negócios, vinculando e atribuindo diversas teorias adquiridas ao longo da graduação de Engenharia de Produção.

1.4 DELIMITAÇÕES DO TRABALHO

A intenção deste trabalho é contribuir para o conhecimento da organização a partir dos dados de seus registros de venda, buscando apresentar um modelo para encontrar as relações entre os produtos registrados nas transações de venda junto aos dias da semana, ou seja, encontrar combinações de itens que mais possuem aderência a determinado dia da semana.

A respeito da implementação deste trabalho a partir dos resultados, esta não será a finalidade deste estudo. O objetivo é obter aprendizado ao acadêmico, apresentando, posteriormente, os resultados ao gestor da loja de conveniência à qual o estudo foi aplicado. Sobre isto, sendo do interesse do proprietário, as implementações podem ser executadas. Em caso de uma possível aplicação do estudo, é plausível uma continuação desse trabalho.

A pesquisa será aplicada em uma loja de conveniência localizada as margens de uma rodovia brasileira de grande fluxo, da qual atende um público diversificado e não possui como padrão registrar os dados de clientes, o que torna inviável determinar uma classificação específica de consumidores. Este estabelecimento não se trata de uma loja franquia, da qual é comum neste segmento. A base de dados coletada abrange o período de um ano, considerado de 01/01/2019 até 31/12/2019.

Uma delimitação se refere ao software utilizado para a aplicação do algoritmo. Embora possam ser utilizadas inúmeras linguagens de programação para modelagem, neste estudo será utilizado o software *RStudio*. Este software é de livre acesso, construído sobre a linguagem de programação R, possuindo todos os pacotes abertos relacionados a ciência de dados, sendo assim, possui uma extensa variedade estatística.

Como base para o desenvolvimento deste estudo, configuração e especificação do modelo, utilizou-se o método proposto por Silva, Peres e Boscaroli (2016), sendo inibida a etapa com cenários em diferentes testes, uma vez que estes não se aplicam a este estudo.

1.5 ESTRUTURA DO TRABALHO

O atual trabalho está estruturado em seis capítulos, que são descritos nesta secção. Na etapa inicial, o capítulo um é formado pelos aspectos que motivam a realização deste trabalho, contextualizando o problema, expondo a questão de pesquisa, os objetivos, as justificativas, delimitações e a estrutura do trabalho.

No capítulo dois é apresentado o referencial teórico, no qual são detalhados os conceitos teóricos necessários abordados para a estruturação desta monografia e o contexto do tema em estudo. Apresenta-se também, a revisão da literatura a partir de livros, artigos, periódicos, teses e dissertações.

No capítulo três é exposto o método de condução da pesquisa, no qual é subdividido em método científico, método de pesquisa e método de trabalho. Posteriormente, neste mesmo capítulo é introduzida a teoria referente à coleta e análise dos dados. A partir da análise dos dados posterior à coleta, averiguou-se a possibilidade de subdividir os dias da semana em períodos, relacionando estes às regras geradas, possibilitando informações com maior teor de detalhe.

No capítulo de número quatro são executadas todas as etapas da metodologia proposta. Em primeiro plano é apresentada a organização. Após, são realizadas as etapas de compreensão e preparação dos dados, a modelagem do algoritmo, sua validação e posterior utilização do modelo, passo a passo, e as regras geradas. Com as regras geradas, apresenta-se as relações junto aos dias da semana.

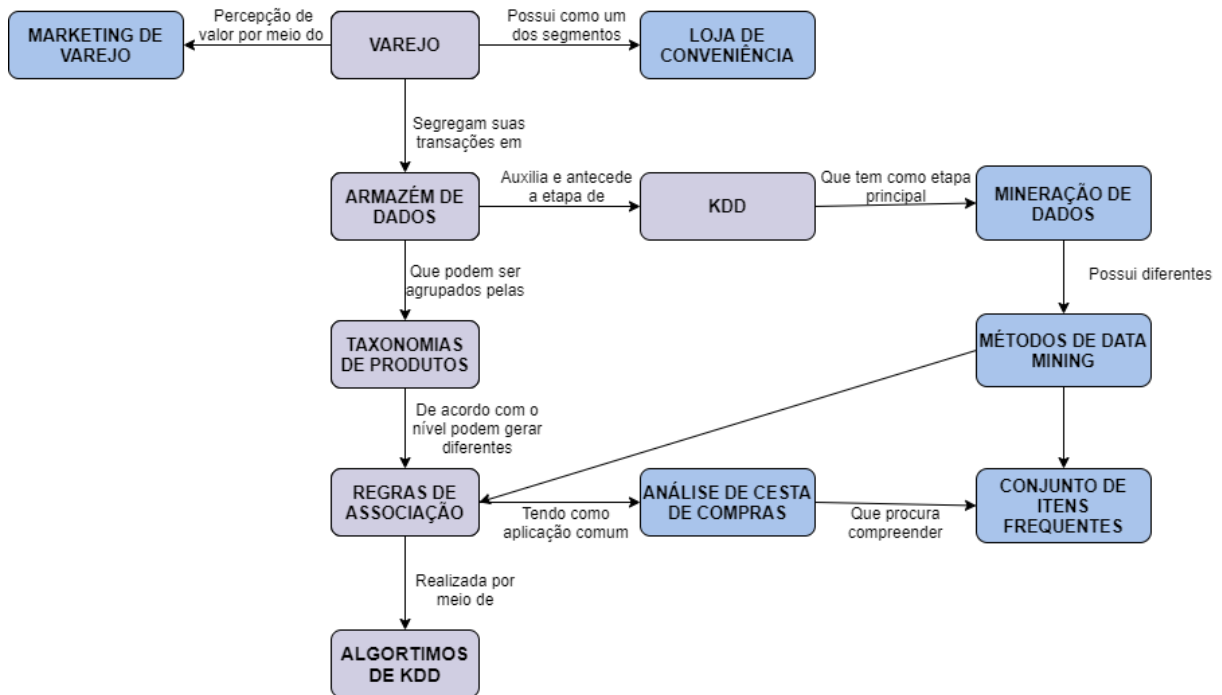
As discussões de resultados são apresentadas no capítulo de número cinco, contemplando as contribuições de pesquisa, tanto para a organização como para a

academia. E no sexto capítulo, contemplam-se as conclusões, sugestões para trabalhos futuros e considerações finais.

2 REFERENCIAL TEÓRICO

Na construção deste referencial teórico serão abordados os conceitos que sustentam este estudo. Estes conceitos são aqui introduzidos conforme demonstrado na Figura 2 por meio do mapa conceitual do escopo deste referencial teórico. Um mapa conceitual é uma ferramenta gráfica que serve para representar, organizar, construir e avaliar conhecimentos, que tem por objetivo representar relações significativas entre conceitos na forma de proposições. (SILVA; CLARO; MENDES, 2017). Uma proposição é constituída de dois ou mais termos conceituais unidos por palavras para formar uma unidade semântica. (NOVAK; GOWIN, 1996).

Figura 2 - Mapa conceitual do escopo do referencial teórico



Fonte: Elaborado pelo autor.

O mapa conceitual desenvolvido, conforme Silva, Claro e Mendes (2017) salientam, não é autoexplicativo, assim, visa destacar os tópicos utilizados neste capítulo para contextualização do assunto abordado nesta monografia.

Primeiramente, será contextualizado o mercado e segmento de negócio específico no qual a ferramenta será aplicada, na sequência será abordado o conceito de *Data Mining*, com explicações referentes às suas tarefas. Após, é apresentado o formato como a pesquisa irá considerar a avaliação dos níveis de categorias de

produto, e então contextualizado a escolha do algoritmo a ser selecionado para o posterior modelo.

2.1 VAREJO

A definição que reúne e sintetiza as descritas é a de Mattar (2011, p. 1), que diz que “varejo consiste nas atividades de negócio envolvidas na venda de qualquer produto ou prestação de qualquer serviço a consumidores finais, para utilização ou consumo pessoal, familiar ou residencial”. De acordo com Levy e Weitz (2000), os varejistas são empresas que conseguem satisfazer as necessidades e desejos dos consumidores, oferecendo a variedade certa de produtos, na quantidade certa, a um preço razoável, na hora e no local desejados. Assim, agregam valor aos produtos e serviços.

Segundo Faleiros (2009), as atividades de varejo são concebidas por vendas de produtos ao consumidor final que, em geral, são comercializados em pequenas quantidades e destinados somente para uso pessoal, cuja venda é realizada em estabelecimentos especializados e não especializados. Os estabelecimentos especializados são postos de combustíveis, farmácias, lojas de móveis, tecidos, roupas ou sapatos; e os estabelecimentos não especializados são hipermercados, supermercados e lojas de departamentos.

A história do varejo no Brasil começou no período colonial, quando surgiram os primeiros armazéns. A cultura de varejo se perdurou na sociedade nas épocas das plantações de café, algodão, cana de açúcar e muito mais, em que os consumidores adquiriam os produtos em pequenas quantidades. Inicialmente, em 1950, a venda em balcão deixou de existir, dando espaço para a liberdade do consumidor em escolher sozinho suas próprias mercadorias. Com o passar do tempo, o varejo passou por uma grande evolução influenciada pela globalização e pelo impacto das tecnologias e da internet. A tecnologia e a internet transformaram a relação entre vendedor e consumidor, permitindo com que as vendas acontecessem virtualmente, por meio de sites e redes sociais.

Com toda essa evolução, o varejo tem se tornado um mercado cada vez mais dinâmico e competitivo, em busca constante por estratégias e investimentos para alcançar novos clientes, fidelizar clientes antigos, aumentar vendas e, conseqüentemente, aumentar o faturamento do estabelecimento. (CONHEÇA, 2018).

O comércio varejista tem se destacado na economia brasileira nos últimos anos. De acordo com matéria do Portal Consumo (2020), o setor varejista seguiu uma taxa média de crescimento de cerca de 2% comparado ao ano anterior, resultados estes que seguem na mesma taxa de evolução dos 3 anos anteriores.

Por se tratar de um dos setores mais dinâmicos da economia, é essencial que o varejo acompanhe constantemente as necessidades e mudanças do mercado. A crescente competição no mercado varejista acarreta maiores níveis de exigências dos consumidores quanto a serviços e produtos oferecidos. A estrutura de varejistas encontra-se muito diversificada. Fator de destaque do setor, segundo Guidolin, Costa e Nunes (2009), é a influência direta no elo final da cadeia de valor dos produtos, pois o varejo representa um canal direto que liga a produção e o consumo, sendo altamente impactado pelas variações na demanda e mudanças nas preferências do consumidor. Nesse contexto, o autor salienta que o maior ativo do comércio varejista é o contato direto com os consumidores, pois com a identificação das preferências do mercado aliada a tradução da demanda percebida para o setor industrial, é possível impulsionar as vendas e, de certa forma, direcionar o rumo da indústria.

Como os varejistas são a unidade de marketing mais próxima dos consumidores da cadeia de suprimentos, a necessidade de varejistas e a importância das lojas aumentam ainda mais o processo de obtenção de informações dos clientes, o compartilhamento dessas informações na cadeia de suprimentos e o desenvolvimento de estratégias que ofereçam alto valor aos consumidores. Conforme Guidolin, Costa e Nunes (2009), o maior ativo do varejo é o monitoramento constante do comportamento do consumidor, e um dos desafios do varejo é lidar com este processo de compra mais complexo e imprevisível. As relações de consumo têm mudado drasticamente com o processo exponencial de informatização junto às novas tecnologias, interferindo em todos os processos executados pelo varejo, das estratégias de mercado à comunicação com os consumidores.

Em sua publicação referente aos desafios do mercado varejista, Morgado (2017, p. 3) salienta que:

O varejo está enfrentando um momento de grandes desafios e possibilidades, mas há espaço para uma verdadeira reinvenção do setor. Para chegar lá, cada empresa tem que compreender o quanto a jornada de compras foi impactada pelas novas tecnologias e quais adaptações precisa fazer. Após esse entendimento, deve criar estratégias para que a equipe use da melhor forma a riqueza de dados e informações que hoje em dia os sistemas são capazes de gerar.

No mercado de varejo, tornou-se vital encontrar o mercado alvo, para então buscar a diferenciação frente aos concorrentes aliando os processos às ferramentas avançadas de tecnologia. Sendo o mercado alvo deste estudo, no subtópico a seguir é mais bem definido o segmento de lojas de conveniência.

2.1.1 Lojas de Conveniência

O conceito de loja de conveniência foi criado em 1927 como apelo ao sentido de economia de tempo. (GRAVES, 2017). Conforme Hazoff e Sauaia (2010), lojas de conveniência são estabelecimentos varejistas para compras convenientes e rápidas, e que operam em sistema de auto atendimento, oferecendo facilidades a consumidores, como sortimento de produtos orientados as marcas líderes, rapidez de atendimento, e em alguns casos sendo operadas em sistema 24 horas. Estas lojas são reconhecidas por atender as necessidades específicas dos seus consumidores, e não se enquadram no modelo de pequeno varejo.

O grande impulso nas lojas de conveniência são suas localizações, principalmente no Brasil, onde estão dispostas dentro de postos de combustível, mas não sendo necessariamente para assim ser designada. Porém, sua expansão está altamente vinculada a ocorrer nestes ambientes, principalmente com lojas definidas em função da bandeira do posto de combustível. (HAZOFF; SAUAIA, 2010). No Brasil, a maioria das conveniências são franqueadas de bandeiras distribuidoras de combustíveis.

Nos últimos anos observou-se um crescimento de grandes redes de varejo - que possuem em seu portfólio redes de hipermercados -, que expandiram para este segmento mais enxuto que visa atender as necessidades rápidas dos consumidores. Estes consumidores têm cada vez menos tempo devido ao crescente envolvimento na vida profissional, e esta globalização fez com que os mercados se aproximassem, incentivando o aumento da tecnologia. (ARIAS, 2019).

De acordo com a Associação Brasileira de Supermercados (ABRAS), que classifica os varejistas segundo o formato das lojas e espaço físico dos empreendimentos, quantidade média de itens comercializados, número de caixas de atendimento, tipos e quantidades de seções, uma loja de conveniência pode ser designada conforme Quadro 3.

Quadro 3 – Características de Lojas de Conveniência

Formato de Loja	Área de Vendas m ²	Nº médio de itens	Nº de caixas de atendimento	Seções
Loja de Conveniência	50 - 200	1.000	1 a 2	Mercearia, frios, laticínios, bazar, <i>snacks</i>

Fonte: Adaptado de ABRAS, 2010.

Este formato de loja frequentemente possui um valor 15% acima da média de preços do mercado. Porém, este não é um viés impeditivo de venda, levando em consideração que os clientes, ali, compram por praticidade. Atualmente, lojas de conveniência respondem por 16% das vendas no varejo brasileiro, percentual ainda abaixo dos 22% que é o índice médio no mundo. (MACHADO, 2019). E um fator a ser considerado é a relevância do mercado de conveniência. As categorias presentes no interior destas lojas podem representar cerca de 80% de todos os itens de vendas do varejo tradicional.

O mercado de lojas de conveniência vive em constante ajuste às necessidades do cliente, que, cada vez mais, busca velocidade e praticidade no seu dia a dia. Para fidelizar o cliente ao canal, as lojas têm o desafio de aliar praticidade a conforto, segurança, cordialidade no atendimento e, claro, produtos de qualidade. O sucesso das lojas também é reflexo do desenvolvimento de ofertas alinhadas às preferências do consumidor do século atual, que leva em conta a saúde e o bem-estar, além de apostas em fatores como tecnologia, logística e experiência de compra do cliente. Portanto, reforçar o contato com o cliente e propiciar uma diferente jornada ao consumidor é uma ênfase a ser dada no setor. Introdução de alta tecnologia e alterações na legislação tem transformado o setor nos últimos tempos. Um diferencial frente a concorrência para uma loja de conveniência será a implementação de técnicas sofisticadas de análise de dados (por exemplo, *big data*) em tempo real, integrando os pontos de venda com as equipes de operação, gestão de categorias de produtos e precificação, de forma a ir refinando as ofertas e colocando os itens certos nos PDVs.

Sendo tendência no país todo, neste ramo de lojas de conveniência, a estratégia é agregar cada vez mais serviços as lojas para se transformarem em verdadeiros centros de serviços, pensando no conceito de torná-lo um lugar completo ao consumidor. Assim, são agregados serviços e módulos a este formato de loja, como padarias, cafeterias gourmet, caixas eletrônicos e um completo *hall* de freezers

contendo bebidas geladas. Com a incessante busca por conveniência em um mundo cada vez mais demandante do nosso tempo, a tecnologia vem ajudando imensamente a realizar tarefas cotidianas.

Há um trabalho bastante analítico por parte dos gestores destes estabelecimentos, que devem identificar as categorias específicas que os consumidores compram em situações de “reposição” e “emergência/impulso”, bem como em fazer o ajuste fino na diversidade e profundidade a ser oferecida em cada categoria. (CARRERA; ABREU; WOODS; LEONE; SONNEVELD; XAVIER, 2017). Dado o espaço reduzido deste tipo de loja, é preciso ser cirúrgico na composição dos itens nas gôndolas. E este vem sendo um entrave na entrada de grandes players varejistas neste segmento, pois tentam ofertar todo o seu elevado mix de hipermercados em um ambiente compacto.

Junto ao avanço tecnológico, o varejo online torna-se um concorrente direto das lojas de conveniências, que devem buscar soluções para capturar oportunidades de se integrar ao consumidor para satisfazer suas necessidades, construir a atenção dos clientes por meio de tecnologia, e trazer valor para as experiências em lojas. E adicionar valor para o cliente, significa aumento de vendas para a loja. Para tornar perceptível este valor, varejistas utilizam o marketing como ferramenta para suas promoções. Este tipo de ação é apresentado no subtópico seguinte.

2.1.2 Marketing de Varejo

O termo caracteriza uma série de ações voltadas para as vendas, que podem ser aplicadas por empresas varejistas para elevar seus lucros. Conforme Santos (2020), o marketing de varejo é um conjunto de medidas que reúnem esforços voltados para os consumidores. Quando bem empregado, ele pode atrair novos clientes, promover seus produtos, fidelizar os clientes que a empresa já tem e fortalecer a marca. Tais medidas impactam diretamente no aumento das vendas e no aumento dos lucros. Sem essas estratégias, por outro lado, o processo pode se tornar mais difícil e lento. Na prática, se vale dos elementos básicos do marketing tradicional, como produto, praça, preço e promoção. E para um aumento considerável do elemento praça, propagando o alcance para novos públicos, a realização de promoções de vendas gera a possibilidade de aumento de compras dos

consumidores, pois clientes satisfeitos com preços e produtos tendem a indicar a marca para outras pessoas.

As promoções de vendas consistem no conjunto de estratégias para atrair e conquistar clientes. As promoções podem ser um fator decisivo no destaque e permanência da empresa no mercado atual, sendo uma ferramenta de marketing importante para se adquirir competitividade e para conquistar o cliente. (DA SILVA *et al.*, 2016).

Segundo Kotler e Keller (2013, p. 570), a promoção de vendas conceitua-se como “uma variedade de incentivos de curto prazo para encorajar a experimentação ou a compra de um produto ou serviço”. O mesmo autor complementa que, as promoções de vendas podem ser realizadas objetivando resultados de curto prazo ou então para aumento de vendas em declínio. Promoção de vendas é a técnica de promover vendas. Não implica propriamente em vender, mas empenhar-se por meio de qualquer ideia ou ação para que isso aconteça. Em suma: prepara o caminho para a execução de vendas em massa. (FERRACCIÙ, 2008). Segundo Riva (2014), o termo Promoção de Vendas começa a entrar em desuso enquanto a expressão Marketing Promocional ganha força.

Ferracciù (2008) define o Marketing Promocional como: “Uma operação de planejamento estratégico ou tático combinando, sinérgica e sincronicamente, as ações de promoção de vendas com uma ou mais disciplinas das outras comunicações multidisciplinares de marketing”. Portanto, compreende-se que, ações promocionais de caráter mais estratégico, na qual impactam em vendas indiretas a curto ou longo prazo, podem ser subentendidas como marketing promocional. Assim, podemos alocar a definição de promoções de venda como um fator tático, e o marketing promocional como um fator estratégico.

O marketing promocional visa colocar ações de impacto no mercado de forma ágil e pontual, dificultando a reação do concorrente. Segundo Riva (2014), por meio da tecnologia, o marketing promocional pode ser reinventado a todo instante, e complementa que por tal motivo é mais eficiente no que se propõe a fazer. O objetivo primário deste tipo de marketing está na venda imediata, criando na organização o senso de “queremos vender hoje”. É importante que a companhia compreenda que os efeitos cruzados das promoções de varejo dependem da força do relacionamento complementar. As promoções de um produto podem influenciar positivamente a venda de outro item, gerando assim o efeito promocional.

2.2 ARMAZÉM DE DADOS

Conhecido como *Data Warehouse* (DW), se refere ao processo de coleta e pré-processamento dos dados armazenados em um ou mais bancos de dados operacionais, com o objetivo de servir de fonte para Sistemas de Suporte à Decisão, provendo uma plataforma sólida de consolidados dados históricos para análise. Ou seja, para alcançar este objetivo, reúne dados diferentes de toda a organização. (LINOFF; BERRY, 2011). O resultado desse processo é a criação de um depósito de dados, uma coleção de dados integrados, consolidados e possivelmente estruturados no tempo (dados históricos), que auxiliam e antecedem a etapa da descoberta de conhecimento em banco de dados (KDD).

Os sistemas de DW possibilitam reconciliar diversos relatórios porque eles têm a mesma origem e definições subjacentes. Segundo Linoff e Berry (2011), este tipo de sistema não apenas reduz a necessidade de explicar resultados conflitantes, mas também fornece visões consistentes dos negócios em diferentes unidades organizacionais e tempo.

Os sistemas de DW representam uma poderosa ferramenta para integrar dados operacionais dispersamente distribuídos em uma forma analítica abrangente para permitir a tomada de decisões. Em outras palavras, os DW são bancos de dados que auxiliam na tomada de decisão com base no armazenamento de dados sobre toda a empresa, de forma unificada e organizada. (SZAFIR-GOLDSTEIN; DE SOUZA, 2005).

Um DW foca na extração, modelagem e análise de dados para a tomada de decisões, onde em muitos casos necessita de ferramentas tecnológicas de suporte à estas decisões. (AGARWAL; DHAR 2014). No processo conhecido como ETL (*Extraction, Transformation and Loading*), que une e possibilita a condução dos dados de um banco de dados ou aplicação ao DW, incluem-se as funções de extração de dados, a qual podem ser realizadas em diferentes bases de dados. Posteriormente, ocorre a limpeza dos dados, detectando erros, e os corrigindo quando possíveis. Na transformação de dados também podem ocorrer conversões dos dados do formato original para o formato do DW. No processo de *loading* ocorre a classificação dos dados, verificação da integridade e resumo consolidado dos dados, para que então sejam realizadas as cargas de dados, unindo-os em um DW. O processo de limpeza e transformação de dados devem ser pontualmente realizados pois estes garantem a solidez e confiança do DW e a qualidade dos resultados das minerações.

Muitas organizações estão utilizando as informações de armazéns de dados para analisar a cesta de compras dos clientes, buscando avaliar promoções realizadas e propondo desenvolver novas que possibilitem o aumento do lucro. (HAN; KAMBER; PEI, 2012). Segundo Junior e Callegari (2002), existem três razões para se construir um DW antes do funcionamento efetivo do *Data Mining*. O primeiro é o ganho de tempo, onde com o DW evita-se o desperdício deste bem com a junção, separação e integração de dados. Segunda razão, é o fato de *Data Warehouses* conterem coleções de dados e seus históricos, e a terceira razão é o motivo do DW possuir dados detalhados e sumarizados.

2.2.1 Descoberta de Conhecimento em Banco de Dados

O valor dos dados armazenados está tipicamente ligado à capacidade de se extrair conhecimento do mais alto nível a partir deles, informação útil que sirva para apoio à tomada de decisão, e/ou para exploração e melhor entendimento do fenômeno gerador dos dados. O KDD é um processo não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados. (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Goldschmidt, Passos e Bezerra (2015, p. 3), alegam que:

"A complexidade (ou não trivialidade) do processo de KDD está na dificuldade em perceber e interpretar adequadamente inúmeros fatos observados durante a realização de processo e na dificuldade em conjugar dinamicamente tais interpretações de forma a decidir que ações devem ser realizadas em cada caso".

O processo de KDD pressupõe a existência de um conjunto de dados, do qual pode envolver n atributos, representando assim um hiperespaço (espaço n dimensional). Quanto maior for o valor de n e o número de registros disponíveis, maior o conjunto de dados a serem analisados. (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). Sucessíveis refinamentos são realizados por meio de repetitivos processamentos sobre os dados. O propósito do KDD está na busca por padrões.

O processo de KDD pode ser orientado por dois tipos de padrões, preditivos e descritivos. Padrões preditivos consistem na busca por resolução de um problema específico de prever os valores de um ou mais atributos em função dos valores de outros atributos. Nos padrões descritivos, o objetivo está em apresentar informações

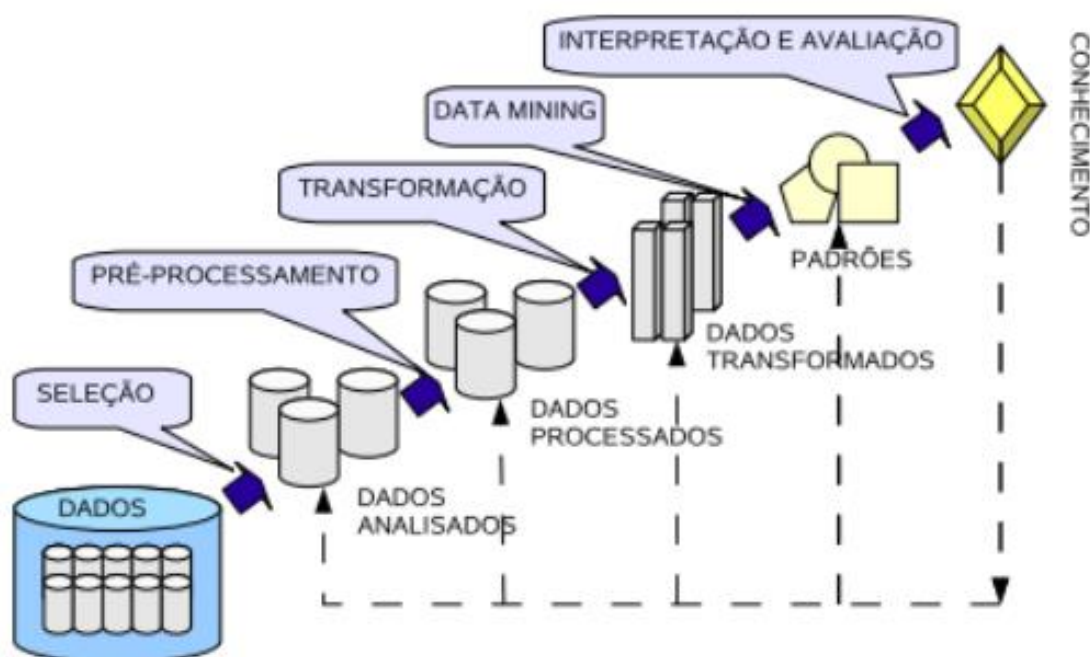
interessantes das quais um especialista do negócio possa se interessar, ao qual se encontra com objetivo deste presente trabalho. No processo de KDD, o analista de dados visualiza os dados como um todo e decide onde explorar baseado no que vê, norteado em sua própria experiência e no conhecimento fornecido pelo especialista do domínio.

Segundo Dantas (2008), o processo de KDD inicia com a análise do entendimento do domínio da aplicação e dos objetivos a serem realizados. No processo seguinte, o foco passa a ser a escolha ou seleção da massa de dados a ser minerada, podendo ser um conjunto de dados ou um subconjunto de variáveis onde a extração será realizada. A fase de *Data Cleaning* e Pré-Processamento tem por objetivo assegurar a qualidade dos dados envolvidos no KDD realizando operações básicas como a remoção de ruídos, que podem ser, por exemplo, atributos nulos. A fase seguinte consiste na Seleção e Transformação dos dados em que serão selecionados os atributos realmente interessantes ao usuário, além de transformados utilizando o padrão ideal para aplicar algoritmos de mineração.

Após a realização das fases anteriores, a Mineração de Dados (*Data Mining*) é iniciada. Esta fase é a mais importante do KDD, sendo realizada através da escolha do método e do algoritmo mais compatível com o objetivo da extração, a fim de encontrar padrões nos dados que sirva de subsídios para descobrir conhecimentos ocultos e úteis.

A Avaliação ou Pós-Processamento é a fase que identifica, entre os padrões extraídos na etapa de Data Mining, padrões interessantes ao critério estabelecido pelo usuário, podendo voltar à fase inicial para novas iterações. Ao término da avaliação, o conhecimento descoberto deverá ser implantado e incorporado ao sistema, sempre documentando e publicando os métodos, a fim de apresentar o conhecimento descoberto ao usuário. (DANTAS et al., 2008).

Figura 3 - Etapas do KDD



Fonte: Lima (2014, p. 17).

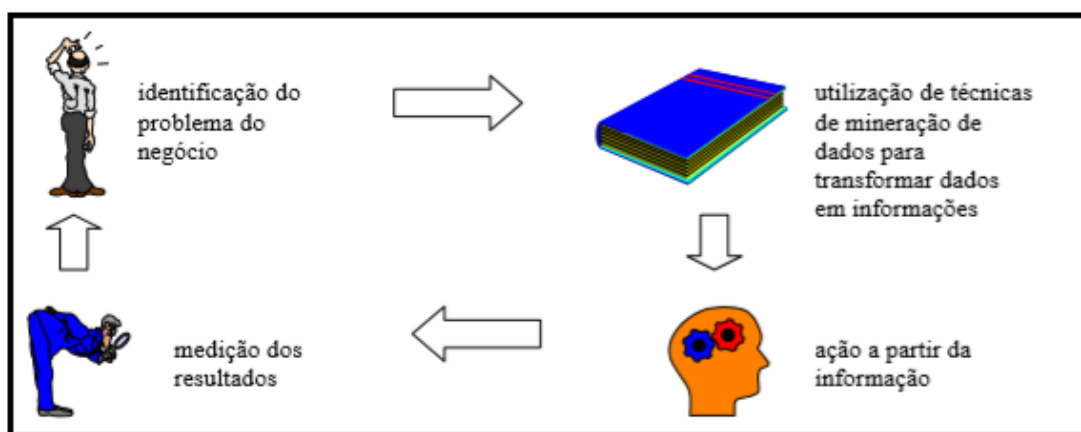
Todo modelo de conhecimento deve ser avaliado com relação ao cumprimento das expectativas definidas nos objetivos da aplicação. Os históricos sobre como os modelos de conhecimento foram gerados também se enquadram como resultados do processo de KDD. Estes históricos são de fundamental importância no controle do processo, pois permitem uma análise crítica e uma revisão das ações realizadas. (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

2.2.2 Mineração de Dados

Definida como a principal etapa do processo de KDD, a Mineração de Dados é o processo de descoberta de conhecimento significativo, como padrões, associações, alterações, anomalias e estruturas significativas de grandes quantidades de dados armazenados em bancos de dados, *data warehouses* ou outros repositórios de informações a partir do desenvolvimento de métodos e técnicas específicas de pesquisas nestas bases. (GOLFARELLI; RIZZI, 2011). A mineração de dados é um ciclo virtuoso, sendo uma etapa em um processo que requer ganho de conhecimento, ou seja, um processo contínuo que traz resultados a toda hora.

Em muitos casos a meta é construir automaticamente um modelo de *software* que prediga um valor de saída dado um conjunto de valores de entrada. Uma variedade de técnicas podem ser utilizadas, e cada uma tem sua estrutura própria. (GONÇALVES, 1999).

Figura 4 - Ciclo Virtuoso da Mineração de Dados



Fonte: Gonçalves (1999, p. 17).

Nas etapas de busca pela informação através dos dados devem ser utilizados o conhecimento sobre o domínio do assunto. Este fator auxilia na etapa de pré-processamento, podendo selecionar padrões de dados, validar valores destes dados para atributos, e quais os critérios de diferenciação sobre as informações. A partir destas diferenciações, a MD pode conter tarefas descritivas, onde o objetivo é derivar padrões (correlações, tendências, grupos, trajetórias, anomalias) que resumam os relacionamentos subjacentes nos dados. Muitas vezes são exploratórias em sua natureza e frequentemente requerem técnicas de pós-processamento para validar e explicar resultados. Uma outra tarefa da mineração de dados denomina-se preditiva, onde o objetivo destas tarefas é prever o valor de um determinado atributo baseado em valores de outros atributos. O atributo a ser previsto é comumente conhecido como a variável dependente ou alvo, enquanto os atributos utilizados para fazer a previsão são conhecidos como as variáveis independentes ou explicativas. Os métodos de mineração de dados podem ser vistos como consistindo em três componentes algorítmicos principais: representação do modelo, avaliação do modelo, e pesquisa.

Um conceito importante na etapa de DM refere-se à capacidade que determinados algoritmos têm de aprender a partir de exemplos (aprendizado indutivo). (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). Quanto a estes aprendizados, uma

das divisões é o aprendizado supervisionado, que compreende a abstração de um modelo de conhecimento a partir dos dados apresentados na forma de pares ordenados (entrada, saída desejada). A segunda parte da divisão é o aprendizado não-supervisionado, que é o processo de aprendizado que busca identificar regularidades entre os dados a fim de agrupá-los em função das similaridades que apresentam entre si, ou seja, não existe a informação de saída.

2.2.2.1 Métodos de *Data Mining*

Existem diferentes métodos de DM para encontrar respostas ou extrair conhecimento em repositório de dados. Dantas et al (2008), cita como sendo os mais importantes para o KDD: Classificação, Modelos de Relacionamento entre Variáveis, Análise de Agrupamento, Sumarização, Modelo de Dependência, Regras de Associação e Análise de Séries Temporais. Abaixo, encontra-se um breve resumo dos métodos mais importantes.

A Classificação associa ou classifica um item a uma ou várias classes categóricas pré-definidas, utilizando comumente uma técnica estatística chamada análise discriminante, objetivando envolver a descrição gráfica ou algébrica das características diferenciais das observações de várias populações, além da classificação em uma ou mais classes pré-determinadas.

Os Modelos de Relacionamento entre Variáveis associam um item a uma ou mais variáveis de predição de valores reais, consideradas variáveis independentes ou exploratórias. Regressão linear simples, múltipla e modelos lineares por transformação são as técnicas mais utilizadas para verificar a existência do relacionamento funcional entre duas variáveis quantitativas.

A Análise de Agrupamento, ou Cluster, associa um item a uma ou várias classes categóricas (ou clusters), determinando as classes pelos dados, independentemente da classificação pré-definida. Os clusters são definidos por meio do agrupamento de dados baseados em medidas de similaridade ou modelos probabilísticos, visando detectar a existência de diferentes grupos dentro de um determinado conjunto de dados e, em caso de sua existência, determinar quais são eles.

A Sumarização determina uma descrição com dispersão reduzida para um dado subconjunto no pré-processamento dos dados, frequentemente utilizadas na

análise de descobrimento de dados. Podemos citar como exemplos simples de sumarização de dados, as medidas de posição e variabilidade.

O Modelo de Dependência descreve dependências entre variáveis. Modelos de dependência existem em dois níveis: estruturado e quantitativo. O nível estruturado especifica, geralmente em forma de gráfico, quais variáveis são localmente dependentes. O nível quantitativo especifica o grau de dependência, usando alguma escala numérica.

As Regras de Associação determinam relações entre campos de um banco de dados, contribuindo para a tomada de decisão, recebendo assim, na Seção 2.4, um maior aprofundamento. Por fim, a Análise de Séries Temporais determina características sequenciais, como dados com dependência no tempo. Seu objetivo é modelar o estado do processo extraído e registrando desvios e tendências no tempo.

2.3 TAXONOMIAS DE PRODUTOS

Taxonomias são esquemas hierárquicos de classificação, nos quais os dados são organizados em estruturas de árvores. (DOTSIKA, 2009). De acordo com Martins (2006) uma taxonomia pode ser definida pelas seguintes regras:

i) cada item da taxonomia deve possuir apenas um pai. Essa característica define o processo de generalização de cada item;

ii) um item pode possuir n filhos. Este é um conceito básico e obrigatório para todas as taxonomias;

iii) se um item possuir filhos ele não pode ser excluído da taxonomia para que não seja descartada toda a sua especialização. Para realizar a exclusão, é necessário realocar os filhos definindo seus novos pais ou excluindo-os da taxonomia;

iv) os itens do nível mais alto da taxonomia (generalização máxima ou raiz) não possuem pai;

v) os itens do nível mais baixo da taxonomia (especialização máxima ou folha) não possuem filhos.

A construção de taxonomias pode ser feita de várias formas:

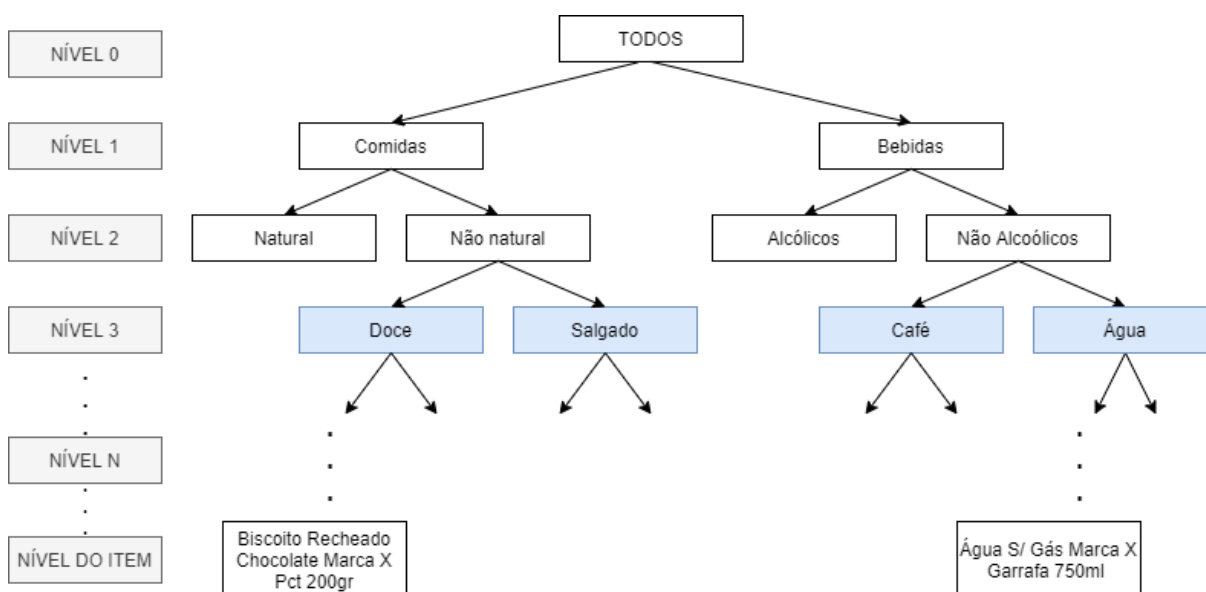
a) *Top-down*: a qual os elementos são montados a partir dos níveis superiores, onde progressivamente são adicionados à estrutura os elementos dos níveis inferiores. Apesar de se poder controlar melhor o nível de detalhe da

taxonomia resultante, determinados elementos de níveis inferiores podem ser erroneamente posicionados em níveis superiores;

- b) *Bottom-up*. A estrutura é montada a partir dos níveis compostos pelos elementos inferiores, sendo os elementos intermediários e superiores adicionados progressivamente à estrutura. Este método resulta em uma taxonomia com um alto nível de detalhe;
- c) *Middle-out*. A estrutura é montada a partir dos níveis intermediários, acrescentando-se elementos nos níveis inferiores e superiores. Esta abordagem permite conciliar os benefícios do detalhamento com os níveis superiores melhor posicionados hierarquicamente. (FILHO, 2015).

Com base no método *Bottom-up* são selecionados os produtos a nível de venda para então criar uma hierarquia customizada conforme Figura 5, ou seja, desenvolver uma árvore de categoria de produto.

Figura 5 - Árvore de Categoria de Produto



Fonte: Adaptado de (GRIVA et al., 2018).

Com a estipulação desta árvore, é determinado o nível ao qual será utilizado para o processo seguinte, que é a verificação de associações dos produtos nas compras dos clientes. Esta divisão em níveis impacta diretamente nos resultados da mineração de dados, segundo Agrawal (1995), é necessário selecionar o nível correto de granularidade da categoria de produto, pois isso pode afetar os resultados das regras de associação, e conseqüentemente todo o sistema de recomendação. A

complexidade de uma regra se refere ao número de itens que ela contém. Quanto mais tipos diferentes de itens nas transações, mais tempo leva para gerar regras de uma dada complexidade. A complexidade desejada das regras afeta a especificidade ou a generalidade dos itens. (LINOFF; BERRY, 2011). Haverá muito mais transações suportando uma determinada regra em níveis mais altos da taxonomia do que em níveis mais baixos. À medida que um item sobe na hierarquia, as quantidades de itens na análise são reduzidas. E à medida que a quantidade de itens é reduzida, conseqüentemente crescem as quantidades de regras de associações geradas na análise. Com isso, na próxima seção, são descritos os conceitos referentes a regras de associação.

2.4 REGRAS DE ASSOCIAÇÃO

O processo de extração de regras de associação foi proposto inicialmente por Agrawal et al. (1993) e representa um padrão ocorrido em combinações de itens com determinada frequência em uma base de dados. Regras de associação de acordo com a taxonomia são mais generalizadas pois retratam dados referentes as categorias ou linhas de produtos, ou seja, a partir de uma dada característica de produto, por exemplo, regata é uma roupa leve e calção de banho também, e roupa leve é um tipo de roupa.

De acordo com Silva, Peres e Boscaroli (2016) as regras de associação são técnicas para tomadas de decisões, onde estas regras são analisadas a fim de encontrar sincronização entre dados. Estas regras apresentam coocorrências de produtos nas vendas registradas em base de dados, isto significa que a ocorrência de um produto X em uma transação de venda acarreta como consequência a presença do produto Y na mesma transação com uma certa probabilidade. As regras de associação podem prever qualquer um dos atributos, não apenas uma classe especificada, e podem prever mais de um valor de atributo por vez. Geralmente são limitadas àqueles com mais de um certo nível mínimo de precisão.

O interesse é uma das medidas mais utilizadas em regras de associação, a qual pode ser dividida em objetiva e subjetiva. Medidas de interesse objetivas identificam de forma estatística a força das regras de associação, uma vez que ela é utilizada para selecionar e classificar os padrões de acordo com o seu potencial interesse para o usuário. Medidas de interesse subjetivas são baseadas em conceitos

como os de utilidade e imprevisibilidade. A utilidade de uma regra ou a possibilidade do usuário em obter proveito da sua utilização é indicador de interesse dela. Estas regras necessitam ser analisadas se são interessantes ou não, e se não implicam uma na outra. A complexidade de uma regra depende do número de itens nela contidos, embora Silva, Peres e Boscaroli (2016) afirmam que os resultados obtidos na descoberta de regras de associação são considerados de fácil interpretação, pois tratam de regras que podem ser expressas em “linguagem natural”, cuja semântica está explícita

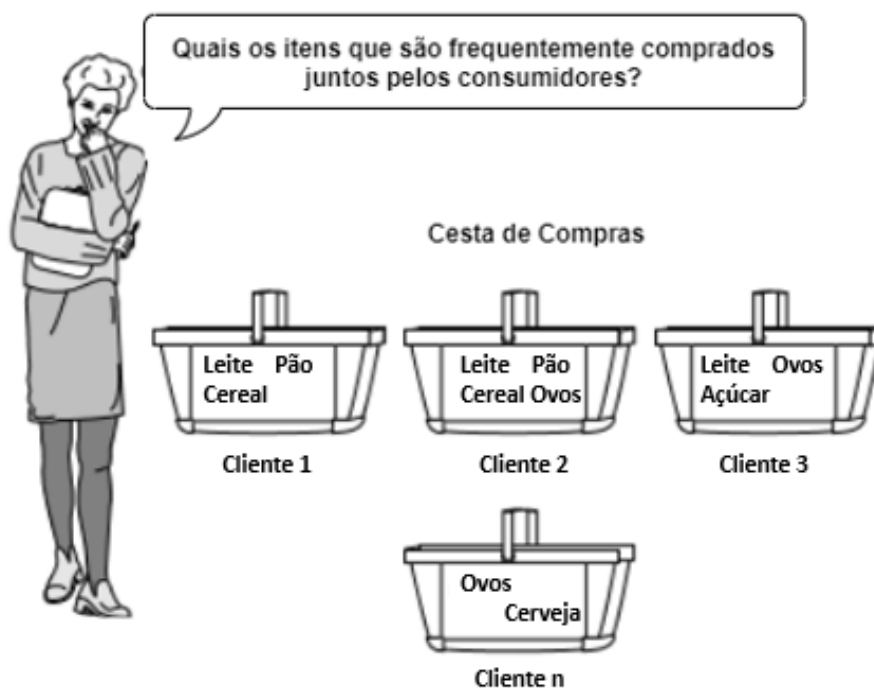
As informações podem ser usadas como base para decisões sobre atividades de marketing, como preços promocionais ou veiculações de produtos. O significado de uma regra associativa pode ser calculado na presença de dois parâmetros, a saber, suporte e confiança. Suporte é a porcentagem de combinações de itens no banco de dados. Enquanto confiança (valor de certeza) é um valor para determinar a força dos relacionamentos entre itens nas regras de associação. Maiores especificações serão atribuídas no tópico 2.4.1. Uma das aplicações de regras de associação mais comuns são as relacionadas com análise de cestas de compras (*Market Basket Analysis*).

2.4.1 Análise de Cestas de Compras

A análise de cesta de compras, conhecida por *Market Basket Analysis* é entendida como o uso de técnicas de associações para encontrar grupos de itens dos quais consumidores colocam em suas cestas de compras e estes tendem a ocorrer juntos em transações, como pode ser visualizado na Figura 6. Linoff e Berry (2011) alegam que o MBA se refere a um conjunto de problemas de negócio relacionados para compreender dados transacionais de pontos de venda. Agrawal, Imielinski, Swami (1993) demonstraram que a busca por padrões em banco de dados transacionais tiveram a sua origem na análise de cestas de compras, sendo representada por transações de uma base de dados, formadas por um identificador e por um ou mais itens presentes, que são os produtos comprados por clientes em visita à estabelecimentos comerciais. Com isso, a avaliação dos dados busca encontrar padrões no comportamento das compras que são realizadas. O processo de prospecção de possíveis padrões realizados com o uso das regras de associação é muito importantes para identificar relacionamentos úteis, que na maiorias dos casos

estão ocultos e devem ser avaliados por especialistas. (AGRAWAL; IMIELIŃSKI; SWAMI, 1993). *Market Basket Analysis* é uma técnica muito utilizada para conhecimento de informações a fim de serem úteis em campanhas de marketing, com objetivo de melhorar a eficácia e as táticas de vendas.

Figura 6 - Análise de Cesta de Compras



Adaptado de Linoff e Berry (2011).

De acordo com Fracalanza (2009), para esta técnica, a entrada de dados é um conjunto de transações correspondentes as compras de diferentes clientes, onde a cesta de compra de um cliente é composta por itens que foram adquiridos na mesma compra, sendo ignorados as quantidades e o preço de cada item. A análise da cesta de compra fornece *insights* sobre a mercadoria, informando quais produtos tendem a ser comprados juntos e quais são mais passíveis de promoção. (LINOFF; BERRY, 2011).

Os resultados desta análise são apresentados em forma de regras de associação, as quais são consideradas as representações mais populares de padrões descobertos em bases de dados, e são definidas em três características diferentes: útil, trivial e inexplicável. As regras úteis são aquelas as quais é possível tomar decisões rápidas e agir sobre as informações. (LINOFF; BERRY, 2011). Por exemplo, quando se descobre que fraldas costumam ser vendidas com alguma cerveja na sexta

feira, pode-se reposicionar um dos produtos de forma que os dois fiquem próximos ou então ofertando um desconto em um dos itens na sexta-feira, induzindo que a compra de ambos aconteça.

De acordo com Linoff e Berry (2011), os resultados triviais são regras geradas e conhecidas por qualquer pessoa familiarizada com o negócio, onde correlaciona-se dois ou mais itens anunciados juntos que raramente são vendidos separadamente. Como exemplo, em uma grande loja de varejo, um cliente comprar um grande aparelho eletrônico e um contrato de manutenção, sendo que dificilmente ocorreria somente a venda de um contrato de manutenção. Como regra, o resultado é válido e bem suportado pelos dados, mas como o objetivo da mineração de dados é encontrar padrões significativos, essa regra é inútil (LINOFF; BERRY, 2011) pois não apresenta associações desconhecidas. Problemas mais sutis – como venda de hambúrgueres e pãezinhos – podem ser inseridas na mesma categoria.

Embora seja um perigo para qualquer técnica de mineração de dados, a análise de associação é particularmente suscetível de reproduzir o sucesso de campanhas de marketing anteriores por causa de sua dependência de dados não resumidos do ponto de venda. Os resultados da análise de associação podem simplesmente estar medindo o sucesso de campanhas de marketing anteriores. (LINOFF; BERRY, 2011). Exceções às regras de alta confiança apontam para áreas nas quais as operações de negócios, a coleta de dados, e o processamento precisam ser melhoradas.

Resultados inexplicáveis parecem não ter um motivo claro para entendimento e não sugerem uma ação, como por exemplo, na inauguração de lojas de ferragens, os itens com maior índice de vendas serem os desentupidores de vasos sanitários. Com investigações mais profundas, uma possibilidade é que novas lojas de ferragens abram muitas vezes perto de subdivisões e os proprietários de imóveis necessitam estocar desentupidores, ou então é resultado apenas de uma anomalia de algumas lojas. Neste caso, qualquer que seja a causa, é duvidoso que uma análise mais densa apenas dos dados da cesta de mercado possa dar uma explicação coerente. (LINOFF; BERRY, 2011).

2.4.1.1 Conjunto de Itens Frequentes

Seja D o conjunto de transações T contidas em um banco de dados, onde cada transação T em D representa um conjunto de itens contidos em I , sendo $I = \{I_1, I_2, \dots,$

In}. (LAROSE & LAROSE, 2015). Cada transação está associada a um identificador TID. Supondo que exista um conjunto específico de itens A e outro conjunto de itens B. Então uma regra de associação assume a forma: se A, então B (ou seja, $A \Rightarrow B$), onde o antecedente A e o conseqüente B são subconjuntos adequados de I, e A e B são mutuamente exclusivos.

Os padrões de itens encontrados na análise são representados na forma de regras de associação, as quais são compostas por dois conjuntos de itens, denominados lado esquerdo (LHS) e lado direito (RHS). O primeiro representa os itens antecedentes, enquanto o segundo representa os itens conseqüentes, interpretados na forma: se há a compra do item antecedente, então há a compra do item conseqüente. (ANSELMO, 2017).

Suporte e confiança são duas medidas de interesse de regras. Eles refletem, respectivamente, a utilidade e a certeza das regras descobertas. (HAN; KAMBER; PEI, 2012). O suporte mede a quantidade ou a proporção de transações que contêm todos os itens da regra. Os suportes para uma regra de associação específica $A \Rightarrow B$ é a proporção de transações em D que contêm A e B, sendo representado pela equação abaixo.

$$\text{Suporte} = P(A \cup B) = \frac{\text{Número de transações contendo A e B}}{\text{Número total de transações}} \quad (1)$$

A confiança mede a quão boa é uma regra em prever o lado direito, comparando com que frequência o lado direito aparece quando a condição no lado esquerdo for verdadeira. (LINOFF; BERRY, 2011). A confiança da regra de associação $A \Rightarrow B$ é uma medida da precisão da regra, determinada pela porcentagem de transações em D contendo A que também contêm B. Outra maneira de dizer isso é que confiança é a razão entre o número de transações com todos os itens e o número de transações com apenas os itens "se" (*if*), como mostra a equação:

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

$$\text{Confiança} (A \Rightarrow B) = \frac{\text{Número de transações contendo A e B}}{\text{Número de transações contendo A}} \quad (3)$$

Normalmente, as regras de associação são consideradas interessantes se satisfizerem um limite mínimo de suporte e um limite mínimo de confiança. Regras de

baixo suporte não se tornam interessantes, pois raramente os itens são comprados juntos pelos clientes. A confiança demonstra a confiabilidade da relação que a regra apresenta, mostrando que, maior a confiança da regra $A \Rightarrow B$, então maior será a probabilidade de B estar presente nas transações que contenham A. (TAN; STEINBACH; KUMAR, 2009).

Para identificação de regras de associação, podem ser adotadas outras medidas de interesse, e uma das mais utilizadas junto ao suporte e confiança é o *lift*, que tem seu uso para avaliar a dependência entre um item e outro. Ou seja, procura indicar o quanto mais frequente se torna B, dado a ocorrência de A.

$$Lift(A \Rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)} = \frac{Confiança(A \Rightarrow B)}{Suporte(B)} \quad (4)$$

O *lift* é o julgamento padrão. O *lift* de uma regra $A \Rightarrow B$ é uma medida para a dependência estatística entre A e B. Se o *lift* $(A \Rightarrow B) = 1$, os itens A e B são independentes entre si. Se o *lift* $(A \Rightarrow B) > 1$, A e B se correlacionam positivamente, ou seja, são dependentes positivamente. Caso o *lift* $(A \Rightarrow B) < 1$, então A e B se correlacionam negativamente, ou seja, são negativamente dependentes e é melhor que tal regra não seja usada. Quanto maior for o *lift*, mais interessante tende a ser a regra, pois o item A está aumentando B em maior taxa.

A Tabela 1 representa algumas transações de uma loja de conveniência para simplificar as teorias mencionadas anteriormente.

Quadro 4 - Transações de uma loja de conveniência

TID	Itens
1	refrigerante, pão de queijo
2	refrigerante, salgadinho, bombom
3	água, carregador veicular, chiclete
4	pão de queijo, croissant, refrigerante
5	refrigerante, cerveja lata
6	refrigerante, pão de queijo, cerveja lata

Fonte: Adaptado de Kerkchoff.

Na tabela é possível identificar que das cinco transações citadas, três contém os produtos refrigerante e pão de queijo. Assim, o suporte para a regra {refrigerante \Rightarrow pão de queijo} é calculado pela equação:

$$\text{Suporte}(\text{refrigerante} \Rightarrow \text{pão de queijo}) = \frac{3}{6} = 50\% \quad (5)$$

Para a confiança considera-se que três das transações contém o refrigerante e o pão de queijo juntos, e cinco das transações contém o refrigerante. Utiliza-se a equação abaixo para identificar a confiança da regra {refrigerante \Rightarrow pão de queijo}:

$$\text{Confiança}(\text{refrigerante} \Rightarrow \text{pão de queijo}) = \frac{3}{5} = 60\% \quad (6)$$

O valor da confiança explicita que 60% dos consumidores dos quais compram pão de queijo, também compram refrigerante.

Se avaliado o *lift* neste caso, encontra-se a equação 7:

$$\text{Lift}(\text{refrigerante} \Rightarrow \text{pão de queijo}) = \frac{0.6}{0.5} = 1,2 \quad (7)$$

Conforme teoria apresentada, refrigerante e pão de queijo estão positivamente correlacionados, sendo dependentes positivamente.

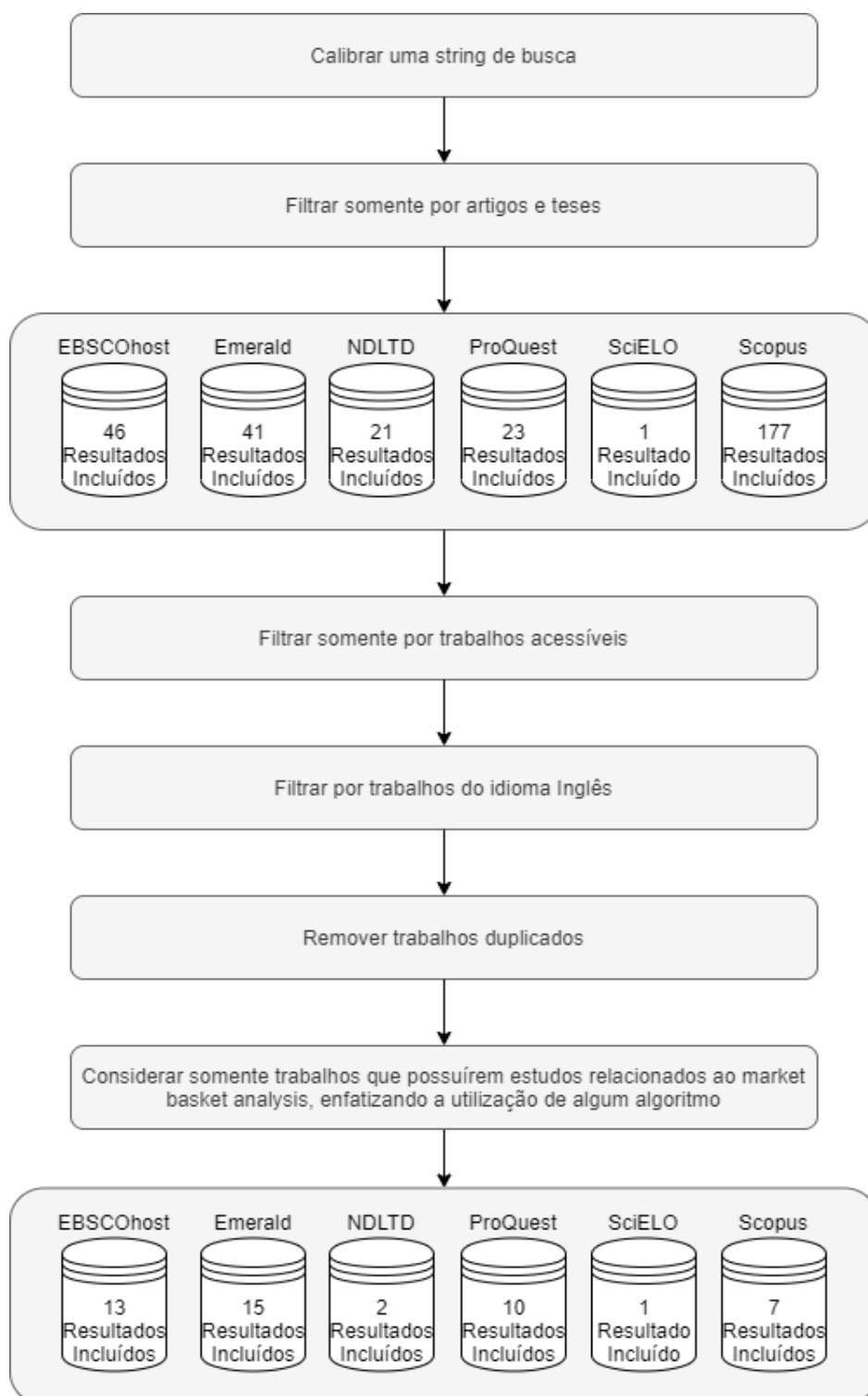
Silva, Peres e Boscaroli (2016) reforçam que quanto menor for o suporte e confiança estipulados para a análise, maior é o trabalho do algoritmo na busca dos *itemsets*. Com base nos exemplos citados por Silva, Peres e Boscaroli (2016), neste trabalho serão considerados como limiares mínimos, 1% e 25% para suporte e confiança, respectivamente, dos quais serão inseridos na elaboração do método de trabalho conforme explicitado no tópico 3.3, que será executado vide algoritmo a ser definido na próxima secção.

2.5 ALGORITMOS DE KDD

Para a realização das tarefas de mineração de dados, existem inúmeros algoritmos que podem ser utilizados, assim como diferentes algoritmos podem ser utilizados para o processo de análise de cesta de compras, como: *Apriori*, *FP-Growth*, *Basic*, *Genex*, entre outros. (HAN, KAMBER, PEI, 2012). Para a seleção do algoritmo a ser posteriormente utilizado, foi realizada uma busca na literatura, considerando o fator de utilização do algoritmo juntamente ao processo de *Market Basket Analysis*. Para esta pesquisa, realizou-se um teste no *Google Acadêmico* a fim de encontrar a *string* ideal de busca. Dada a falta de estudos na língua portuguesa, orientou-se a busca pela *string* na linguagem inglesa. Realizados os testes, a *string* inserida foi

“*data mining*” AND “*market basket analysis*” AND “*algorithm*” para que a partir desta fossem encontrados artigos e teses relevantes. Inicialmente, conforme ilustrado na Figura 7, foram encontrados 309 trabalhos. Sobre estes resultados, buscou-se pela melhor definição dos resultados dados os filtros inseridos.

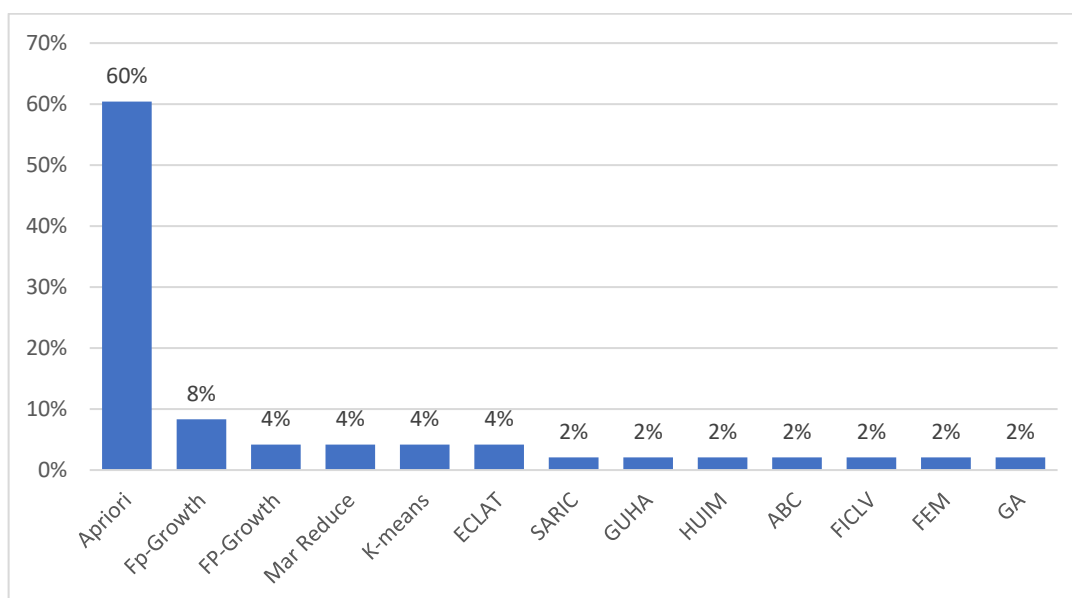
Figura 7 - Etapas de Busca por Algoritmos de KDD



Fonte: Elaborado pelo autor.

A partir da busca, foi possível identificar que dentre os algoritmos de *Market Basket Analysis*, em 60% dos trabalhos verificados havia a utilização do algoritmo *Apriori*. No Gráfico 3 é possível identificar os resultados sobre a pesquisa realizada.

Gráfico 3 - Algoritmos encontrados na pesquisa



Fonte: Elaborado pelo autor.

Apriori é um algoritmo seminal proposto por R. Agrawal e R. Srikant em 1994 para minerar conjuntos de itens frequentes para regras de associações booleana. O nome do algoritmo é baseado no fato de que o algoritmo usa conhecimento prévio de propriedades frequentes do conjunto de itens. (HAN; KAMBER; PEI, 2012). A ideia básica do algoritmo *Apriori* é encontrar todos os conjuntos frequentes na transação, e a necessidade frequente desses conjuntos frequentes é maior ou igual ao grau de suporte mínimo do conjunto. Em seguida, as fortes regras de associação são geradas pelo conjunto de frequências, que deve ser maior que o suporte mínimo e a confiança mínima estipulada.

O algoritmo *Apriori* emprega uma abordagem iterativa conhecida como pesquisa em nível, em que os conjuntos de itens k são usados para explorar os conjuntos de itens $(k + 1)$. Primeiro, o conjunto de itens de um item frequente é encontrado para digitalizar o banco de dados para acumular um montante para cada item e coletando os itens que satisfazem o suporte mínimo, caso não satisfaçam o suporte mínimo estipulado, o conjunto de itens é eliminado. O conjunto resultante é indicado por L1. Em seguida, L1 é usado para encontrar L2, o conjunto de conjuntos

de dois itens frequentes, que é usado para encontrar L3 e assim por diante, até que não sejam encontrados conjuntos de itens, ou seja, até que não encontre mais os valores mínimos para o suporte estabelecido. (HAN; KAMBER; PEI, 2012).

Em cada iteração, o algoritmo avalia o suporte dos dados novamente, acrescentando os itens que apresentam suporte igual ou superior ao estabelecido, no conjunto pré-definido na passagem anterior, assim, gerando novos conjuntos de itens. A descoberta de cada conjunto de itens requer uma varredura completa do banco de dados.

Realizada a verificação dos conjuntos de itens, o algoritmo gera o conjunto de grupos frequentes que tenham atendido o suporte mínimo, posteriormente, realiza a etapa seguinte que consiste na geração de regras de associação levando em consideração o valor mínimo de confiança estipulado. Assim, é possível identificar e apresentar as regras fortes do respectivo conjunto de dados, ou seja, associações que atingem o suporte mínimo e confiança mínima.

O objetivo do referencial teórico foi apresentar os principais tópicos que norteiam essa pesquisa. O próximo capítulo abordará os procedimentos metodológicos utilizados neste estudo.

3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo são expostos os procedimentos utilizados no desenvolvimento desta pesquisa, descrevendo o método científico, método de pesquisa, método de trabalho, e as técnicas para coleta e análise de dados.

3.1 MÉTODO CIENTÍFICO

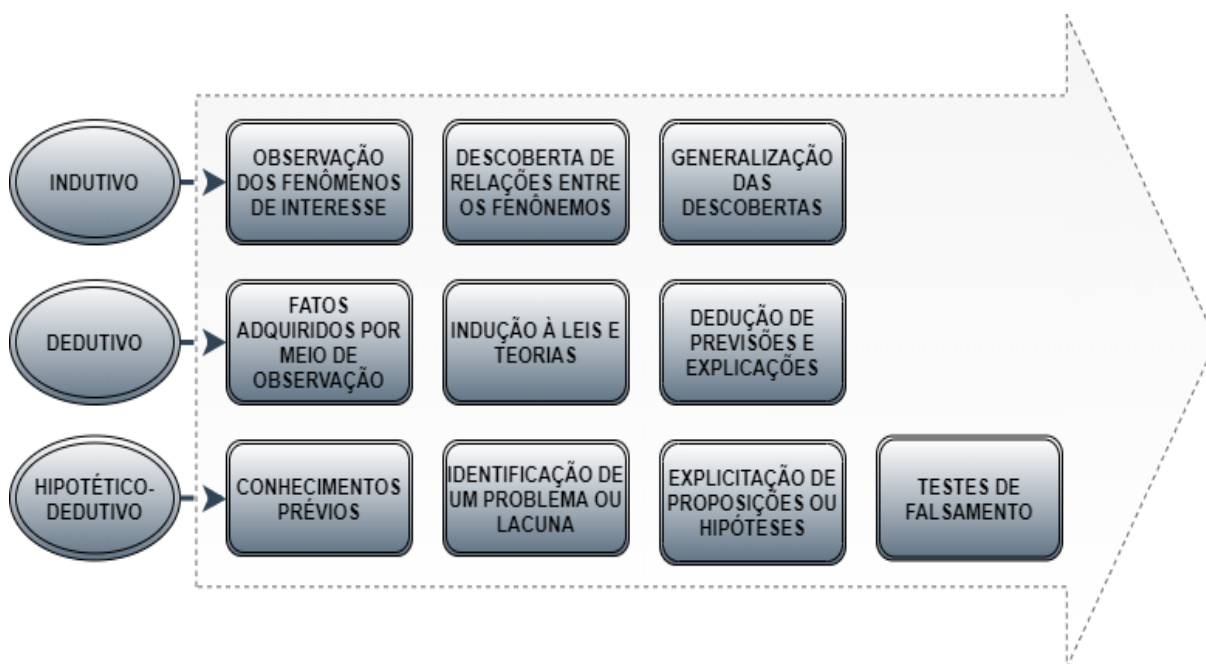
A pesquisa pode ser definida como uma investigação sistemática, cujo objetivo central costuma ser o desenvolvimento ou refinamento de teorias, e em alguns casos, a resolução de problemas. A pesquisa é muitas vezes necessária diante da falta de informação adequada e sistematizada para responder a um determinado problema. (DRESCH; LACERDA; ANTUNER JR, 2015).

O método científico é o conjunto de processos ou operações mentais que devem ser empregadas na investigação. É a linha de raciocínio adotada no processo de pesquisa. (SILVA; MENEZES, 2005). Sendo assim, o método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar os objetivos por meio de conhecimentos válidos e verdadeiros, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões de cientistas. (MARCONI; LAKATOS, 2010). Os métodos científicos podem ser divididos em Indutivos, Dedutivos e Hipotético-Dedutivos.

O método indutivo foca na construção de conhecimento através da observação de determinado objeto de pesquisa. Já o método dedutivo refere-se à explicação ou previsão de fenômenos com base em leis e teorias universais. E o método hipotético-dedutivo possui base no conhecimento prévio, identifica um problema e testa hipóteses que resultarão em previsões e explicações. (DRESCH; LACERDA; ANTUNER JR, 2015). As fases de cada método encontram-se na Figura 8.

Nesta pesquisa utiliza-se o método dedutivo, uma vez que se caracteriza por ser orientado pelo raciocínio lógico, sendo baseado em leis e teorias para desenvolvimento do conhecimento. Este método utilizado será apresentado na subsecção a seguir.

Figura 8 - Métodos Científicos



Fonte: Adaptado de Dresch, Lacerda e Antunes Jr (2015).

Para escolha do método de pesquisa adequado para cada estudo, o pesquisador deve considerar alguns aspectos, tais como: (i) se o método assegura que o problema de pesquisa será respondido/endereçado; (ii) se o método é legitimado pela comunidade científica; (iii) se o método evidencia, de maneira clara, os procedimentos que devem ser adotados para a condução da pesquisa. (LACERDA et al., 2013). Por meio da observação destes elementos, o intuito é garantir a robustez da pesquisa e seus resultados.

3.2 MÉTODO DE PESQUISA

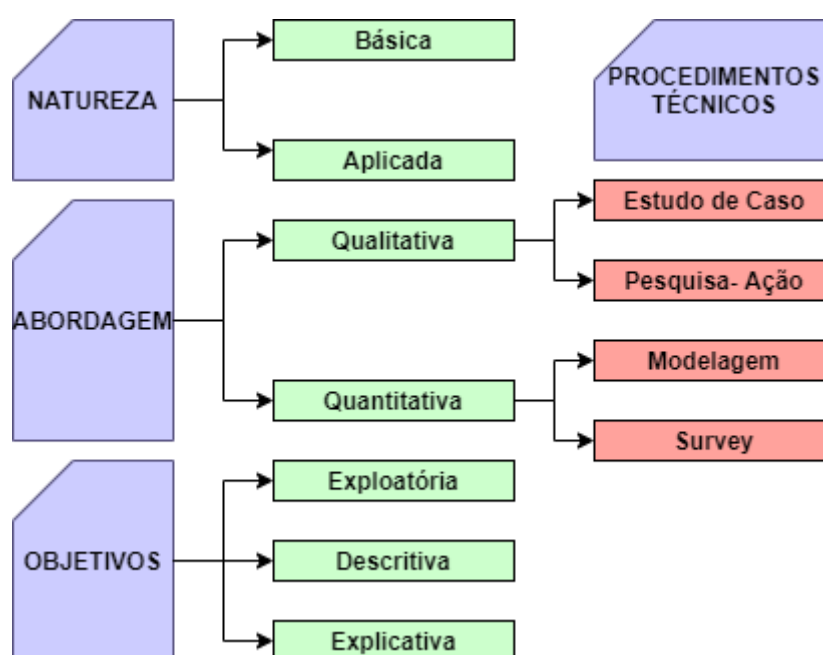
Pode-se definir pesquisa como o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa é requerida quando não se dispõe de informação suficiente para responder ao problema, ou então quando a informação disponível se encontra em tal estado de desordem que não possa ser adequadamente relacionada ao problema. (GIL, 2017).

A definição do método de pesquisa torna-se especialmente importante pois auxilia o pesquisador a garantir que a sua investigação irá responder o problema de pesquisa. Além disso, a adequada aplicação do método de pesquisa possibilita que a investigação seja reconhecida pela comunidade científica como uma pesquisa séria e

válida para a área de interesse. (DRESCH; MIGUEL, 2015). Silva e Menezes (2005) concluem que a pesquisa consiste em ser um procedimento reflexivo e crítico de busca de respostas para problemas ainda não solucionados.

A pesquisa desenvolve-se ao longo de um processo que envolve inúmeras fases, desde a adequada formulação do problema até a satisfatória apresentação dos resultados. (GIL, 2017). Na figura 9 são mostradas quanto as diferentes classificações que uma pesquisa possui.

Figura 9 - Classificações de pesquisa



Fonte: Adaptado de Miguel (2010).

Uma pesquisa pode ser classificada como, pesquisa básica ou pesquisa aplicada. A pesquisa básica, tem como objetivo principal garantir o progresso científico, sem preocupação de utilizar o conhecimento gerado na prática. A pesquisa aplicada tem como principal interesse, que seus resultados auxiliem os profissionais na solução de problemas do dia a dia. (DRESCH; LACERDA; ANTUNER JR, 2015). No contexto deste trabalho, utiliza-se a pesquisa aplicada, a se considerar como objetivo, a geração de conhecimentos através da aplicação da ferramenta DM para compreender itens vendidos em conjunto em uma loja de conveniência em determinados dias da semana.

Quanto à abordagem, a pesquisa pode ser subdividida em *Qualitativa* ou *Quantitativa*. A considerar que opiniões e informações podem ser traduzidas em

números a fim de classificá-las e analisá-las, pesquisas quantitativas apropriam-se do uso de recursos e de técnicas estatísticas. Pesquisas qualitativas objetivam-se na interpretação de fenômenos e na atribuição de significados aos mesmos (SILVA; MENEZES, 2005). Essa pesquisa possui abordagem quantitativa, uma vez que dados numéricos são apresentados e analisados.

Do ponto de vista dos objetivos, de acordo com Silva e Menezes, podem classificar-se de acordo como elaborado no Quadro 5.

Quadro 5 - Classificação dos objetivos da pesquisa

Tipo de Pesquisa	Descrição
Exploratória	Objetiva tornar o objeto de estudo mais familiar visando à construção de hipóteses
Descritiva	Objetiva descrever as características de determinada população ou fenômeno ou o estabelecimento de relações a partir de variáveis.
Explicativa	Objetiva identificar os fatores que determinam ou contribuem para a ocorrência dos fenômenos, aprofunda o conhecimento da realidade porque explica a razão, o “porquê” das coisas

Fonte: Elaborado pelo autor a partir de Silva e Menezes (2005).

Esta pesquisa pode ser classificada como descritiva, onde por meio das variáveis analisadas, podem gerar relações que tendem a ocorrer no estabelecimento, a fim de propor cenários que possam trazer maior retorno financeiro futuramente.

Quanto ao método de pesquisa, Dresch, Lacerda e Antunes JR (2015) destacam quatro utilizados nas pesquisas relativas à Engenharia de Produção, descritos no Quadro 6.

De acordo com Neto e Pureza (2012), a gestão de sistemas de produção de bens ou serviços na Engenharia de Produção envolvem um conjunto de decisões acerca das atividades desenvolvidas nos níveis estratégicos, táticos ou operacionais. Os autores também afirmam que a utilização de modelos possibilita um maior entendimento dos cenários envolvidos, auxiliando no processo de tomadas de decisões gerenciais. Um modelo deve ser suficientemente detalhado para captar elementos essenciais e representar o sistema real. Por outro lado, deve ser suficientemente simplificado (abstraido) para ser tratável por métodos de análise e resolução conhecidos.

Quadro 6 - Métodos de Pesquisa

Tipo de Pesquisa	Descrição
Estudo de Caso	Busca melhor compreender os fenômenos em seu contexto real, através de descrições detalhadas baseada em fontes de dados. Fundamentam-se na comparação dos dados coletados, buscando identificar o surgimento de categorias teóricas que possam ainda servir de base para a preposição de novas teorias.
Pesquisa-ação	Objetiva a resolução de um problema coletivo. Os pesquisadores e participantes representativos da situação ou do problema estão envolvidos de modo cooperativo ou participativo. É um método fundamentalmente empírico com uma abordagem qualitativa.
Survey	Objetiva desenvolver conhecimento em uma área específica, através de coleta e análise de dados, para avaliar o comportamento das pessoas e ou o ambiente em que se encontram. Entre seus objetivos está a geração de dados confiáveis que possibilitem uma análise estatística robusta, sendo assim, possui uma abordagem quantitativa.
Modelagem	Busca o melhor entendimento dos problemas, através de representações simplificadas da realidade, permitindo compreensão do ambiente estudado.

Fonte: Adaptado de Dresch, Lacerda e Antunes JR (2015).

Neto e Pureza (2012) classificam os modelos como concretos ou abstratos. Por modelos concretos podemos associar a utilização de maquetes, protótipos e materiais físicos que representem em determinada escala, a realidade. Modelos abstratos apropriam-se de técnicas analíticas, aplicação de matemática, estatística, e técnicas experimentais como a simulação do sistema real. (NETO; PUREZA, 2012). Neste trabalho, é utilizado o modelo abstrato, a ser considerado a existência de abordagem quantitativa, que compreende um conjunto de variáveis de controle que variam em um domínio específico e variáveis de desempenho que inferem a qualidade das decisões obtidas a partir de relações causais e quantitativas definidas entre essas variáveis por meio da utilização de ferramentas analíticas na avaliação dos resultados.

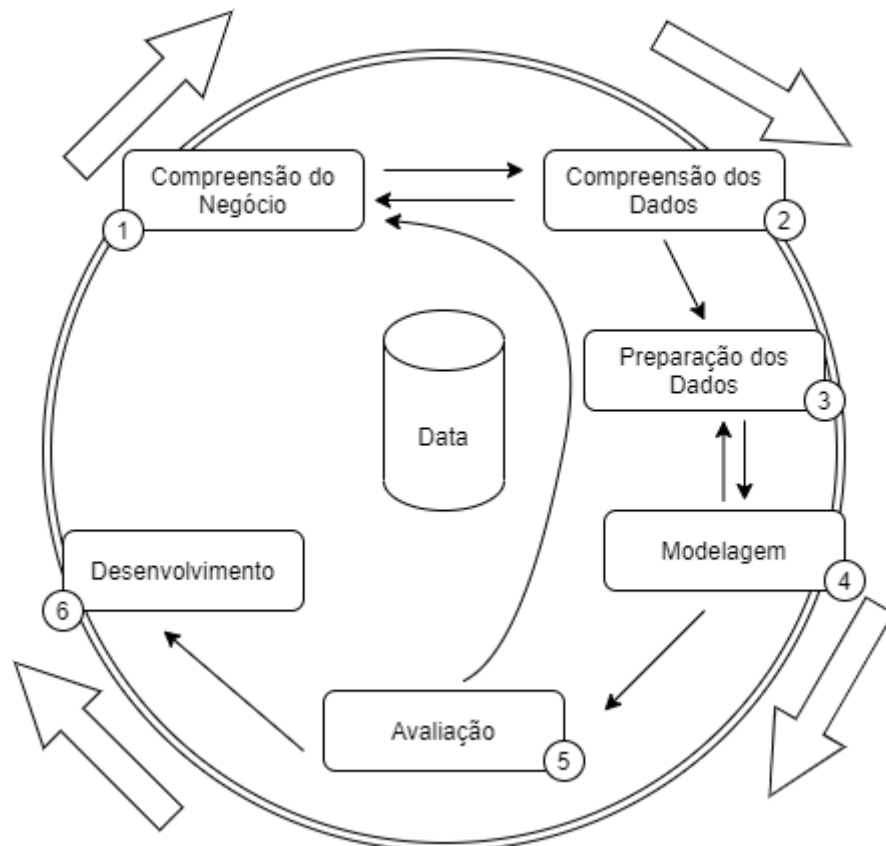
De acordo com Neto e Pureza (2012), as pesquisas baseadas em modelos quantitativos podem ser classificadas em axiomáticas ou empíricas. As pesquisas axiomáticas são dirigidas a problemas idealizados, onde procura-se obter soluções para o modelo. Por outro lado, a pesquisa empírica é dirigida por descobertas e medidas interpretativas, com foco em aplicar modelos científicos, obtidos através de pesquisas teóricas, em processos reais, com a validação dos modelos científicos obtidos pela pesquisa teórica quantitativa, e com a utilidade e o desempenho das

soluções resultantes. (NETO; PUREZA, 2012). A partir destas subdivisões, este trabalho pode ser definido como pertencente a classe empírica, uma vez que objetiva modelar um sistema para compreender as relações existentes entre as compras efetuadas e os dias da semana que ocorrem.

Pesquisas empíricas podem ser classificadas em descritivas ou normativas. Pesquisas empíricas do tipo descritiva visam criar modelos que levam a compreensão de processos, enquanto as de tipo normativa, visam o desenvolvimento de políticas, estratégias e ações que melhorem a situação corrente, baseada em modelos que prescrevam uma decisão para o problema. (NETO; PUREZA, 2012). Dado estas classificações, este trabalho possui caráter normativo, por estar diretamente atribuído a estratégias e ações futuras.

Para a condução deste trabalho, utilizou-se como base as etapas do Cross-Industry Standard Process for Data Mining (CRISP-DM) propostas por Larose e Larose (2015). Na Figura 10 encontram-se os passos do método.

Figura 10 - Método CRISP-DM



Fonte: Adaptado de Larose e Larose (2015).

O CRISP-DM fornece um processo padrão não proprietário e disponível gratuitamente para ajustar a mineração de dados à estratégia geral de solução de problemas de uma unidade de negócios ou pesquisa. (LAROSE; LAROSE, 2015). Esta metodologia ocorre em eventos distintos e reúne as melhores práticas para que o DM seja o mais produtivo e eficiente possível propondo modelos de melhorias ou solução de problemas.

O método define o ciclo de vida de um projeto em 6 fases:

- a) Compreensão do negócio;
- b) Compreensão dos dados;
- c) Preparação dos dados;
- d) Modelagem;
- e) Avaliação;
- f) Desenvolvimento.

No subcapítulo seguinte, o método de trabalho a partir desta modelagem será detalhado, ou seja, como a pesquisa foi desenvolvida.

3.3 MÉTODO DE TRABALHO

O método de trabalho consiste na sequência de passos lógicos que o pesquisador seguirá para atingir os objetivos propostos, onde o método de pesquisa selecionado deve ser desdobrado e detalhado com fundamento no método científico determinado. O método de trabalho deve conter a descrição das etapas e as técnicas de coleta e análise de dados, explicitando as razões que motivam tais escolhas. (DRESCH; LACERDA; ANTUNER JR, 2015). A partir do método selecionado para a condução do modelo, são detalhadas as etapas deste modelo da Figura 10.

- Etapa 1 - necessário esclarecer o problema por meio da questão que norteia a pesquisa a fim de definir os objetivos do sistema. Foi identificado a possibilidade de um estudo que visa compreender a relação na compra conjunta de itens em uma loja de conveniência juntamente aos dias em que estas ocorrem com maior frequência. Assim, a partir dos combos gerados, diferentes promoções diárias, em um dos itens do combo poderiam ser realizadas, a fim de estimular a venda do outro item tendente.
Com a finalidade de buscar métodos que sustentem os futuros resultados, iniciou-se com o estudo da literatura através dos bancos de dados, considerando artigos, dissertações e livros sobre os temas relacionados. Por meio da pesquisa buscou-se identificar trabalhos sobre varejo, em especial varejo de lojas de conveniência, e *Data Mining*, buscando evidenciar trabalhos em que a ferramenta auxiliava para os processos de tomadas de decisões. A partir dessa pesquisa desenvolveu-se as justificativas e os objetivos.
- Etapa 2 – Refere-se a compreensão dos dados. É aqui que começa de fato o trabalho de mineração de dados, pois o profissional deve ser capaz de identificar quais são os dados importantes para a resolução do problema e analisar sua qualidade. (COMO, 2018). Esta etapa consiste na coleta dos dados da loja de conveniência, das quais foram extraídos dados das transações do ano de 2019. Nessa etapa, caso não sejam identificados os dados, deve ser retornado à etapa anterior para reavaliação dos objetivos.
- Etapa 3 - cobre todos os aspectos da preparação do conjunto de dados final, que deve ser usado para as fases subsequentes, desde os dados iniciais, brutos e sujos. (LAROSE; LAROSE, 2015). Dentro desta etapa realizou-se o tratamento dos dados, onde são selecionados somente itens diretamente relacionados a loja de conveniência, sendo excluídas operações relativas à pista de abastecimento de combustível. Posterior ao tratamento, são avaliadas as categorias dos produtos e selecionado o nível de categoria de produto a ser utilizado na análise.
- Etapa 4 - contempla a construção do modelo, o *software RStudio* foi selecionado, pois além de ser um software gratuito com recursos completos de ciência de dados, é construído sobre a linguagem R, linguagem de conhecimento prévio do autor, e que possui todos os pacotes abertos relacionados a ciência de dados. De acordo com as especificações do *software*,

é gerado um modelo de *Market Basket Analysis*. Primeiramente, é realizado um teste para validação do modelo, para, se existente alguma inconsistência ou *bug* no modelo, o mesmo deve ser corrigido. Para este teste, é utilizado o conjunto de dados denominado *Online Retail*, disponível em diferentes comunidades que englobam a ciência de dados, para que assim o modelo seja validado. Validado o modelo, os valores ideais de suporte e confiança são inseridos conforme definidos no tópico 2.4.1.1 com base em Silva, Peres e Boscaroli (2016). Então ocorrem o carregamento dos dados tratados e o processamento do modelo. Em caso de alguma divergência, deve ser retornado ao passo anterior.

- Etapa 5 - ocorrem as análises de dados, consistindo na avaliação das associações de produtos geradas, realizando a geração das cestas de compras de acordo com as medidas de interesse para que então sejam relacionados aos dias da semana. Este relacionamento junto aos dias da semana é realizado via cálculos estatísticos proporcionais para encontrar em qual dia da semana um combo tende a ter maior saída. Esses modelos devem ser avaliados quanto à qualidade e eficácia, antes de implantá-los para uso no campo. (LAROSE; LAROSE, 2015). Caso o modelo não performe conforme o esperado, deve ser retornado à primeira etapa. Após a execução do modelo, as análises das regras geradas relacionadas aos dias da semana devem ser documentadas e os insights gerados devem ser apresentados para os envolvidos nas tomadas de decisão.
- Etapa 6 - consiste em definir uma estratégia para a implantação do projeto, ou seja, executar e manter o modelo no dia a dia para acompanhamento da análise com intuito de aprimorar o operacional do negócio e os ganhos do estabelecimento. Esta etapa não será considerada no estudo deste trabalho.
Na próxima secção, são introduzidas as coletas de dados e as breves introduções analíticas.

3.4 COLETA DE DADOS

Com a finalidade de analisar a cesta de compras, sobre a qual deseja-se descobrir as regras de associação juntamente aos dias da semana, foram coletadas todas as transações da loja de conveniência realizadas no caixa do estabelecimento.

O período estipulado junto ao proprietário do negócio foi de 01/01/2019 até 31/12/2019. Os dados foram coletados juntamente ao responsável do estabelecimento, definido este período dado a linearidade do faturamento entre os meses de um ano completo. Não se buscou avançar ao período do ano de 2020 por ser um período atípico, do qual o mercado teve de lidar com uma pandemia, que trouxe consigo efeitos socioeconômicos.

Os dados do estabelecimento, caracterizam-se como dados estruturados, onde linhas armazenam uma ocorrência de um evento caracterizado por um conjunto de colunas que representam características que descrevem um exemplar (instância) daquele evento. (SILVA; PERES; BOSCARIOLI, 2016). Em grande parte, estes dados são caracterizados, segundo Silva, Peres e Boscaroli (2016) como resultantes de processos de geração de dados inerentes a sistemas transacionais, onde segregam-se em conjuntos de tabelas que possuem relação entre si.

De acordo com Han, Kamber e Pei (2012), os dados tendem a ser incompletos e inconsistentes, onde uma rotina de limpeza de dados objetiva preencher os valores ausentes, suavizar os ruídos e corrigir inconsistências nos dados. Estas limpezas de dados fazem parte do processo de transformação de dados, para que a base de dados possua dados úteis e de qualidade oferecendo condições ao processo de descoberta de conhecimento, pois conforme Han, Kamber e Pei (2012) reforçam, dados com pouco feitiço tendem a trazer resultados ruins e de baixa qualidade. Para a realização deste processo, serão realizadas as atividades de limpeza de dados, integração e transformação dos dados.

3.5 ANÁLISE DOS DADOS

Conforme Dresch; Lacerda; Antunes Jr (2015), a análise de dados busca apresentar a relevância de informações coletadas com o objetivo de dar sentido ao conjunto de informações coletadas. Neste trabalho, após a coleta de dados, os dados e fatores a serem analisados são: nível da categoria a ser utilizado no modelo, o resultado da validação do modelo, as regras de associação geradas a partir das transações do ponto venda em estudo, e a partir do resultado, relacionar estes aos dias da semana ao qual ocorrem, para buscar compreender diferentes fenômenos acontecidos no dia a dia.

Inicialmente, após a coleta de dados e prévia análise, verificou-se a necessidade do escalamento dos produtos em níveis, uma vez que diversos produtos se assemelham, diferenciados somente por marca. Esta escala definida será tratada no item *4.3.3 Transformação dos Dados*.

Os dados utilizados em uma avaliação prévia do modelo foram submetidos a mesma análise a ser realizada com os dados reais do estudo. Por meio da aplicação do algoritmo *Apriori* e suas regras especificadas, a certificação do modelo teste se dá quando o mesmo converge ao resultado esperado, as regras de associação, ou seja, o realce da tendência geral de combos gerados sobre a base de dados dos pontos de venda. Nesta validação também é rastreada possíveis erros apresentados no modelo ou necessidade de alteração de algum parâmetro.

Após a seleção do nível ideal da categoria selecionada, o arquivo, transformado para o formato “*csv*” será importado para o software *RStudio* para a realização da mineração dos dados, com o objetivo de encontrar a associação mais frequentes entre os itens do ponto de venda.

De acordo com Han, Kamber e Pei (2012), somente a utilização de suporte e confiança como medidas de interesse, não trazem o correto respaldo quanto as regras interessantes. Uma medida de correlação é necessária, e neste trabalho apropria-se do *lift*, que é uma medida que busca identificar as dependências entre os itens, conforme explicitação no subtópico *2.4.1.1 Conjunto de Itens Frequentes*. No modelo proposto, foi utilizado um comando ao qual, filtra-se através do *lift*, em ordem decrescente, com intuito de ressaltar os itens que possuem maior dependência. Silva, Peres e Boscaroli (2016) salientam que itens que associações com $lift = 1$ indicam que não existem relação entre si, portanto, devem ser desconsideradas.

Para a análise dos dados, são considerados os combos gerados pelo modelo do *software*, e a partir disso, será desenvolvido um modelo básico de análise gráfica no Microsoft Power BI, visando analisar quais os dias da semana em que estas vendas associadas possuem maior aderência. Após uma análise inicial dos dados, observou-se a possibilidade de realização de diferentes abordagens mais profundas e de auxílio ao gestor do estabelecimento, onde uma destas é a verificação dos períodos do dia com maior frequência das associações geradas no modelo.

Após a apresentação dos procedimentos metodológicos, no próximo capítulo são expostas as aplicações do modelo proposto e seus resultados.

4 ANÁLISE DOS RESULTADOS

Neste capítulo, em primeiro momento é apresentada a organização em estudo, e na sequência são expostas as demais etapas da metodologia CRISP-DM apresentada anteriormente. Ou seja, explicita-se a compreensão e preparação dos dados, a modelagem, que consiste em desenvolvimento do modelo, validação deste e a utilização do modelo por meio dos dados, após isso, a etapa de avaliação é executada. Nesta etapa, os resultados são obtidos e as relações junto aos dias da semana e aos períodos do dia são analisadas. A análise quanto a relação das regras geradas e o dia da semana foi realizada a parte, em um BI. Optou-se em não inserir a informação do dia da semana no modelo *A priori*, pois se assim tratado, este dado seria apresentado como um produto, o qual não é objetivo da análise. A sexta etapa conforme metodologia selecionada, não será aplicada pois este trabalho contém apenas a análise a partir dos dados, e por hora, não a implementação do modelo no dia a dia do negócio.

4.1 ORGANIZAÇÃO ANALISADA

Nesta seção, são apresentadas as principais informações sobre a organização analisada. A pesquisa foi realizada na loja de conveniência de um posto revendedor varejista de combustíveis, da região do Vale do Rio dos Sinos, no estado do Rio Grande do Sul. A empresa possui mais dois estabelecimentos na região, porém, esta análise teve o foco somente à unidade principal, que é a de maior prestígio na região. Em sua estrutura física, além da divisão da loja de conveniência, a empresa possui um setor denominado como pista, que caracteriza o abastecimento de combustíveis, e pelo fornecimento de serviços como lavagem, troca de óleo e lubrificantes. Como mencionado anteriormente, o foco do estudo concentra-se somente em análise dos dados do setor da loja de conveniência

A unidade conta com um quadro de 35 funcionários, dos quais, 8 são exclusivamente da loja de conveniência. Dividida em 100 metros quadrados, e na busca do bem-estar e fidelização dos clientes, a loja possui um ambiente interno para usufruto dos clientes, já visando a venda de itens de consumo imediato, oferecendo atendimento das 06:00 horas da manhã às 23:00 horas.

As atividades relacionadas à venda, compra e estoque de produtos é realizada por meio de um *software* muito comum a este tipo de negócio, sendo possível o administrador conter os dados de controle do negócio.

A empresa está vinculada a uma grande distribuidora do mercado, portanto, possui a bandeira desta rede exibida como sua marca. Assim, agentes assessores desta rede frequentemente prestam suporte para evolução do negócio. Através desta assessoria, o gestor do negócio possui conhecimento sobre volumetria e faturamento dos itens (nível 7 da árvore de categorias) com maior procura, e também referente as categorias dos produtos (nível 2 da árvore de categorias), porém, desconhece a informação de quais os itens que costumam ser vendidos juntos e qual sua relação frente ao dia da semana. Portanto, a partir da aplicação da ferramenta de Data Mining sobre os dados coletados diariamente, este gestor é capaz de melhor avaliar as relações ocorridas em suas vendas, tomando decisões e posteriores ações sempre visando o principal objetivo de uma empresa, o lucro.

Na próxima secção é apresentado o processo de compreensão dos dados.

4.2 COMPREENSÃO DOS DADOS

Os dados foram coletados conforme período estipulado e apresentado no tópico 3.4 *Coleta de Dados*. As transações do ponto de venda em análise encontram-se no sistema de gestão do estabelecimento, ao qual apenas o gestor do negócio possui acesso aos relatórios. Os dados foram exportados em formato “xlsx”. Por limitações do sistema, foi necessária a extração de um arquivo referente a cada período mensal, para posterior envio ao autor, via e-mail, para realização das análises. Dada o formato em que se encontravam os arquivos, mesmo sendo provenientes da mesma fonte, possuem uma forma inadequada para serem submetidos a algoritmos. (SILVA; PERES; BOSCARIOLI, 2016).

Nesta etapa, a partir da verificação dos dados existentes, surge como possibilidade uma posterior análise relacionada aos períodos do dia, o que tende a vir trazer mais informações do que apenas a realização da análise juntamente ao dia.

Os dados foram coletados no formato conforme Figura 11.

Figura 11 - Dados Originais

Relatório de Vendas por Data de Emissão										GERADO EM: 3/27/2020 4:46 PM
Tipo Relatório: Detalhado Venda a Venda (Com Mercadorias/Serviços) Data Emissão: 01/01/2019 a 31/01/2019 Unidade de Negócio: 1(31.424.135/0001-30)										
Data de Emissão										
Hora	Número	Série	Tipo	Integ. XML	Cliente	Caixa	Vista/Prazo	Desconto	Valor Total	
07:04:00	363904	1	NFC-e	Não	CONSUMIDOR FINAL	1623	Vista	0,00	13,09	
Descrição										
Quantidade Preço C. Preço Y. Orig. Total Item										
100396 - ENERG MONSTER 473ML 1,000 6,8330 10,99										
99952 - BALA FRUITTELLA GELATO FRUTAS VERMELHAS 1,000 0,9530 2,10 2,10										
07:06:01	363905	1	NFC-e	Não	CONSUMIDOR FINAL	1623	Vista	0,00	140,01	
Descrição										
Quantidade Preço C. Preço Y. Orig. Total Item										
6 - VP - GASOLINA ADITIVADA V-POWER 31,400 4,0673 4,459 140,01										
07:08:09	363906	1	NFC-e	Não	CONSUMIDOR FINAL	1623	Vista	0,00	70,99	
Descrição										
Quantidade Preço C. Preço Y. Orig. Total Item										
100225 - GELO 3KG 3,000 2,0000 6,500 19,50										
2 - GC - GASOLINA COMUM 4,696 3,9623 4,259 20,00										
100497 - CIGARRO ROTHMANS MINISTER BLUE SPECIAL 3,000 5,0476 5,500 16,50										
101013 - CHOC BIS LACTA AO LEITE 126G 1,000 3,0300 6,990 6,99										
101257 - CHOC LACTA SHOT 30G 1,000 3,0000 8,000 8,00										
07:08:46	363907	1	NFC-e	Não	CONSUMIDOR FINAL	1623	Vista	0,00	50,00	
Descrição										
Quantidade Preço C. Preço Y. Orig. Total Item										
5 - VP - GASOLINA ADITIVADA V-POWER 11,213 4,0673 4,459 50,00										
07:10:00	363908	1	NFC-e	Não	CONSUMIDOR FINAL	1623	Vista	0,00	100,53	
Descrição										
Quantidade Preço C. Preço Y. Orig. Total Item										
20 - CNU - CACAU NATURAL VEGETAL 20,000 3,2426 3,2426 100,53										

Fonte: Empresa estudada.

Compreendidos os dados originais, e seguindo o método CRISP-DM, no tópico seguinte é apresentada a preparação dos dados.

4.3 PREPARAÇÃO DOS DADOS

Nesta fase são abrangidas todas as atividades para construir o conjunto final de dados a partir da base de dados brutos inicial. Para o atingimento dos objetivos, os *databases* devem ser preparados seguindo as atividades comentadas no tópico 3.3 *Método de Trabalho*.

4.3.1 Limpeza dos dados

Também conhecido como *data cleaning*, têm o objetivo de amenizar dois importantes problemas advindos de processos imprecisos de aquisição de dados: a existência de valores ausentes e a existência de valores ruidosos. (SILVA; PERES; BOSCARIOLI, 2016). Sobre os dados coletados, ao que tange quanto a valores ausentes, a informação referente ao cliente da transação foi excluída, uma vez que o percentual de registro do nome do cliente em transações era muito baixo. Ao considerar valores ruidosos, itens com as mesmas descrições, porém diferentes

unidades de medidas tomaram apenas uma das opções como referencial, sendo necessária a conversão correta do preço de aquisição do produto pelo estabelecimento.

Como o intuito do estudo é realizar a análise perante itens de venda na conveniência, todos os registros referentes à pista de abastecimento foram excluídos.

4.3.2 Integração de dados

Também conhecido como *data integration*, consiste em realizar ações que permitam integrar, adequadamente, dados provenientes de diversas fontes de dados. (SILVA; PERES; BOSCARIOLI, 2016). Uma vez que foram extraídos relatórios mês a mês, observou-se problema na redundância dos dados dada as variações de período, com presença de nomenclaturas diferentes para códigos equivalentes, ou então códigos duplicados para itens de mesma descrição.

4.3.3 Transformação dos dados

A etapa de transformação de dados consiste na remoção de dados irrelevantes, ou na transformação do conjunto atual de recursos para um novo conjunto de dados apropriados para mineração e análise (AGARWAL; DHAR, 2014), para que o processo resultante possa ser mais eficiente e os padrões encontrados sejam mais facilmente compreendidos. (HAN; KAMBER; PEI, 2012). Nos dados coletados, verificou-se a necessidade da remoção de dados que poderiam gerar ruídos e não seriam utilizados na análise, sendo de forma inicial, ou em algum avanço do estudo.

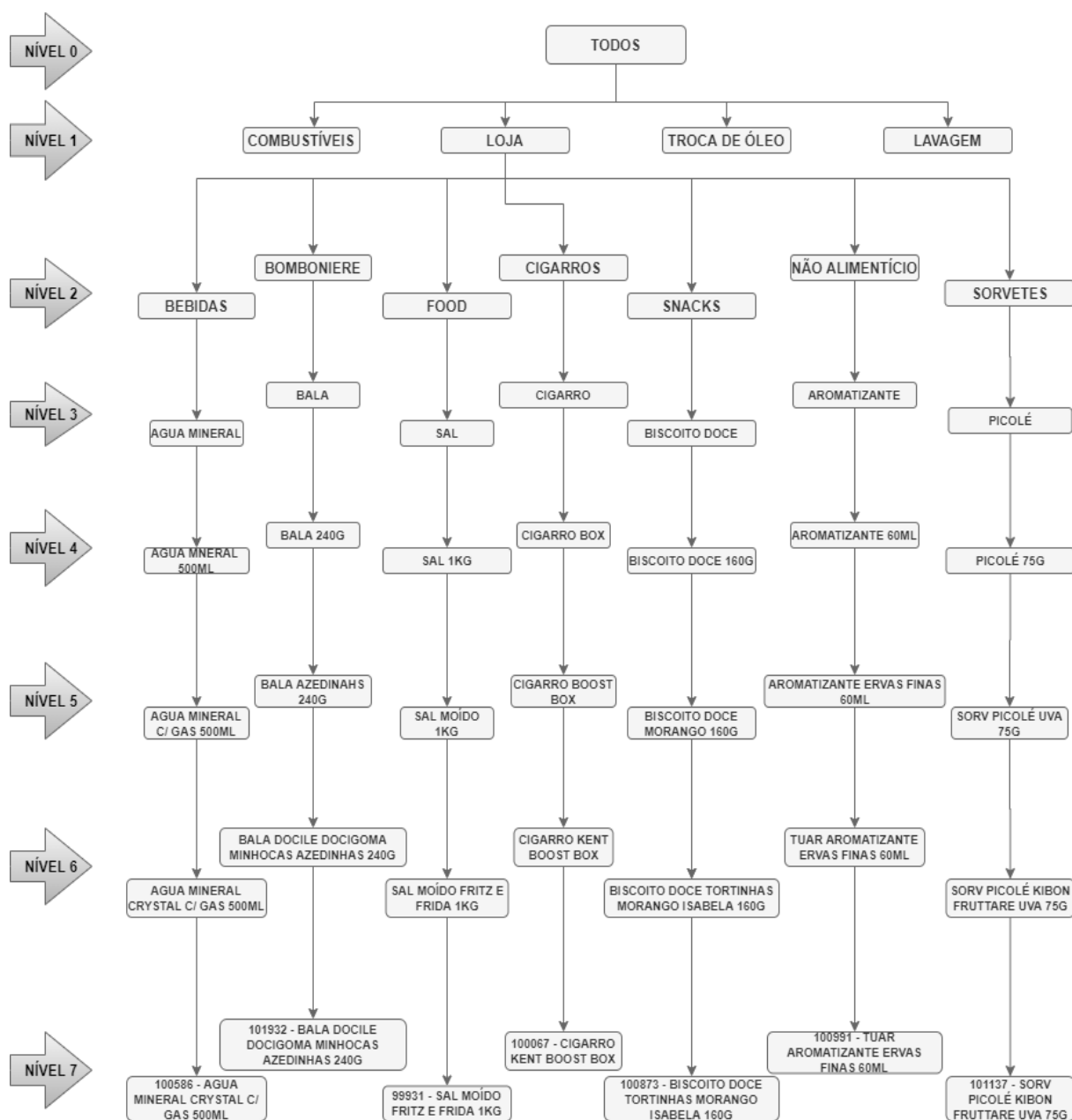
Para a obtenção de uma base consistente, foram agrupados os dados dos 12 meses em um único arquivo do tipo “xlsx”, e as seguintes informações necessitaram de tratamento, a saber:

- a) dados do estabelecimento foram excluídos;
- b) informações de cupons fiscais foram excluídas;
- c) informações de quantidade, preço de compra, preço de venda e desconto relativo aos produtos das transações foram eliminados;
- d) a data, em formato numérico, foi alterada para a nomenclatura do referido dia da semana.

Para que houvesse uma análise com maior coerência, foi necessária a classificação ideal das taxonomias dos produtos, possibilitando a obtenção de resultados, dada a grande diversidade de SKUs (*Stock Keeping Unit*) existentes. Para que os dados se tornem íntegros, é necessária a análise da categoria a ser utilizada. Esta é uma análise do nível de especificação, onde o objetivo é encontrar até que ponto as relações entre as variáveis independentes e dependentes são válidas. (MARCONI; LAKATOS, 2010). Conforme apresentado no tópico 2.3 *Taxonomias de Produto*, foi abrangido o método *Bottom-up*. A escolha do método se deu ao fato dos dados conterem informações no nível mais inferior, ou seja, ao nível de produto. Os níveis, conforme explicitados na Figura 12, ficaram assim divididos:

- Nível 7 (nível de produto): cadastro original do produto no padrão como tal é vendido no ponto de venda;
- Nível 6: nível ao qual é retirado o código identificador da loja. A medida tomou-se necessária dado a existência de itens que, constituem o mesmo produto, porém possuíam cadastrados duplicados;
- Nível 5: nível ao qual, a partir do nível 6, foram retiradas as marcas dos produtos;
- Nível 4: nível onde os produtos foram reorganizados tomando como padrão sua embalagem, ou seja, informações como sabores ou característica principal, foram retiradas;
- Nível 3: neste nível foram ignoradas as embalagens e características como sabores, sendo assim considerado somente a espécie do produto;
- Nível 2: Nível do qual os produtos possuem uma família de caracterização a partir do nível 1;
- Nível 1: nível onde distinguiu-se somente a categoria quanto as divisões de atividades da companhia;
- Nível 0: nível geral dos produtos do estabelecimento, ou seja, nível ao qual não é mais possível atribuir nenhum item pai.

Figura 12 - Níveis de Categorias de Produto



Fonte: Elaborado pelo autor.

Dadas as avaliações iniciais, onde a partir do modelo selecionado foram realizados testes de validações de valores, a fim de validar o nível de categoria a ser selecionada por meio da quantidade de regras geradas a partir da medida estipulada, aquele que melhor se adequou foi a do nível 3. Sendo assim, este foi o nível escolhido para as posteriores etapas. Os dados conforme classificação do estabelecimento são divididos previamente em 7 categorias que se referem a itens de vendas da loja de conveniência, correspondendo a 742 SKUs (*Stock Keeping Unit*) totais, que são designados como nível 7. No nível 3 de categorias, são abordados 145 valores

diferentes. A partir do nível de produto, em um Excel, este nível 3 foi relacionado ao nível de produto para posterior utilização do modelo. O arquivo final, transformado, é demonstrado na Figura 13.

Figura 13 - Dados Tratados

	A	B	C	D	E
1	Data de Emissão	DiaSemana	Hora	Descricao	Numero
2	01/01/2019	TERCA-FEIRA	07:04:00	ENERGETICO	363904
3	01/01/2019	TERCA-FEIRA	07:04:00	BALA	363904
4	01/01/2019	TERCA-FEIRA	07:08:09	GELO	363906
5	01/01/2019	TERCA-FEIRA	07:08:09	CIGARRO	363906
6	01/01/2019	TERCA-FEIRA	07:08:09	CHOCOLATE	363906
7	01/01/2019	TERCA-FEIRA	07:08:09	CHOCOLATE	363906
8	01/01/2019	TERCA-FEIRA	07:16:34	CIGARRO	363910
9	01/01/2019	TERCA-FEIRA	07:17:08	GELO	363911
10	01/01/2019	TERCA-FEIRA	07:17:39	CAFE	363912
11	01/01/2019	TERCA-FEIRA	07:21:29	GELO	363914
12	01/01/2019	TERCA-FEIRA	07:23:15	CAFE	363916
13	01/01/2019	TERCA-FEIRA	07:24:30	CAFE	363918
14	01/01/2019	TERCA-FEIRA	07:24:30	REFRIGERANTE	363918
15	01/01/2019	TERCA-FEIRA	07:25:40	REFRIGERANTE	363919
16	01/01/2019	TERCA-FEIRA	07:25:45	SALGADO PASTEL	363920
17	01/01/2019	TERCA-FEIRA	07:25:45	CAFE	363920
18	01/01/2019	TERCA-FEIRA	07:25:45	DONUTS	363920
19	01/01/2019	TERCA-FEIRA	07:26:32	CAFE	363921
20	01/01/2019	TERCA-FEIRA	07:31:30	CAFE	363925
21	01/01/2019	TERCA-FEIRA	07:31:30	CAFE	363925
22	01/01/2019	TERCA-FEIRA	07:40:01	SALGADO PAO DE QUEIJO	363931

Fonte: Elaborado pelo autor.

Com a realização da limpeza, integração e transformação dos dados, o arquivo no formato “xlsx” foi convertido para o formato “csv”, estando pronto para ser importado no software a ser utilizado. A modelagem dos dados é a concentração do tópico seguinte.

4.4 MODELAGEM

Neste tópico é desenvolvido e aplicado o modelo algorítmico que busca alcançar os objetivos propostos.

4.4.1 Desenvolvimento do Modelo

Para o desenvolvimento do modelo foi selecionado o software gratuito *RStudio*, que se trata de um software de desenvolvimento integrado para R, que é uma linguagem de programação focada em análises estatísticas e gráficos. Estas análises podem ser realizadas por meio de pacotes disponibilizados pelo R. No estudo em questão, são utilizados dois pacotes, o *arules* e o *arulesViz*. O primeiro fornece a infraestrutura para representar, manipular e analisar dados e padrões de transações (HAHSLER; GRUEN; HORNIK; BUCHTA, 2005), e uma das manipulações é a função do algoritmo *Apriori*, que consta dentro deste pacote. O segundo pacote estende às regras várias técnicas de visualização para regras de associação e conjuntos de itens, incluindo também visualizações interativas para exploração de regras.

No *RStudio*, são necessárias as digitações de comandos para que o software importe os dados e seja capaz de explorar as transações para entregar os resultados de análises. A partir disso, para se obter as regras de associações, deve ser construído o modelo conforme apresentado na Figura 14.

Figura 14 - Construção do Modelo

```
1 install.packages("arules")
2 install.packages("arulesViz")
3 library(arules)
4 library(arulesViz)
5 dados <- read.csv("dados.csv", sep = ";")
6 View(dados)
7 transactions <- as(split(dados[, "Descricao"], dados[, "Numero"]), "transactions")
8 summary(transactions)
9 itemFrequencyPlot(transactions, top=10)
10 rules <- apriori(transactions, parameter = list(supp = 0.01, conf = 0.25))
11 inspect(rules)
12 rules_high_lift <- head(sort(rules, by="lift"))
13 inspect(rules_high_lift)
14 plot(rules_high_lift, method="graph", control=list(type="items", arrowSize=1, alpha=1))
15 write(rules_high_lift, file="associacoes.csv", sep = ";", dec=".", row.names= FALSE)
```

Fonte: Elaborado pelo autor.

Os comandos de número 1 e 2, `install.packages("arules")` e `install.packages("arulesViz")`, realizam a instalação dos pacotes em referência, pois não fazem parte das funções iniciais do *software*. Na sequência, os comandos `library(arules)` e `library(arulesViz)` são utilizados para leitura dos pacotes e posterior utilização sobre os dados importados.

Após instalar e carregar os pacotes, o arquivo de dados a ser utilizado deve ser importado, para isso, utiliza-se o comando 5, `read.csv()`, onde deve ser indicado entre aspas o nome do arquivo juntamente à sua extensão, e o separador, através do trecho de comando `sep =";"`, que identifica que as colunas do arquivo são separadas por ponto e vírgula. No sexto comando, é possível verificar como os dados foram importados no *software*. No comando `transactions <- as(split(dados[, "Descricao"], dados[, "Numero"]), "transactions")`, os dados serão transformados do formato data frame em um conjunto de transações do formato *long*. Ou seja, os itens da coluna "Descricao" que possuírem o mesmo dado na coluna "Numero", serão agrupados em uma única linha do conjunto, e na sequência, todos os itens repetidos em uma mesma transação serão excluídos.

No comando número 8, é possível obter uma visão geral sobre o conjunto de dados, que informa a quantidade total de transações, a quantidade que um item é identificado, e a informação de quais são os itens mais frequentes dentro do conjunto de dados importado. O comando `itemFrequencyPlot(transactions, top=10)` ativa um gráfico com os 10 itens mais frequentes na base de dados. A quantidade de itens a ser mostrada no gráfico pode ser manipulada pelo usuário do algoritmo.

A linha de número 10, refere-se ao comando `apriori()`, que terá o algoritmo aplicado sobre o conjunto *transactions*. Nesta linha de comando, as medidas de suporte e confiança, por meio dos parâmetros `supp` e `conf` são determinadas, para que a aplicação determine quais as associações que podem ser mantidas e quais serão descartadas. Na sequência, o comando `inspect()` é aplicado para que seja possível a visualização de todas as regras geradas dados os parâmetros do comando anterior, assim como a apresentação das medidas de interesse para cada uma das regras. O comando de número 12 é aplicado para que por meio da função `sort()`, as regras geradas no comando anterior sejam ordenadas em formato decrescente de acordo com a regra estipulada, que neste caso é o *lift*. O comando 13 é aplicado para que esta ordenação seja exibida no console.

No comando seguinte, por meio da função *plot()*, é gerado o gráfico de correlação entre os itens gerados nas regras. Por fim, o comando 15 é aplicado para que, com intuito de facilitar a análise de regras geradas, seja gerado um arquivo no formato “*csv*” e então exportado de acordo com as parametrizações inseridas.

4.4.2 Teste de Validação do Modelo Construído

Com o intuito de validar a estrutura do modelo quanto à possíveis erros de sintaxe na utilização do *RStudio*, será utilizado um *data set* teste, disponível no *UCI Repository on Machine Learning*, que é um repositório online com inúmeras bases de dados a serviço da comunidade interessada em *Machine Learning*. O *data set* a ser utilizado será o *Online Retail*. Para considerar o modelo como válido, será necessário avaliar a documentação existente, a fim de verificar as regras de associação dadas como resultado. Nesta situação, se os resultados do processo de validação coincidirem com a documentação, o documento estará validado.

Para concretizar a validação, utilizando o modelo construído no tópico anterior, na quinta linha de comando foi inserido o nome do arquivo em análise, neste caso “*OnlineRetail*”. Na linha de comando número 7, foram inseridos os nomes das colunas do arquivo a serem trabalhadas – *Description* e *InvoiceNo* -, conforme Figura 15.

Figura 15 - Validação do Modelo

```
1 install.packages("arules")
2 install.packages("arulesViz")
3 library(arules)
4 library(arulesViz)
5 dados <- read.csv("dados.csv", sep = ";")
6 View(dados)
7 transactions <- as(split(dados[, "Description"], dados[, "InvoiceNo"]), "transactions")
8 summary(transactions)
9 itemFrequencyPlot(transactions, top=10)
10 rules <- apriori(transactions, parameter = list(supp = 0.01, conf = 0.25))
11 inspect(rules)
12 rules_high_lift <- sort(rules, by="lift")
13 inspect(rules_high_lift)
14 plot(rules_high_lift, method="graph", control=list(type="items", arrowSize=1, alpha=1))
15 write(rules_high_lift, file="associacoes.csv", sep = ";", dec=".", row.names= FALSE)
```

Fonte: Elaborado pelo autor.

Nesta etapa de validação do modelo, nenhum *bug* foi observado e não existiram erros de sintaxe, fazendo com que o código fosse validado e o modelo aprovado para posterior utilização dos dados coletados no estabelecimento. Após a aplicação dos primeiros comandos, é possível identificar, por meio do comando *summary* os cinco itens mais frequente na base de teste analisada, conforme Figura 16. Os itens, correspondem respectivamente a *White Hanging Heart T-Light Holder*, que aparece em 2.302 transações, sendo o produto mais frequente, e na sequência há o *Regency Cakestand 3 Tier*, frequente em 2.169 transações. Após, o *Jumbo Bag Red Retrosport* é contabilizado 2.135 vezes, e *Party Bunting* aparece em um total de 1.706 transações. Completando os cinco itens mais frequentes, há o *Lunch Bag Red Retrosport*, presente em 1.607 transações. O valor de 521.121 representa a quantidade de vezes que todos os outros itens aparecem na base de teste analisada.

Figura 16 - Comando *summary(transactions)*

```
> summary(transactions)
transactions as itemMatrix in sparse format with
25900 rows (elements/itemsets/transactions) and
4224 columns (items) and a density of 0.004854042
```

```
most frequent items:
WHITE HANGING HEART T-LIGHT HOLDER
                                2302
      REGENCY CAKESTAND 3 TIER
                                2169
      JUMBO BAG RED RETROSPOT
                                2135
      PARTY BUNTING
                                1706
      LUNCH BAG RED RETROSPOT
                                1607
      (other)
                                521121
```

Fonte: Elaborado pelo autor.

Com a aplicação do comando com o algoritmo *Apriori*, sustentado pelas medidas de parâmetros de suporte e confiança, aplicou-se novamente o comando *summary()*, a fim de verificar a totalidade de regras geradas com tais medidas. Como o explicitado na Figura 17, foram geradas 780 regras de associação a partir das 25.900 transações contidas na base de dados.

Figura 17 - Comando *summary(rules)*

```
> summary(rules)
set of 780 rules

rule length distribution (lhs + rhs):sizes
  2   3   4
549 219  12

      Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
      2.000  2.000   2.000   2.312  3.000   4.000

summary of quality measures:
  support      confidence      coverage
Min.   :0.01000  Min.   :0.2505  Min.   :0.01143
1st Qu.:0.01081  1st Qu.:0.3527  1st Qu.:0.02046
Median :0.01185  Median :0.4662  Median :0.02834
Mean   :0.01329  Mean   :0.4813  Mean   :0.03047
3rd Qu.:0.01456  3rd Qu.:0.5907  3rd Qu.:0.04004
Max.   :0.03216  Max.   :0.8991  Max.   :0.08243

      lift      count
Min.   : 2.848  Min.   :259.0
1st Qu.: 8.651  1st Qu.:280.0
Median :11.129  Median :307.0
Mean   :13.008  Mean   :344.2
3rd Qu.:15.491  3rd Qu.:377.0
Max.   :60.260  Max.   :833.0

mining info:
      data ntransactions support confidence
transactions          25900    0.01    0.25
```

Fonte: Elaborado pelo autor.

Com a aplicação dos demais comandos, as regras foram ordenadas de forma decrescente do valor do *lift*, e então exportadas para um Excel. O Quadro 7 apresenta

uma amostra das regras geradas a partir da aplicação do algoritmo *Apriori*. A coluna *Rules* apresenta em cada uma das linhas as regras geradas, constando o item antecedente (LHS) e o conseqüente (RHS). Na segunda, terceira e quarta linha, aparecem respectivamente as informações de suporte, confiança e *lift*, sendo que o *lift* é quem determina a ordem desta lista. Na quinta coluna, o *Count* representa a quantidade de vezes que aquela associação se encontra na base analisada.

Quadro 7 - Amostra das Regras de Associação Geradas em Teste

<i>Rules</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>	<i>Count</i>
{REGENCY TEA PLATE PINK} => {REGENCY TEA PLATE GREEN}	0,010888	0,89808	60,2603	282
{REGENCY TEA PLATE GREEN } => {REGENCY TEA PLATE PINK}	0,010888	0,73057	60,2603	282
{REGENCY TEA PLATE PINK} => {REGENCY TEA PLATE ROSES}	0,010502	0,86624	49,0933	272
{REGENCY TEA PLATE ROSES} => {REGENCY TEA PLATE PINK}	0,010502	0,59518	49,0933	272
{POPPY'S PLAYHOUSE LIVINGROOM} => {POPPY'S PLAYHOUSE BEDROOM}	0,010116	0,79393	48,2700	262
{POPPY'S PLAYHOUSE BEDROOM} => {POPPY'S PLAYHOUSE LIVINGROOM}	0,010116	0,61502	48,2700	262
{REGENCY TEA PLATE GREEN} => {REGENCY TEA PLATE ROSES}	0,012432	0,83419	47,2772	322
{REGENCY TEA PLATE ROSES} => {REGENCY TEA PLATE GREEN}	0,012432	0,70459	47,2772	322
{POPPY'S PLAYHOUSE LIVINGROOM} => {POPPY'S PLAYHOUSE KITCHEN}	0,010154	0,79697	46,9125	263

Fonte: Elaborado pelo autor.

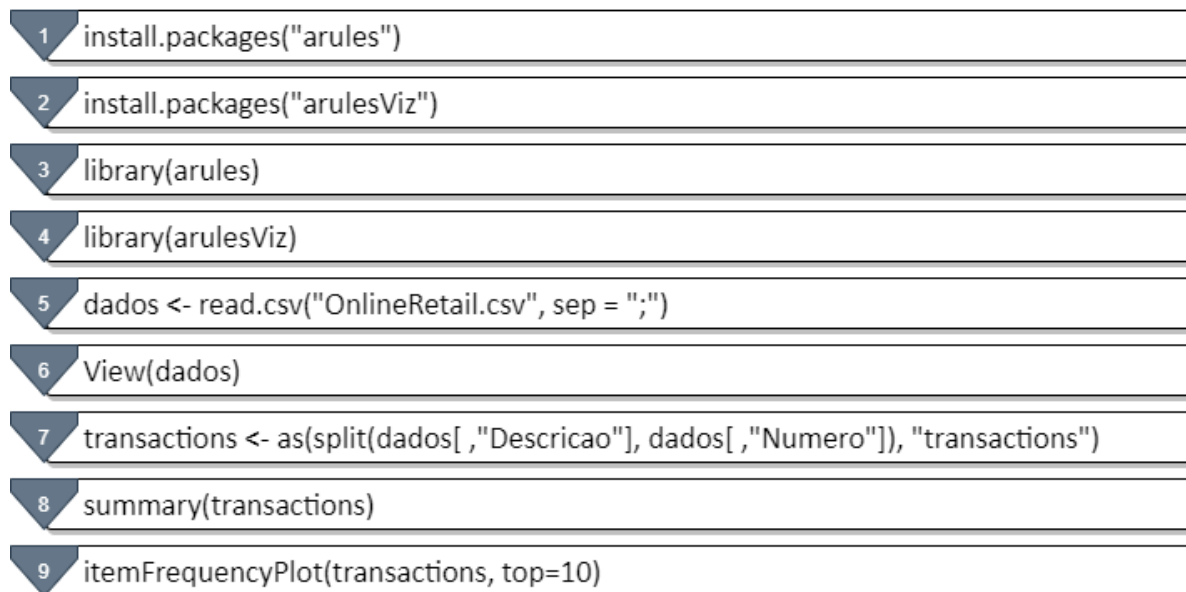
Finalizados todos os comandos sobre a base de teste, observou-se que o modelo não apresentou nenhum erro, e entregou as regras de associação conforme documentos encontrados no repositório, o qual era o resultado esperado. Assim, o modelo encontra-se qualificado para a utilização das transações de venda da empresa foco.

4.4.3 Utilização do Modelo

Para a utilização do modelo, foram carregados no software o arquivo transformado e salvo em formato “csv”, conforme demonstrado anteriormente na Figura 13.

No software, os primeiros comandos, de instalação dos pacotes do software a serem utilizados, e a leitura destes pacotes foram executados, conforme Figura 18, onde na sequência deste procedimento, foi realizado o comando de importação dos dados para o *software* e a visualização da importação.

Figura 18 - Comandos iniciais



```
1 install.packages("arules")
2 install.packages("arulesViz")
3 library(arules)
4 library(arulesViz)
5 dados <- read.csv("OnlineRetail.csv", sep = ";")
6 View(dados)
7 transactions <- as(split(dados[, "Descricao"], dados[, "Numero"]), "transactions")
8 summary(transactions)
9 itemFrequencyPlot(transactions, top=10)
```

Fonte: Elaborado pelo autor.

Com a importação dos dados do estabelecimento, o comando sete foi executado para o agrupamento dos itens de acordo com as transações, e remoção de itens duplicados em uma transação. O comando de *summary(transactions)* foi executado, onde é reforçado que existem 172.192 transações nesta base de dados, havendo 139 códigos de itens presentes nas transações, como indica o Quadro 8.

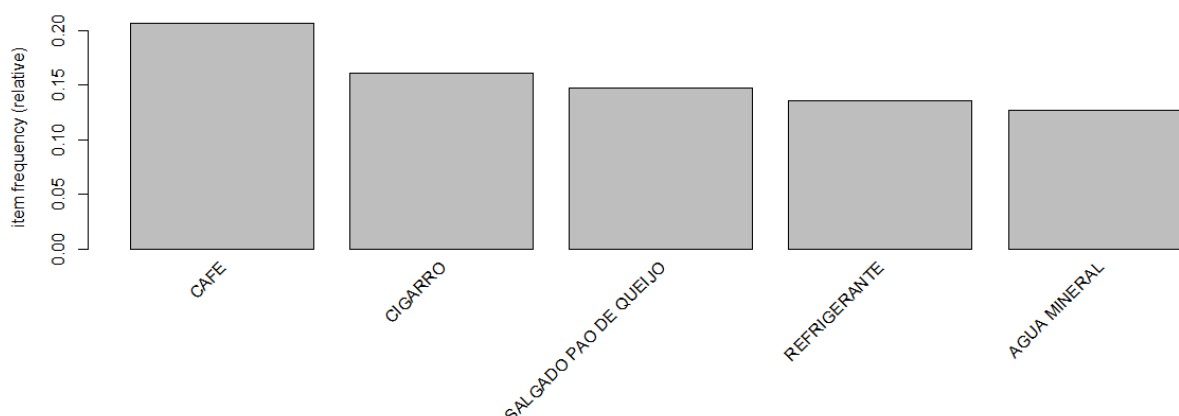
Quadro 8 - Resultados das principais regras

Total de Transações	Quantidade de Itens	Café	Cigarro	Salgado Pão de Queijo	Refrigerante	Água Mineral	Outros
172.192	139	35.526	27.641	25.433	23.286	21.943	160.207

Fonte: Elaborado pelo autor.

O comando abordado previamente, apresenta os itens mais frequentes na base de dados, ou seja, os mais vendidos. O item mais recorrente é o café, presente em 35.526 transações, na sequência, há o cigarro e o salgado pão de queijo, com 27.641 e 25.433 aparições, respectivamente. O quarto item mais frequente é o refrigerante, contabilizado 23.286 vezes, seguido da água mineral, presente em 21.943 transações. No Gráfico 4 é apresentado o resultado do comando de número nove com os cinco itens mais frequentes.

Gráfico 4 - Cinco itens mais frequentes



Fonte: Elaborado pelo autor.

No décimo comando, presente na Figura 19, o algoritmo *Apriori* é aplicado juntamente dos parâmetros de suporte e confiança. Neste estudo, foi definido os valores de 1% para o suporte e 25% para confiança. Ou seja, os resultados desta análise deverão estar presentes juntamente em no mínimo 1% das transações, e em pelo menos 25% das transações que o antecedente (item da esquerda) ocorre, o precedente (item da direita) ocorre. Estes valores foram definidos a partir das análises realizadas nos modelos exemplificados por Silva, Peres e Boscaroli (2016).

Figura 19 - Comando Finais

```

10 rules <- apriori(transactions, parameter = list(supp = 0.01, conf = 0.25))
11 summary(rules)
12 inspect(rules)
13 rules_high_lift <- sort(rules, by="lift")
14 inspect(rules_high_lift)
15 plot(rules_high_lift, method="graph", control=list(type="items", arrowSize=1, alpha=1))
16 write(rules_high_lift, file="associacoes.csv", sep = ";", dec=".", row.names= FALSE)

```

Fonte: Elaborado pelo autor.

As regras geradas a partir do algoritmo *Apriori* devem satisfazer aos dois requisitos: ocorrer de acordo com a frequência (suporte) e a confiança mínima, ou seja, as regras devem ser fortes. (SILVA; PERES; BOSCARIOLI, 2016).

Após aplicada a regra, o comando `summary(rules)` foi ativado para demonstrar a quantidade de regras geradas, conforme Figura 20. Na primeira linha o numeral 2 indica que as regras geradas possuem dois itens associados, e o numeral 8 na segunda linha representa que a partir dos parâmetros, foram geradas 8 regras de associação.

Figura 20 - Resultado do comando `summary(rules)`

```

> summary(rules)
set of 8 rules

rule length distribution (lhs + rhs):sizes
2
8

```

Fonte: Elaborado pelo autor.

O comando `inspect(rules)` é executado para que as regras sejam mostradas no console, e o comando número treze e quatorze são executados para, respectivamente, ordenar as regras pelo *lift* e apresentar os resultados de acordo com esta ordenação. As regras geradas e ordenadas de acordo com o *lift* podem ser verificadas na Figura 21.

Figura 21 - Comando *inspect(rules_high_lift)*

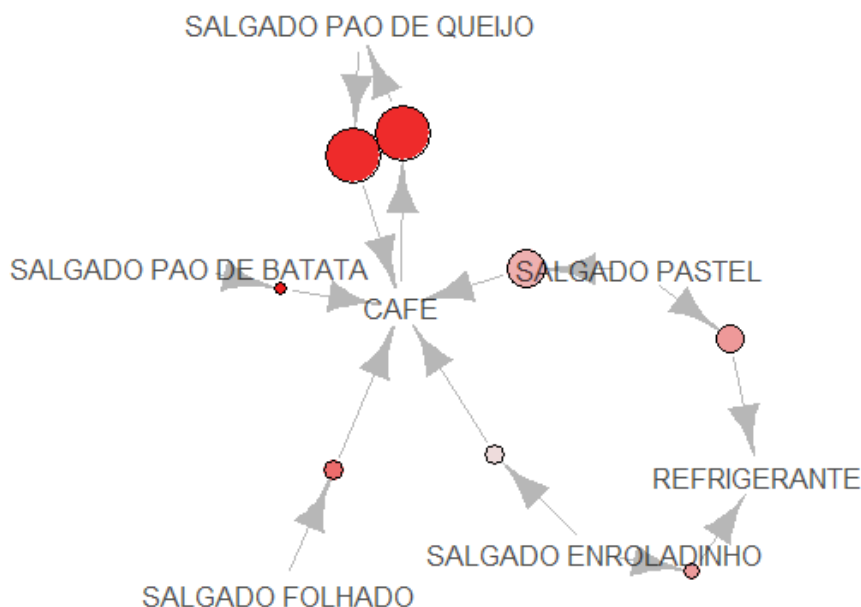
```
> inspect(rules)
  lhs                rhs                support  confidence coverage  lift    count
[1] {SALGADO PAO DE BATATA} => {CAFE}                0.01034311 0.4871444 0.02123211 2.361154 1781
[2] {SALGADO FOLHADO}      => {CAFE}                0.01858391 0.4324909 0.04296948 2.096253 3200
[3] {SALGADO ENROLADINHO} => {REFRIGERANTE}       0.01544787 0.2617337 0.05902133 1.935431 2660
[4] {SALGADO ENROLADINHO} => {CAFE}                0.01879298 0.3184099 0.05902133 1.543310 3236
[5] {SALGADO PASTEL}       => {REFRIGERANTE}       0.03156941 0.2632956 0.11990104 1.946981 5436
[6] {SALGADO PASTEL}       => {CAFE}                0.04468268 0.3726630 0.11990104 1.806271 7694
[7] {SALGADO PAO DE QUEIJO} => {CAFE}                0.07050269 0.4773326 0.14770140 2.313597 12140
[8] {CAFE}                 => {SALGADO PAO DE QUEIJO} 0.07050269 0.3417216 0.20631621 2.313597 12140
```

Fonte: Elaborado pelo autor.

O modelo gráfico destas associações geradas é apresentado no Gráfico 5, onde, quanto maior o tamanho do círculo, maior o suporte da regra, e quanto maior o realce do tom da cor, maior é o *lift* entre a regra dos itens competentes.

Gráfico 5 – Associações geradas

Graph for 8 rules



Fonte: Elaborado pelo autor.

O último comando executado foi o *write(rules_high_lift, file="associacoes.csv", sep = ";", dec=".", row.names= FALSE)*, para que estas regras geradas sejam salvas e armazenadas em um arquivo no formato "csv". A partir destas regras geradas, no próximo tópico as regras serão analisadas, para que então sejam avaliadas perante os dias da semana.

4.5 AVALIAÇÃO

Nesta fase, todos os resultados do modelo são analisados para avaliar até qual o ponto que o modelo desenvolvido atende aos objetivos do negócio.

4.5.1 Análise das Regras de Associação

Conforme apresentado no tópico anterior e ilustrado na Figura 24, utilizadas as medidas mínimas de 1% e 25% para suporte e confiança, respectivamente, foram geradas 8 regras de associação, para que o passo seguinte de relação junto aos dias da semana fosse gerado. As regras geradas inicialmente são explicitadas no Quadro 9, onde o *lift* foi ordenado em ordem decrescente, portanto, as primeiras associações são aquelas das quais possuem maior dependência entre si.

Quadro 9 - Regras de Associação

#	Regras	Lift	Suporte	Confiança	Count
1	{SALGADO PÃO DE BATATA} => {CAFÉ}	2,361	0,010	0,487	1781
2	{SALGADO PÃO DE QUEIJO} => {CAFÉ}	2,314	0,071	0,477	12140
3	{CAFÉ} => {SALGADO PÃO DE QUEIJO}	2,314	0,071	0,342	12140
4	{SALGADO FOLHADO} => {CAFÉ}	2,096	0,019	0,432	3200
5	{SALGADO PASTEL} => {REFRIGERANTE}	1,947	0,032	0,263	5436
6	{SALGADO ENROLADINHO} => {REFRIGERANTE}	1,935	0,015	0,262	2660
7	{SALGADO PASTEL} => {CAFÉ}	1,806	0,045	0,373	7694
8	{SALGADO ENROLADINHO} => {CAFÉ}	1,543	0,019	0,318	3236

Fonte: Elaborado pelo autor.

Os itens que possuem maior dependência entre si são o *Salgado Pão de Batata* e o *Café*, estando presentes juntos em cerca de 1% das transações da loja em estudo, ou seja, a cada 100 transações ocorridas nos estabelecimentos, 1 contém estes itens. A confiança de 45% representa que o café estará presente em 45% das transações que um salgado pão de batata ocorre. A segunda e terceira regra gerada fazem relação entre os mesmos itens, neste caso, pão de queijo e café. A diferença que observada é a seguinte: é mais frequente encontrarmos café nas operações que

possuem o pão de queijo, do que encontrar pão de queijo em operações que possuem café.

Dos itens mais frequentes, observados no subitem *4.4.3 Utilização do Modelo*, é possível analisar que o cigarro, segundo item mais frequente na base de dados, e a água mineral, quinto item mais frequente na base de dados, não possuem nenhuma regra de associação forte. Em contra partida, o item mais frequente na base de dados encontra-se presente em 6 das 8 regras geradas, ou seja, presente em 75% das regras geradas, realçando para a sua alta demanda. Em uma análise comercial inicial, poderíamos dizer que este item tende a ser um agregador de valor às promoções, ou seja, pode ser um item que tende a puxar vendas de outros itens. O pão de queijo, terceiro item com mais saídas, está presente em 25% das regras. Outro item frequente, o refrigerante, é encontrado em 25% das transações.

Um fator interessante encontrado nestas regras fortes é que, sempre, há associação contendo algum dos salgados juntamente a uma bebida. Estes itens, salgados, se fazem importante pelo fato de serem produzidos no próprio local, ou seja, possuem maior valor agregado. E sendo itens de consumo imediato, eles podem ser consumidos no próprio local, o que pode levar o cliente a uma nova transação.

As regras podem ser classificadas como úteis, conforme definição no tópico 2.4.1 Análise de Cestas de Compras, uma vez que a partir destas regras geradas é possível tomar decisões e ações rápidas. Em análise de negócio, há duas análises fundamentais que podem ser realizadas ao observar todo o negócio da empresa. Estas associações podem ser utilizadas para evitar a promoção destes dois itens em um mesmo dia, pois estes dois itens em conjunto já tendem a ter uma venda considerada, ou seja, ao ofertar um item, não deve ser ofertado um item à ele associado pois a probabilidade dos dois produtos serem vendidos simultaneamente em uma transação é grande.

Em outro campo de visão, ao considerar as demais áreas da empresa, por meio dos combos gerados, podem ser utilizados estes combos para puxar um pequeno gatilho promocional em um serviço com maior valor agregado. Como exemplo, ao cliente realizar a compra dos itens pão de batata e café, oferecê-lo 10% de desconto em uma lavagem de carro no mesmo dia. Uma vez que há um funcionário e máquina exclusiva para esta operação de lavagem, portanto é necessária uma demanda contínua para que o custo deste serviço seja mais bem diluído. Portanto, em uma política de ganhos, quanto mais deste serviço for realizado no dia, menor é o custo

operacional, e para isso apropria-se da cesta de compra do cliente, onde a oferta em uma das regras geradas, irá instigar no mesmo uma vantagem econômica em outro serviço. E para o negócio, a rotatividade deste outro serviço agregador é essencial dado o considerado lucro líquido operacional.

A fim de compreender melhor os fenômenos ocorridos nas vendas, no subitem a seguir, estes combos são relacionados aos dias da semana, para que uma análise mais profunda sobre a gestão do negócio possa ser realizada.

4.5.2 Análise das Regras de Associação x Dia da Semana

Conhecidas as regras de associação, buscou-se compreender estes fenômenos em relação ao dia da semana. Esta análise foi realizada sobre o *software* de BI, com intuito de ser possível realizar de forma dinâmica diferentes avaliações. Isso foi possível a partir da mesma base utilizada no *software RStudio*.

No modelo executado, foi realizada a análise de cada uma das regras separadamente, e então registrou-se os percentuais em uma tabela, conforme Tabela 1.

Tabela 1 - Regras de Associação x Dia da Semana

REGRAS	SEGUNDA	TERÇA	QUARTA	QUINTA	SEXTA	SÁBADO	DOMINGO
PÃO DE BATATA >> CAFÉ	13,03%	13,87%	15,22%	16,34%	15,78%	14,54%	11,23%
SALGADO PÃO DE QUEIJO >> CAFÉ	14,33%	19,95%	13,94%	14,59%	15,26%	12,97%	8,95%
CAFÉ >> SALGADO PÃO DE QUEIJO	14,33%	19,95%	13,94%	14,59%	15,26%	12,97%	8,95%
SALGADO FOLHADO >> CAFÉ	14,11%	14,55%	14,21%	14,27%	17,54%	13,80%	11,53%
SALGADO PASTEL >> REFRIGERANTE	15,56%	15,10%	15,21%	15,29%	19,06%	11,74%	8,04%
SALGADO ENROLADINHO >> REFRIGERANTE	15,41%	15,15%	13,99%	16,46%	17,57%	12,16%	8,51%
SALGADO PASTEL >> CAFÉ	14,71%	14,32%	14,87%	16,08%	16,36%	13,43%	10,23%
SALGADO ENROLADINHO >> CAFÉ	13,94%	13,16%	13,84%	14,86%	17,00%	15,08%	12,11%

Fonte: Elaborado pelo autor

Sobre estes resultados gerados, foi possível observar que em grande parte, 87% dos casos, a maior saída destes itens associados tende a ser realizada em dias próximos ao fim de semana. E em 62% dos casos, ou, em cinco das oito regras geradas, a maior volumetria destas saídas ocorreu nas sextas-feiras. Estas regras foram: Café ⇒ Salgado Pastel, Salgado Pastel ⇒ Refrigerante, Salgado Enroladinho ⇒ Refrigerante, Salgado Pastel ⇒ Café, Salgado Enroladinho ⇒ Café, com 17,54%, 19,06%, 17,57%, 16,36% e 17,00%, respectivamente.

Na cesta Pão de Batata ⇒ Café, o dia com maior volumetria é a quinta-feira, com 16,34% das ocorrências frente aos dias da semana. A regra Salgado Pão de Queijo ⇒ Café e seu inverso, possuem os mesmos valores, uma vez que estão sendo analisadas as relações dos dois itens em conjunto relacionados ao dia da semana, e estas tiveram volume expressivo de vendas na terça-feira, com percentual de 19,95% relacionado aos dias da semana.

Como exemplificado no subitem 4.5.1 *Análise das Regras de Associação*, se formos buscar uma associação diária como gatilho para a oferta de outro serviço, é necessário comparar qual a associação que possui maior volumetria em um dia da semana. Nesta situação, é mostrado na Tabela 2 que há uma associação diferente para cada dia da semana, e nos fins de semana, sábado e domingo, há maior volumetria da mesma regra, Salgado Enroladinho ⇒ Café. Na busca por uma relação para uma oferta promocional com um serviço, seria possível obter uma oferta diferente para cada dia da semana, exceto aos finais de semana.

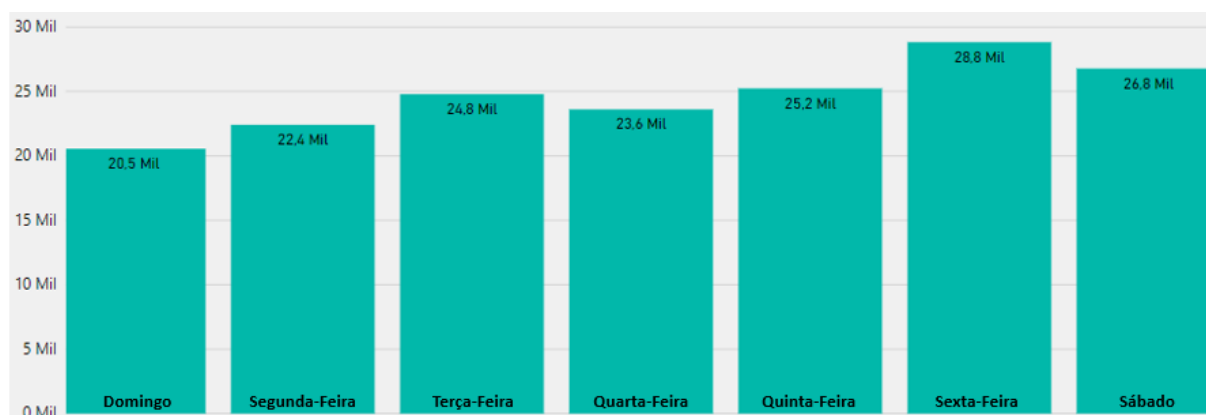
Tabela 2 - Associação mais vendida em cada dia

DIA DA SEMANA	%	ASSOCIAÇÃO
SEGUNDA-FEIRA	15,56%	SALGADO PASTEL ⇒ REFRIGERANTE
TERÇA-FEIRA	19,95%	PÃO DE QUEIJO ⇒ CAFÉ
QUARTA-FEIRA	15,22%	PÃO DE BATATA ⇒ CAFÉ
QUINTA-FEIRA	16,46%	SALGADO ENROLADINHO ⇒ REFRIGERANTE
SEXTA-FEIRA	19,06%	SALGADO PASTEL ⇒ REFRIGERANTE
SÁBADO	15,08%	SALGADO ENROLADINHO ⇒ CAFÉ
DOMINGO	12,11%	SALGADO ENROLADINHO ⇒ CAFÉ

Fonte: Elaborado pelo autor.

Se avaliado o domingo em maior detalhe, pode ser observado que a representatividade do percentual é menor em comparação aos demais dias, isso também pode ser relacionado pelo fato de ser um dia no qual as pessoas costumam manter-se em suas residências, portanto o fluxo comercial não é tão elevado como em dias de semana. Esta comparação é demonstrada no Gráfico 6, que representa a volumetria de transações de cada dia de acordo com a base de dados.

Gráfico 6 - Volume de vendas em decorrência dos dias



Fonte: Elaborado pelo autor.

A partir das análises realizadas, observa-se que, para melhor administração do negócio, uma vez que estas associações estão atreladas à itens que são de produção própria, seria um fator diferencial na análise estipular períodos dentro do dia. Assim, procurou-se estabelecer períodos do dia, para que sejam analisadas as regras associadas por meio do *Data Mining* em comparações diretas a estes períodos. Analisadas as informações geradas frente ao negócio, e considerado que a base de dados contém horários de ocorrência das transações, torna-se plausível a divisão do dia em diferentes períodos, o que tenderia a trazer maiores informações e poder de decisão ao gestor do negócio.

Como as regras de associação geradas contém itens que dependem de produção própria, a produção pode ser direcionada para determinados momentos do dia, a fim de obter melhor aproveitamento da equipe de funcionários. A partir disso, foram subdivididos nos turnos existentes no dia, e em períodos dentro de cada turno, de acordo com Quadro 10.

Quadro 10 - Períodos

Hora Início	Hora Fim	Turno	Período
05:45:00	08:00:00	MANHÃ	MANHÃ1
08:00:00	10:00:00		MANHÃ2
10:00:00	12:00:00		MANHÃ3
12:00:00	14:00:00	TARDE	TARDE1
14:00:00	16:00:00		TARDE2
16:00:00	18:00:00		TARDE3
18:00:00	20:00:00	NOITE	NOITE1
20:00:00	23:00:00		NOITE2

Fonte: Elaborado pelo autor.

Sobre esta divisão, é realizada a análise sobre cada uma das regras de associações, para melhor compreender a divisão das vendas, e quais informações podem ser coletadas nesta análise. A geração de resultados apresenta-se no Quadro 11 para a primeira regra gerada, as demais podem ser visualizadas no Apêndice A.

Quadro 11 - Regra de Associação comparada aos períodos

(Continua)

SALGADO PÃO DE BATATA ⇒ CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA-FEIRA	MANHÃ	MANHÃ1	48	20,70%	61,20%	2,70%
		MANHÃ2	67	28,90%		3,70%
		MANHÃ3	27	11,60%		1,50%
	TARDE	TARDE1	12	5,20%	25,90%	0,70%
		TARDE2	11	4,70%		0,60%
		TARDE3	37	15,90%		2,10%
	NOITE	NOITE1	22	9,50%	12,90%	1,20%
		NOITE2	8	3,40%		0,40%
	TERÇA-FEIRA	MANHÃ	MANHÃ1	59	23,90%	70,00%
MANHÃ2			82	33,20%	4,60%	
MANHÃ3			32	13,00%	1,80%	
TARDE		TARDE1	6	2,40%	19,40%	0,30%
		TARDE2	14	5,70%		0,80%
		TARDE3	28	11,30%		1,60%
NOITE		NOITE1	21	8,50%	10,50%	1,20%
		NOITE2	5	2,00%		0,30%
QUARTA-FEIRA		MANHÃ	MANHÃ1	77	28,70%	64,90%
	MANHÃ2		74	27,60%	4,10%	
	MANHÃ3		23	8,60%	1,30%	
	TARDE	TARDE1	5	1,90%	23,90%	0,30%
		TARDE2	21	7,80%		1,20%
		TARDE3	38	14,20%		2,10%
	NOITE	NOITE1	20	7,50%	11,20%	1,10%
		NOITE2	10	3,70%		0,60%
	QUINTA-FEIRA	MANHÃ	MANHÃ1	80	27,50%	70,80%
MANHÃ2			101	34,70%	5,60%	
MANHÃ3			25	8,60%	1,40%	
TARDE		TARDE1	6	2,10%	19,60%	0,30%
		TARDE2	11	3,80%		0,60%
		TARDE3	40	13,70%		2,20%
NOITE		NOITE1	18	6,20%	9,60%	1,00%
		NOITE2	10	3,40%		0,60%

Fonte: Elaborado pelo autor

Quadro 11 - Regra de Associação comparada aos períodos

(Conclusão)

SALGADO PÃO DE BATATA ⇒ CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEXTA-FEIRA	MANHÃ	MANHÃ1	69	24,60%	67,60%	3,90%
		MANHÃ2	86	30,60%		4,80%
		MANHÃ3	35	12,50%		2,00%
	TARDE	TARDE1	14	5,00%	21,40%	0,80%
		TARDE2	15	5,30%		0,80%
		TARDE3	31	11,00%		1,70%
	NOITE	NOITE1	24	8,50%	11,00%	1,30%
		NOITE2	7	2,50%		0,40%
	SÁBADO	MANHÃ	MANHÃ1	48	17,70%	73,80%
MANHÃ2			99	36,50%	5,50%	
MANHÃ3			53	19,60%	3,00%	
TARDE		TARDE1	14	5,20%	20,70%	0,80%
		TARDE2	21	7,70%		1,20%
		TARDE3	21	7,70%		1,20%
NOITE		NOITE1	9	3,30%	5,50%	0,50%
		NOITE2	6	2,20%		0,30%
DOMINGO		MANHÃ	MANHÃ1	25	12,50%	67,00%
	MANHÃ2		67	33,50%	3,70%	
	MANHÃ3		42	21,00%	2,30%	
	TARDE	TARDE1	10	5,00%	16,00%	0,60%
		TARDE2	14	7,00%		0,80%
		TARDE3	8	4,00%		0,40%
	NOITE	NOITE1	27	13,50%	17,00%	1,50%
		NOITE2	7	3,50%		0,40%

Fonte: Elaborado pelo autor.

Na primeira regra gerada com base no maior *lift*, Salgado Pão de Batata ⇒ Café, é observado que com grande relevância, esta associação ocorre no turno da manhã, sendo o período Manhã2 o de maior compra dos clientes, que compreende o horário das 08:00 horas até as 10:00 horas. Portanto, este é um horário do qual o gestor do negócio e a equipe devem estar atentos à produção e reposição destes itens, para que não ocorra a falta da oferta de produtos aos clientes do local. O período Manhã1 é o segundo período com maior venda desta cesta.

Ao avaliar o resultado do turno apenas, observa-se que a manhã de quinta-feira é quando ocorre a maior contabilização de registros de transações desta associação, informação que vai de encontro com a informação apresentada na Tabela 1, ao avaliar a saída, desta regra específica, somente em relação ao dia da semana.

Ao avaliar a segunda regra gerada, Salgado Pão de Queijo \Rightarrow Café, e a terceira regra gerada, que é o inverso desta regra, é possível averiguar que esta associação tem enorme recorrência no período Manhã² das terças-feiras, onde novamente o resultado é consistente com o dado apresentado na Tabela 1.

A regra de associação Salgado Folhado \Rightarrow Café contém maior contabilização de vendas aos sábados, também no período Manhã². Esta informação é diferente ao avaliado anteriormente considerando somente o dia da semana e não os períodos, onde foi possível observar que a maior relevância foi obtida às sextas-feiras. Na situação em análise, a volumetria de vendas neste dia e horário é muito maior comparada aos demais horários do mesmo dia, o que conseqüentemente pode ser associado que este seja o horário de maior movimento do estabelecimento.

Na regra de associação Salgado Pastel \Rightarrow Refrigerante, é possível observar uma variação no resultado em comparação às associações anteriores. A maior relevância de vendas deste combo encontra-se nas sextas-feiras a tarde, das 16 às 18 horas, o que compreende o período Tarde³. Esta análise pode ser compreendida por serem os últimos horários comerciais da semana na região, onde as pessoas costumam sair de sua rotina, fato que pode ser consumado pela ausência do café pela primeira vez nas associações resultantes. O café, pode ser considerado um produto energético, o que justifica seu maior consumo pela parte da manhã. No caso da associação Salgado Pastel \Rightarrow Refrigerante, esta regra é frequente no turno da tarde, tendo sua venda, em análise geral, sendo o dobro da realizada no turno da manhã, conforme pode ser visualizado na Tabela 3, que representa as médias dos produtos em relação aos turnos. Os combos estão na ordem de acordo com o *lift* da regra gerada, portanto em ordem decrescente desta medida.

Tabela 3 - Combos x Turnos

#	Combos	Manhã	Tarde	Noite
1	Pão de Batata >> Café	67,9%	21,0%	11,1%
2	Pão de Queijo >> Café	69,4%	19,5%	11,1%
3	Salgado Folhado >> Café	73,8%	16,8%	9,4%
4	Salgado Pastel >> Refrigerante	29,0%	50,6%	20,4%
5	Salgado Enroladinho >> Refrigerante	31,2%	45,3%	23,5%
6	Salgado Pastel >> Café	66,2%	21,5%	12,2%
7	Salgado Enroladinho >> Café	62,1%	22,3%	15,6%

Fonte: Elaborado pelo autor.

Ao avaliar o combo Salgado Enroladinho ⇒ Refrigerante é possível observar que o combo possui uma homogeneidade quanto aos períodos de venda em comparação às demais associações. Sua maior contabilização de vendas é encontrada nas quintas-feiras, no período Tarde³, o que diverge quando comparada a mesma associação relacionada ao dia de forma geral. Quando comparado a regra gerada em relação ao dia da semana, sem a determinação do período, o mesmo ocorreu com maior frequência nas sextas-feiras. Indiferente o dia da semana, este item sempre possui maior saída no turno da tarde.

A associação Salgado Pastel ⇒ Café, assim como o combo Salgado Enroladinho ⇒ Café têm, ambas, relevâncias observadas no período Manhã² da sexta-feira, convergindo com a análise conforme Tabela 1, quando avaliado somente os dias de forma generalizada.

No capítulo seguinte, apresenta-se as discussões dos resultados obtidos.

5 DISCUSSÃO DOS RESULTADOS

Realizada as análises dos resultados, é considerado relevante realizar uma discussão por meio dos resultados obtidos. Assim, este capítulo discute as contribuições da pesquisa para a organização e para a academia.

Na busca por encontrar a venda de itens associados, por meio do *Data Mining*, e obter a relação destes com os dias da semana, foi possível observar que as associações geradas de acordo com os parâmetros estipulados, possuem volatilidade. Ou seja, dentre as associações, não há uma que seja “campeã” de vendas em todos os dias da semana, cada dia possui uma associação diferente em destaque. Como consequência, é possível criar campanhas promocionais destes itens em dias diferentes daqueles aos quais já possuem saída conhecida, a fim de estimular maior giro dos itens no estabelecimento.

Um fato que o método aplicado permite identificar é que dentre os cinco itens mais frequentes na base coletada, conforme mostrado no Gráfico 4, dois deles, cigarro e água mineral, não se encontram nas regras geradas. O que pode indicar que estes itens possuem tendência a vender sozinhos ou com os mais variados produtos.

Ao avaliar a relação das regras geradas com os dias da semana, ao qual foi dado como objetivo inicial deste trabalho, identificou-se fatores que poderiam contribuir e impactar consideravelmente nesta variável e nas decisões do gestor. Fatores importantes e observados foram, o período do dia e a análise por turno.

Na análise quanto ao período do dia, observa-se que a maioria dos combos gerados possuem maior saída no período Manhã², horário entre às 8:00 e 10:00 horas da manhã, compreendendo as primeiras horas comerciais do dia. Porém, o que é possível analisar nas regras deste período é que o item consequente é sempre o café, o que torna este produto indispensável no estabelecimento, essencialmente neste horário. Portanto, ao analisar o negócio em si, para que não ocorra uma perda de venda, é essencial que este item possua estoque e esteja disponível para servir ao cliente de forma saborosa. Este item, encontra-se presente em 35.526 transações, o que equivale a 20,6% das vendas do estabelecimento.

Em contra partida, as associações que não possuem a relevância em vendas no período Manhã², estão segregadas no período Tarde³, compreendido das 16:00 às 18:00 horas, e como item consequente, apresentam o refrigerante.

Quando a análise esteve limitada aos turnos do dia, pode manter-se os indicativos comentados anteriormente, onde regras com o conseqüente *Café* tiveram maior relevância de vendas pela manhã, e quando apresentado o conseqüente *Refrigerante*, este teve maior volumetria no período da tarde. Nestes casos de vendas em que o turno da manhã foi relevante, este momento do dia chega a ser responsável, em média, por 67% das vendas dos combos gerados, em alguns casos tendo suas saídas três vezes maiores que o do segundo período com maior relevância.

O período da noite, em que apresenta uma venda relativamente baixa, pode colocar em análise a viabilidade do funcionamento do estabelecimento em determinados horários, onde a partir desta análise, mais uma variável sobre o tempo poderia ser avaliada, que são as vendas hora a hora. Assim, o gestor pode avaliar qual o ponto que justifica manter o estabelecimento aberto, dado todos os custos envolvidos. Esta é uma análise a nível estratégico, uma vez que entram as despesas com pessoas, despesas fixas de estruturas e chances de roubos.

Em uma análise frente a concorrência, há um efeito contrário e que pode reforçar o motivo de manter o local em funcionamento neste turno, pois ao fechá-lo, há uma perda na receita do negócio e logo, abertura de espaço aos concorrentes neste horário. Neste caso, a utilização das regras de associação podem ser um estímulo para trazer clientes nestes períodos de baixa demanda, por meio da criação de promoções. Este é o momento em que, dadas estas necessidades de articulações, a inteligência empresarial e poder de decisão se tornam diferenciais para o destaque do negócio em relação aos concorrentes.

Uma informação resultante que se torna considerável, é que, ao ser considerado somente o período Manhã3, o sábado e o domingo são os dias com maior saída dos itens presentes nas associações. O fator que pode interferir diretamente é que o estabelecimento se localiza na rota de um dos maiores santuários da região em estudo, e este é o horário que abrange o encerramento do evento religioso com maiores espectadores nestes dias.

No capítulo seguinte são abordadas as considerações finais sobre o estudo, bem como suas limitações e sugestões para trabalhos futuros.

6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo principal selecionar um algoritmo de DM que encontrasse as associações nas compras dos clientes, para posterior análise junto aos dias da semana, permitindo avaliar o comportamento das vendas do estabelecimento. Para realizar estas associações, conforme referencial teórico apresentado no capítulo 2, buscou-se na literatura uma tarefa do *Data Mining*, e posteriormente um algoritmo que fosse capaz de realizar esta atividade a partir da base de dados coletada. Com isso, a escolha pelo algoritmo *Apriori* se deu por ser um algoritmo de ampla abrangência na literatura.

O desenvolvimento e validação deste modelo, apresentados no tópico 4.4 *Modelagem* suporta a utilização do algoritmo. Assim, foi possível no capítulo 4 gerar as regras de associação e compreendê-las frente aos dias da semana. Ao decorrer do desenvolvimento deste trabalho, por meio das etapas da metodologia CRISP-DM, foi identificada a capacidade de fornecer informações mais precisas ao gestor ao dividir os dias de semana em turnos e períodos. Com isso, há a capacidade de possibilitar ao gestor maior vantagem competitiva do estabelecimento frente aos concorrentes a partir dos dados coletados diariamente.

Entende-se que os resultados gerados, atendem ao objetivo principal da pesquisa, pois foi capaz de gerar informações mais detalhadas quanto aos padrões de compras dos clientes no ponto de venda e possibilitou trazer em relevância que diferentes fatores do dia a dia podem impactar na geração de lucro da empresa analisada. No capítulo 5, os resultados do estudo realizados foram discutidos, dando ênfase a estes fatores diversos.

Ao que tange ao objetivo específico “**desenvolver uma abordagem de ETL (*Extraction, Transformation and Loading*) sobre os dados não estruturados**” buscou-se apresentá-lo no capítulo 4 por meio da compreensão e preparação dos dados. A extração e posterior correto tratamento são essenciais para a correção de inconsistências nos dados, conforme enfatizam Han, Kamber e Pei (2012). Nesta fase ocorreu o melhor detalhamento das taxonomias do produto, a fim de buscar mais coerência dos resultados ao que tange à frequência das associações. A seleção apropriada do nível de classificação dos itens a ser selecionada para utilização na base interfere diretamente na quantidade de resultados a partir das medidas estipuladas.

Para atender o objetivo **“identificar um algoritmo que possa ser empregado na base de dados da loja de conveniência”**, no capítulo 2 foram debatidos os algoritmos de KDD. Foi possível observar que embora existam inúmeros algoritmos capazes de atender a proposta das regras de associações, o algoritmo *Apriori* é fortemente relevante na literatura. A validação deste modelo, por mais que simples, é essencial, uma vez que qualquer divergência no algoritmo resulta em informações imprecisas e divergentes da realidade do negócio. E tratando-se de um negócio com grande fluxo, decisões não assertivas podem levar a perda de considerável valor econômico.

O objetivo específico **“identificar as associações de compras a partir do algoritmo aplicado”** foi abordado no capítulo 4. No processo em questão, em primeiro momento, desenvolveu-se e validou-se o modelo a ser utilizado, para então executá-lo a fim de compreender os itens mais frequentes na base de dados, ou seja, aqueles com maiores saídas. A partir disso as regras de associações foram geradas conforme medidas de interesse estipuladas, a partir das 172.192 transações identificadas na base e dados. Ao considerar as informações extraídas, é possível gerar uma melhor visão acerca dos padrões de compras dos clientes.

O atingimento do objetivo **“analisar as associações frente aos dias da semana para gerar *insights* do negócio objetivando auxiliar as tomadas de decisões”** foi atingido no capítulo 4, onde a partir das regras geradas, comparou-se estas junto à base de dados, por meio de análises de *business intelligence*. Como resultado foi possível avaliar através das datas, os dias da semana, turnos e períodos de maior ocorrência das associações. Sobre estas análises foi possível observar que o processo de decisão de negócios pode ser direcionado à três diferentes métricas para análise de negócio. A primeira avaliação foi realizada tendo como base as associações geradas, buscando qual a representatividade que cada dia possui em comparação ao período semanal. Nesta avaliação é possível verificar que as cestas geradas possuem uma certa homogeneidade nos percentuais durante os dias de semana, e possuem uma queda aos finais de semana.

A segunda avaliação que pôde ser efetuada foi, a partir do dia da semana, associar a cesta de maior sobressalência dentre as demais. Este há de ser um desdobramento para responder à questão “dado o dia da semana, qual é a regra de associação que melhor se atribui?”. A partir desta verificação, análises trimestrais

poderiam ser executadas para campanhas de marketing mais bem direcionadas nos meses subsequentes.

O terceiro método de avaliação, surgido no decorrer desta monografia, foi subdividir os períodos dos dias, ou seja, estipular as associações que mais ocorrem naquele período. Assim, pode-se concluir que cada cesta gerada possui um comportamento diferenciado ao longo do dia, gerando maiores detalhes sobre os comportamentos dos clientes do estabelecimento. Os padrões de consumo sobre as associações não são lineares, o que enfatiza a importância deste segmento do varejo, fornecendo praticidade e variedade aos consumidores.

Constata-se que a ferramenta de *Data Mining* proporciona um suporte à gestores de negócios no fornecimento de informações, uma vez que dados por si só, existem, porém, informação, só existe a partir de dados. A partir do modelo gerado e seus resultados, estas informações podem auxiliar nas avaliações de giro de estoques, pois conhecendo os padrões de vendas, o gestor é capaz de estipular estoques mínimos e melhor performar suas compras junto aos fornecedores. Com estas informações, o gestor é capaz também de mensurar o giro dos produtos, uma vez que produtos parados em estoque significam perda de dinheiro, e quanto maior a taxa de giro, mais lucro gerado à empresa. Inclusive, estas análises apoiam as previsões de demanda para análises temporais no negócio, tornando os administradores aptos a direcionarem suas tomadas de decisões junto à equipe.

6.1 LIMITAÇÕES DO TRABALHO

No que tange às limitações deste estudo, o período ao qual os dados foram analisados é relevante. Anteriormente a este período, a empresa possuía outro sistema de frente de loja, portanto não havia os registros transacionais anteriores a este período. E o período posterior à data analisada, pode ser considerado como “anormal”, dado o surgimento de uma pandemia, que obrigou o estabelecimento a paralisar seu funcionamento habitual por determinados períodos.

Ainda como limitação quanto aos dados, a ausência de mais classificações dos produtos leva a uma classificação manual do executor da tarefa. Assim, conflitando com o tempo disponível para a realização das atividades, não é possível um brainstorming junto aos gestores do negócio para avaliação de cada item a fim de

gerar novos níveis aos produtos, e introduzir estes dentro do sistema de gestão da empresa.

Por motivo da classificação manual das novas taxonomias dos produtos, não é possível atribuir um preço médio ao nível dos produtos utilizados na mineração.

Han, Kamber e Pei (2012) reforçam que quanto maiores forem as medidas de interesse – suporte e confiança – maior é a utilidade e certeza das associações geradas. Porém, como o mercado varejista é constituído de centenas de SKUs, é necessário a utilização de valores não muito altos, para que assim sejam encontrados resultados plausíveis de análise.

A falta de maiores dados dos consumidores é um fator limitante neste estudo. Conforme apresentação das referências utilizadas no decorrer da monografia, perfis de clientes poderiam ser estipulados, assim, as regras poderiam ser cruzadas frente a estes perfis, assim como realizar o cruzamento entre regras geradas, perfis de clientes e períodos do dia. Este seria um insight pouco encontrado na academia até o momento, uma vez que seria bem focalizado nos padrões específicos dos consumidores, dando aptidão para um prévio conhecimento de reação de mercado a partir de uma ação do negócio.

Além destes, pode-se citar a falta de tempo para colocar o estudo realizado em prática, o que inviabiliza a sexta etapa do método CRISP-DM, que consiste no desenvolver do negócio a partir do modelo gerado, avaliando suas informações geradas.

6.2 SUGESTÕES PARA TRABALHOS FUTUROS

Em um primeiro momento, sugere-se continuar este trabalho, executando a etapa de número seis do método CRISP-DM. Assim, podem surgir campanhas de marketing diretamente para cada dia da semana, a fim de mensurar qual a aderência e conversão destas campanhas. Igualmente, relacionado ao marketing, conforme mencionado durante este trabalho, as cestas geradas podem ser divulgadas em períodos diferentes daqueles das quais mais possuem registros de saída, e assim as promover juntamente à algum outro serviço que o negócio possui, oferecendo um desconto neste serviço prestado.

Igualmente, sugere-se um novo estudo a partir do modelo utilizado para avaliação em torno do layout do estabelecimento para posicionamento estratégico dos

itens, o que pode vir a aumentar o ticket médio de compra dos clientes. A expansão deste estudo aos demais estabelecimentos da companhia é uma sugestão, visto que podem ser realizadas comparações tendo em vista a possibilidade de encontrar diferentes padrões de compras dos clientes de acordo com as regiões em que se localizam.

Indica-se também uma replicação deste estudo, identificando o perfil do consumidor, como faixa etária, sexo, classe social, modo de pagamento, onde, então será possível direcionar ofertas de acordo com os perfis de clientes. Também, outra variável com viés econômico e de importância aos gestores pode ser comparada, que é a geração de associações junto a faixa de lucro que estas cestas deixam ao negócio, pois nem sempre itens com maior demanda possuem retorno financeiro atrativo.

O modelo apresentado neste trabalho possui grande aderência ao mercado varejista, com isso, sugere-se a aplicação deste modelo em varejistas de moda, redes de farmácia, lojas de materiais de construção, varejistas do mercado pet e redes de supermercados, a fim de disseminar a tarefa de *Market Basket Analysis* provendo informações úteis aos gestores.

REFERÊNCIAS

AGARWAL, Ritu; DHAR, Vasant. Big data, Data science, and Analytics: The Opportunity and challenge for is research. **Information System Research**, Maryland, v. 25, n. 3, p. 443-448, Set. 2014.

AGRAWAL, Rakesh; IMIELIŃSKI, Tomasz; SWAMI, Arun. Mining Association Rules Between Sets of Items in Large Databases. *In: ACM SIGMOD Record*, 1993, p. 207-216.

AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Fast Algorithms for Mining Association Rules. *In: 20th INTERNACIONAL CONFERENCE ON VERY LARGE DATABASE*. Santiago Del Chile: Morgan Kaufmann, 1994, p. 487-499.

AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Mining sequential patterns. *In: Eleventh International Conference on Data Engineering*. Taiwan: IEEE, 1995, p. 3-14.

AHLERT, Fabiano Charlier; VACCARO, Guilherme Luís Roehle; COSTA, Filipe Campelo Xavier. **A Compreensão da Motivação de Compras de Clientes utilizando a Análise Fatorial**. n. January, 2009.

ANSELMO, Filomena Clara Gouveia. Regras de Associação - **Market Basket Analysis** - Itens Frequentes e Itens Raros. 2017. 72 f. Tese (Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão) - Mestrado em Economia e Gestão, Universidade do Porto (FEP), Portugal, 2017.

APÓS sete meses de alta, venda do varejo brasileiro caem. Mercado e Consumo, [S.l], 13 fevereiro 2020. Disponível em: <<https://www.mercadoeconsumo.com.br/2020/02/13/apos-sete-meses-de-alta-vendas-do-varejo-brasileiro-caem/>>. Acesso em: 02 mai. 2020.

ARIAS, Luis Fernando Caruncho. **Lojas de Conveniência: O impacto do risco percebido na intenção de compra vs. o nível de automação do serviço**. 2019. 47 f. Dissertação (Mestrado em Marketing) – Lisbon School of Economics & Management, Universidade de Lisboa, Lisboa, 2019.

ARMAZENAMENTO de dados: entenda a importância para a sua empresa!. [S.l.], 24 out. 2017. Disponível em: <https://brasil.softlinegroup.com/sobre-a-empresa/blog/armazenamento-de-dados-entenda-a-importancia-para-a-sua-empresa>. Acesso em: 30 abr. 2020. Blog: Softline, we know we can.

AS LOJAS do futuro passam obrigatoriamente pelas lojas de conveniência. **Brasil Postos**, [S.I.], 19 ago. 2020. Disponível em: <<https://www.brasilpostos.com.br/noticias/loja-de-conveniencia/239616/>>. Acesso em: 12 out. 2020.

ASSAD, Nancy. Inteligência Empresarial Inovadora. **Catho – Carreira & Sucesso**, 2011. Disponível em: <<https://www.catho.com.br/carreira-sucesso/colunistas/nancy-assad/inteligencia-empresarial-inovadora/>>. Acesso em: 10 abr. 2020.

BAHRAMI, Mahdi; GHORBANI, Mazaher; ARABZAD, S. Mohammad. Information technology (IT) as an improvement tool for customer relationship management (CRM). **Procedia-Social and Behavioral Sciences**, v. 41, p. 59-64, 2012.

BEECK, Alexandre van. Como o novo perfil de compras impacta no varejo. **Mercado e Consumo**, São Paulo, 17 dez. 2018. Disponível em: <<https://mercadoeconsumo.com.br/2018/12/11/como-o-novo-perfil-de-compras-impacta-nos-canais-do-varejo/>>. Acesso em: 12 out. 2020.

BERNSTEIN, Abraham; PROVOST, Foster; HILL, Shawndra. Toward Intelligent Assistance for a Data Mining Process: An Ontology-Based Approach for Cost-Sensitive Classification. **IEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING**, [S.I.], v. 17., n. 4, 2005. Disponível em: <https://ieeexplore.ieee.org/document/1401890>. Acesso em: 29 mar. 2020.

CALDAS, Jéssica Acquarone Melo; RIBEIRO, Priscilla Cristina Cabral. **Análise da Contribuição da Tecnologia da Informação em Lojas de Conveniência no Rio de Janeiro**. 2017. 104 f. Projeto Final do Curso de Engenharia de Produção, Universidade Federal Fluminense, Niterói, 2017.

CARRERA, Heitor; ABREU, Juliana; WOODS, Douglas; LEONE, Eduardo; SONNEVELD, Silvia; XAVIER, André. **A corrida pelo varejo de conveniência no Brasil**. The Boston Consulting Group, 2017.

COMO explorar e gerenciar dados com o CRISP-DM. Semantix, São Paulo, 17 dez. 2018. Disponível em: <<https://semantix.com.br/como-explorar-e-gerenciar-dados-com-o-crisp-dm>> Acesso em: 13 out. 2020.

COMO o big data pode ajudar seus clientes do varejo?. Canal Westcon, [S.I.]. Disponível em: <<https://blogbrasil.westcon.com/como-o-big-data-pode-ajudar-seus-clientes-do-varejo>>. Acesso em: 25 abr. 2020.

CONHEÇA a história e evolução do varejo no Brasil. Solução Casting, [S.l.], 14 junho 2018. Disponível em: <<https://www.solucaocasting.com.br/blog/historia-e-evolucao-do-varejo-no-brasil/>>. Acesso em: 12 abr. 2020.

COSTA, Rêuder N. Cerqueira; JUNIOR, João B. Rocha. Processamento eficiente de regras de associação preferenciais. **Revista de Sistemas e Computação**, Salvador, v. 8, n. 2, p. 605-625, 2018.

DA SILVA, Diego Ferreira et al. Promoção De Vendas: Uma Ferramenta Estratégica Para a Fidelização De Clientes Em Tempos De Crise. **Reinpec**, v. 2, n. 1, p. 96–111, 2016.

DANTAS, E. R. G. et al. O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões. In: Simpósio de Excelencia em Gestão e Tecnologia, 5., 2008, Rio de Janeiro. **Anais eletrônicos [...]**. [S.l.], 2008. Disponível em: https://www.aedb.br/seget/arquivos/artigos08/331_331_Artigo_SEGET_EJDR_Versao_Final_010808.pdf. Acesso em: 10 maio 2020.

DAVENPORT, Thomas H; PRUSAK, Laurence. **Conhecimento empresarial**: como as organizações gerenciam o seu capital intelectual. 5. Ed. Rio de Janeiro: Campus, 1998.

DE OLHO no varejo. In: **ABRAS**. [S.l.], 8 fev. 2012. Disponível em: <<https://www.abras.com.br/home/redacao-portal/?materia=4231>>. Acesso em: 09 maio 2020.

DOTSIKA, Felfie. Uniting formal and informal descriptive power: Reconciling ontologies with folksonomies. **International Journal of Information Management**, Swansea, n. 29, p. 407–415, Out. 2009. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0268401209000127>>. Acesso em: 16 maio 2020.

DRESCH, A.; MIGUEL, P. A. C. Análise Dos Principais Métodos De Pesquisa Empregados Para a Condução De Estudos Que Abordam a Inovação No Brasil. **Revista Gestão Inovação e Tecnologias**, v. 5, n. 4, p. 2480–2494, 2015.

DRESCH, Aline; LACERDA, Daniel Pacheco; ANTUNES JR, José Antônio Valle. **Design Science Research**: Método de pesquisa para avanço da ciência e tecnologia. Porto Alegre: Bookman, 2015. Livro eletrônico.

FALEIROS, J. P. M. **Comércio varejista**: uma análise do comportamento do volume de vendas sob uma perspectiva não linear. *Revista do BNDES*, v. 32, p. 158, 2009.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, Providence, v. 17, n. 3, p. 37-54, 1996. Disponível em: <<https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>>. Acesso em: 11 abr. 2020.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2., 1996, Portland. **Anais eletrônicos** [...] Portland: Association for the Advancement of Artificial Intelligence, 1996. Disponível em: . Acesso em: 11 abr 2020.

FERRACCIÙ, João De Simoni Soderini. **Marketing promocional**: a evolução da promoção de vendas. 6. ed. São Paulo: Pearson, 2008. Livro eletrônico.

FERREIRA, Marcelo Sales; CAMPOS, Ivan Carneiro de; MACEDO, Marcelo Alvaro da Silva. Análise da importância do uso estratégico da informação no varejo supermercadista no Brasil. In: Encontro Nacional de Engenharia de Produção (ENEGEP), 26., 2006, Fortaleza. **Anais eletrônicos** [...]. Fortaleza: Associação Brasileira de Engenharia de Produção, 2006. Disponível em: http://www.abepro.org.br/biblioteca/ENEGEP2006_TR530352_7451.pdf. Acesso em: 01 abr. 2020.

FERREIRA, R. V.; CHEROBIM, A. P. M. S. Impacto dos investimentos em TI no desempenho organizacional de empresas de panificação de Minas Gerais: Estudo multicaso. **BASE** - Revista de Administração e Contabilidade da Unisinos, São Leopoldo, v. 9, n. 2, p. 147-161, 2012. Disponível em: <http://revistas.unisinos.br/index.php/base/article/view/base.2012.92.04>. Acesso em: 27 mar. 2020.

FIGUEIRA, Rafael. **Mineração de dados e bancos de dados orientados a objetos**. 1998. 96f. Dissertação (Mestrado em Ciência da Computação) – Faculdade de Ciências da Computação, Universidade Federal do Rio de Janeiro, 1998.

FILHO, Edival Ponciano de Carvalho. **Construção de taxonomias sobre informações compostas por descrições ambíguas com enriquecimento por meio de utilização de dicionários *On-line***. 2015. 157 f. Tese (Doutorado em Engenharia de Sistemas e Computação) – Programa de Pós-Graduação em Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

FRACALANZA, Livia Fonseca. **Mineração de Dados voltada para Recomendação no Âmbito de Marketing de Relacionamento**. 2009. 59 f. Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática, Pontifícia Universidade Católica do Rio de Janeiro (PUCRIO), Rio de Janeiro, 2009.

GAMARRA, C.; GUERRERO, J. M.; MONTERO, E. A knowledge discovery in databases approach for industrial microgrid planning. **Renewable and Sustainable Energy Reviews**, v. 60, p. 615–630, 2016.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017. Livro eletrônico.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2. ed. Rio de Janeiro: Elsevier, 2015. Livro eletrônico.

GOLFARELLI, Matteo; RIZZI, Stefano. Data warehouse testing: A prototype-based methodology. In: ESCHBACH, Robert; POORE, Jesse H. **Information and Software Technology**. n. 11. [S.l.]. 2011. p. 1183-1198.

GONÇALVES, Lorén Pinto Ferreira. **Mineração de Dados em Supermercados: o caso do supermercado “Tal”**. Dissertação (Mestrado em Administração) - Programa de Pós-Graduação em Administração (Convênio UFRS/URCAMP), Universidade Federal do Rio Grande do Sul, Porto Alegre, 1999.

GRAVES, Steven M. Convenience Stores: A Landscape Perspective. **Yearbook of the Association of Pacific Coast Geographers**, Hawai, n. 79, p. 134-152, 2017.

GRIVA, Anastasia et al. Retail business analytics: Customer visit segmentation using market basket data. **Expert Systems With Applications**, Grécia, p. 1-16, 2018.

GUIDOLIN, S. M.; COSTA, A. C. R.; NUNES, B. F. **Conectando indústria e consumidor: desafios do varejo brasileiro no mercado global**. BNDES Setorial, Rio de Janeiro, n. 30, p. 3- 61, 2009.

HAHSLER, Michael; GRÜN, Bettina; HORNIK, Kurt; BUCHTA, Christian. **Introduction to arules** – A computational environment for mining association rules and frequent item sets. Disponível em: <https://cran.r-project.org/web/packages/arules/vignettes/arules.pdf>. Acesso em: 16 jul. 2020.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. 3. ed. San Francisco: Morgan Kaufmann, 2012.

HAZOFF, Waldemar; SAUAIA, Antonio Carlos Aidar. O inimigo silencioso das lojas de conveniência em postos de serviços: um levantamento na zona centro-oeste da cidade de São Paulo. In: Seminário em Administração, 8., 2010, São Paulo. **Anais eletrônicos** [...] São Paulo: Universidade de São Paulo, 2010. Disponível em: <<http://sistema.semead.com.br/13semead/resultado/trabalhosPDF/813.pdf>>. Acesso em: 29 mar. 2020.

JUNIOR, Alberto Martins; CALLEGARI, Flávio Ibelli. **Data Warehouse: conceitos e aplicações no varejo**. [S.l.], [2020?]. Disponível em: <<http://giulianimarketing.pro.br/site/wp-content/uploads/2016/10/Data-Warehouse-conceitos-e-aplica%C3%A7%C3%A3o-no-varejo.pdf>>. Acesso em: 09 maio 2020.

KERCKHOFF, Ana Paula. **Market basket analysis utilizando data mining**. São Leopoldo: Unisinos, 2018.

KOIFFMAN, Pedro. Senso de Urgência um dos segredos do sucesso no varejo. **Ecommerce Brasil**, [S.l.], 16 dez. 2015. Disponível em: <<https://www.ecommercebrasil.com.br/artigos/senso-de-urgencia-um-dos-segredos-do-sucesso-no-varejo/>>. Acesso em: 01 maio 2020

KHOLOD, Marina. Market Basket Analysis of Convenience Store POS Data. **Science Reports - Tohoku University**, v. 71, p. 61–82, 2018.

KOTLER, Phillip; KELLER, Kevin Lane **Administração de marketing**. 14. ed. São Paulo: Pearson Education do Brasil, 2013.

KURNIAWAN, F. et al. Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data. **Knowledge Engineering and Data Science**, Indonesia, v. 1, n. 1, p. 20-25, 2018.

LACERDA, D. P. et al. **Design Science Research: A research method to production engineering**. *Gestão & Produção*, v. 20, n. 4, p. 741–761, 2013.

LAMBOIA, Fabiany; PEREIRA, Luciano Machado. **Data Mining**. Cascavel: Unioeste. 2005, p. 5-20.

LAROSE, Daniel; LAROSE, Chantal. **Data Mining and Predictive Analytics**. 2 ed. New Jersey: Wiley, 2015.

LEVY, M.; WEITZ, B. A. *Administração de varejo*. São Paulo: Atlas, 2000.

LIAO, S. H.; CHEN, C. M.; WU, C. H. Mining customer knowledge for product line and brand extension in retailing. **Expert Systems with Applications**, v. 34, n. 3, p. 1763–1776, 2008.

LIMA, José Wilson dos Santos. *Aplicação de Técnicas de Regras de Associação em Sistemas de Potência*. 2014. 70 f. Trabalho de Conclusão de Curso (Bacharel em Engenharia Elétrica) – Universidade Federal do Pará, Tucuruí, 2014.

LINOFF, Gordon S.; BERRY, Michael J. A. **Data mining techniques**: For marketing, Sales, and Customer Relationship Management. New York: Wiley Publishing, Inc., 2011.

MACHADO, Alexandre. *Conveniência é a bola da vez*. In: MERCADO & CONSUMO. [S.l.], 29 out. 2019. Disponível em: <https://mercadoeconsumo.com.br/2019/10/29/conveniencia-e-a-bola-da-vez/>. Acesso em: 24 maio 2020.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de Metodologia Científica**. Atlas, 2010.

MARTINS, Camila Delefrate. **Construção semi-automática de taxonomias para generalização de regras de associação**. 2006. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2006.

MATHEUS, Renato; PARREIRAS, Fernando. *Inteligencia Empresarial versus business intelligence: abordagens complementares para o apoio à tomada de decisão no Brasil*. In: CONGRESSO ANUAL DA SOCIEDADE BRASILEIRA DE GESTÃO DO CONHECIMENTO, 3., 2004, São Paulo. **Anais eletrônicos** [...]. Minas Gerais: Universidade Federal de Minas Gerais, 2004. Disponível em: <https://zenodo.org/record/12310#.X5n8nYhKjIU>. Acesso em: 21 mar. 2020.

MATTAR, Fauze Najib. **Administração de Varejo**. 1. ed. Rio de Janeiro: Elsevier, 2011.

MERCADO brasileiro disputa varejo de conveniência. Mundo do Marketing, [S.l.], 23 janeiro 2018. Disponível em: <<https://www.mundodomarketing.com.br/ultimas-noticias/37699/mercado-brasileiro-disputa-varejo-de-conveniencia.html>>. Acesso em: 31 mar. 2020.

MICHAELI, Rainer; SIMON, Lothar. An illustration of Bayes' theorem and its use as a decision-making aid for competitive intelligence and marketing analysts. **European Journal of Marketing**, v 42, n. 7/8, p. 804-813, 2008.

MORAIS, Izabelly Soares de. **Introdução a Big Data e Internet das Coisas (IoT)**. 1. Ed. Porto Alegre: SAGAH, 2018. Livro Eletrônico.

MORANDI, M. I. W. M.; CAMARGO, L. F. R. **Revisão Sistemática da Literatura**. In: Design Science Research: Método de Pesquisa para Avanço da Ciência e Tecnologia. 1. ed. p. 141-172. Porto Alegre: Bookman, 2015.

MORGADO, Maurício. **Três grandes desafios para o varejo do futuro**. GV-executivo, São Paulo, v. 16, n. 1, p. 33-35, 2017.

MORITZ, Gilberto de Oliveira; MORITZ, Mariana Oliveira; PEREIRA, Maurício Fernandes. **Planejamento por cenários prospectivos**. 1 ed. São Paulo: Atlas S.A., 2012.

NETO, Reinaldo Morabito; PUREZA, Vitoria. **Modelagem e Simulação**. In: Metodologia de pesquisa em engenharia de produção e gestão de operações. 2. ed. Rio de Janeiro: Elsevier, 2012. p. 169–198.

NOVAK, Joseph D.; GOWIN, D. Bob. **Aprender a aprender**. Lisboa, Plátano Edições Técnicas, 1996.

PAPAVASILEIOU, Vasilios; TSADIRAS, Athanasios. Evaluating time variations to identify valuable association rules in Market basket analysis. **Intelligent Decision Technologies**, Grécia, v. 7, n. 1, p. 81-90, 2013.

PERIZZOLO, Cesar B. **Sistema de Identificação Espacial em Ambientes de Consumo Utilizando Mineração de Dados**. 2005. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2005.

PEZZOTTI, Renato. Consumidor do futuro será mais seletivo e tecnológico, diz estudo da KPMG. *In*: UOL. São Paulo, 21 set 2020. Disponível em: <<https://economia.uol.com.br/noticias/redacao/2020/09/21/consumidor-do-futuro-sera-mais-seletivo-e-tecnologico-diz-estudo-da-kpmg.htm>>. Acesso em: 12 out. 2020.

PRASAD, Pramons; MALIK, Latesh. Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions. **International Journal on Computer Science and Engineering**. v. 3, n. 5, p. 2177–2182, 2011.

RIVA, João et al. **Marketing promocional**: um olhar descomplicado. São Paulo: Cengage Learning, 2014.

SALAZAR, Luis Henrique. **Desenvolvimento de uma Ferramenta para Auxiliar a Execução de Revisões Sistemáticas da Literatura**. 2015. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) - Universidade Federal de Santa Catarina, Florianópolis, 2015.

SANTOS, Nayara. Marketing de Varejo: O Guia Completo de Estratégias. *In*: **LISTENX**. [S.I.], 04 jun 2020. Disponível em: <https://listenx.com.br/blog/qual-a-importancia-do-marketing-no-varejo/>. Acesso em: 30 maio 2020.

SHAHRABI, J.; NEYESTANI, R. S. Discovering Iranians' Shopping Culture by Considering Virtual Items Using Data Mining Techniques. **Journal of Applied Sciences**, Iran, v. 9, n. 13, p. 2351-2361, 2009.

SILVA, Edna Lúcia; MENEZES, Estera Muszkat. **Metodologia da Pesquisa e Elaboração de Dissertação**. 4. ed. Florianópolis: UFSC, 2005. p. 138p, 2005.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI Clodis. **Introdução à Mineração de Dados Com aplicações em R**. 1. Ed. Rio de Janeiro: Elsevier, 2016.

SILVA, Lúcia A. **Atitude do consumidor em relação às marcas próprias de supermercados**: um estudo exploratório 2009. 235 f. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto. Universidade de São Paulo Ribeirão Preto. 2009.

SILVA, Wilson; CLARO, Genoveva Ribas; MENDES, Ademir. Aprendizagem significativa e mapas conceituais. *In*: Congresso Nacional de Educação, 8., 2017, Paraná. **Anais eletrônicos** [...]. Paraná: UNINTER, 2017.

SZAFIR-GOLDSTEIN, Cláudia; DE SOUZA, Cesar Alexandre. Tecnologia da Informação aplicada à gestão empresarial: um modelo para a empresa digital. In: Seminários em Administração (SEMEAD). 6., 2003, São Paulo. **Anais eletrônicos [...]** São Paulo: USP, 2003.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining Mineração de Dados**. Tradução de Acauan P. Fernandes. 1. ed. Rio de Janeiro: Ciência Moderna Ltda, 2009.

VENKITACHALAM, K.; AMBROSINI, V. A triadic link between knowledge management, information technology and business strategies. **Knowledge Management Research and Practice**, v. 15, n. 2, p. 192–200, 2017.

VILLARS, Richard L.; OLOFSON, Carl W.; EASTWOOD, Matthew. Big Data: What It is and Why You Should Care. **IDC White Paper**, 2011.

VITRINE DO VAREJO. Velocidade e conveniência estão entre as prioridades para o consumidor, sua loja está pronta?. [S.l., 2020?]. Disponível em: <<https://vitriedovarejo.com/velocidade-e-conveniencia-estao-entre-as-prioridades-para-o-consumidor-sua-loja-esta-pronta/>>. Acesso em: 10 abr. 2020.

WERNECK, Marcelo; BRONZO, Marcelo; OLIVEIRA, Marco Paulo Valadares de; SOUSA, Paulo Renato de. As capacidades analíticas e seu papel na geração de vantagens competitivas sustentáveis. **Harvard Business Review**, [S.l.], 25 out. 2019. Disponível em: <https://hbrbr.com.br/capacidades-analytics-sustentavel/>. Acesso em: 12 abr. 2020.

YANG, Yinghui; HAO, Chunhui. Product selection for promotion planning. *Knowledge and Information Systems*, [S.l.], v. 29, n. 1, p. 223-236, 2011.

**APÊNDICE A – ASSOCIAÇÕES GERADAS EM COMPARAÇÃO AOS PERÍODOS
DO DIA**

Associação Salgado Pão de Queijo ⇒ Café

(Continua)

PAO DE QUEIJO >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA- FEIRA	MANHÃ	MANHÃ1	467	26,84%	68,91%	3,85%
		MANHÃ2	548	31,49%		4,51%
		MANHÃ3	184	10,57%		1,52%
	TARDE	TARDE1	58	3,33%	19,94%	0,48%
		TARDE2	86	4,94%		0,71%
		TARDE3	203	11,67%		1,67%
	NOITE	NOITE1	149	8,56%	11,15%	1,23%
		NOITE2	45	2,59%		0,37%
TERÇA- FEIRA	MANHÃ	MANHÃ1	614	25,35%	67,18%	5,06%
		MANHÃ2	716	29,56%		5,90%
		MANHÃ3	297	12,26%		2,45%
	TARDE	TARDE1	57	2,35%	20,69%	0,47%
		TARDE2	155	6,40%		1,28%
		TARDE3	289	11,93%		2,38%
	NOITE	NOITE1	222	9,17%	12,14%	1,83%
		NOITE2	72	2,97%		0,59%
QUARTA- FEIRA	MANHÃ	MANHÃ1	450	26,60%	69,21%	3,71%
		MANHÃ2	518	30,61%		4,27%
		MANHÃ3	203	12,00%		1,67%
	TARDE	TARDE1	56	3,31%	20,09%	0,46%
		TARDE2	98	5,79%		0,81%
		TARDE3	186	10,99%		1,53%
	NOITE	NOITE1	151	8,92%	10,70%	1,24%
		NOITE2	30	1,77%		0,25%

(Conclusão)

PAO DE QUEIJO >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
QUINTA-FEIRA	MANHÃ	MANHÃ1	481	27,16%	71,71%	3,96%
		MANHÃ2	595	33,60%		4,90%
		MANHÃ3	194	10,95%		1,60%
	TARDE	TARDE1	57	3,22%	18,46%	0,47%
		TARDE2	74	4,18%		0,61%
		TARDE3	196	11,07%		1,61%
	NOITE	NOITE1	137	7,74%	9,82%	1,13%
		NOITE2	37	2,09%		0,30%
	SEXTA-FEIRA	MANHÃ	MANHÃ1	496	26,77%	66,16%
MANHÃ2			526	28,39%	4,33%	
MANHÃ3			204	11,01%	1,68%	
TARDE		TARDE1	72	3,89%	21,10%	0,59%
		TARDE2	118	6,37%		0,97%
		TARDE3	201	10,85%		1,66%
NOITE		NOITE1	177	9,55%	12,74%	1,46%
		NOITE2	59	3,18%		0,49%
SÁBADO		MANHÃ	MANHÃ1	305	19,37%	73,08%
	MANHÃ2		550	34,92%	4,53%	
	MANHÃ3		296	18,79%	2,44%	
	TARDE	TARDE1	71	4,51%	18,10%	0,58%
		TARDE2	91	5,78%		0,75%
		TARDE3	123	7,81%		1,01%
	NOITE	NOITE1	104	6,60%	8,83%	0,86%
		NOITE2	35	2,22%		0,29%
	DOMINGO	MANHÃ	MANHÃ1	55	5,06%	69,83%
MANHÃ2			407	37,44%	3,35%	
MANHÃ3			297	27,32%	2,45%	
TARDE		TARDE1	129	11,87%	17,85%	1,06%
		TARDE2	33	3,04%		0,27%
		TARDE3	32	2,94%		0,26%
NOITE		NOITE1	59	5,43%	12,33%	0,49%
		NOITE2	75	6,90%		0,62%

Associação Salgado Folhado ⇒ Café

(Continua)

SALGADO FOLHADO >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA-FEIRA	MANHÃ	MANHÃ1	124	27,74%	78,97%	3,91%
		MANHÃ2	153	34,23%		4,82%
		MANHÃ3	76	17,00%		2,39%
	TARDE	TARDE1	17	3,80%	14,77%	0,54%
		TARDE2	15	3,36%		0,47%
		TARDE3	34	7,61%		1,07%
	NOITE	NOITE1	11	2,46%	6,26%	0,35%
NOITE2		17	3,80%	0,54%		
TERÇA-FEIRA	MANHÃ	MANHÃ1	117	25,05%	70,24%	3,69%
		MANHÃ2	138	29,55%		4,35%
		MANHÃ3	73	15,63%		2,30%
	TARDE	TARDE1	12	2,57%	19,49%	0,38%
		TARDE2	27	5,78%		0,85%
		TARDE3	52	11,13%		1,64%
	NOITE	NOITE1	29	6,21%	10,28%	0,91%
NOITE2		19	4,07%	0,60%		
QUARTA-FEIRA	MANHÃ	MANHÃ1	100	22,88%	68,65%	3,15%
		MANHÃ2	155	35,47%		4,88%
		MANHÃ3	45	10,30%		1,42%
	TARDE	TARDE1	20	4,58%	20,82%	0,63%
		TARDE2	22	5,03%		0,69%
		TARDE3	49	11,21%		1,54%
	NOITE	NOITE1	32	7,32%	10,53%	1,01%
NOITE2		14	3,20%	0,44%		
QUINTA-FEIRA	MANHÃ	MANHÃ1	127	27,73%	71,18%	4,00%
		MANHÃ2	151	32,97%		4,76%
		MANHÃ3	48	10,48%		1,51%
	TARDE	TARDE1	13	2,84%	17,90%	0,41%
		TARDE2	21	4,59%		0,66%
		TARDE3	48	10,48%		1,51%
	NOITE	NOITE1	37	8,08%	10,92%	1,17%
NOITE2		13	2,84%	0,41%		

(Conclusão)

SALGADO FOLHADO >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEXTA-FEIRA	MANHÃ	MANHÃ1	154	27,85%	71,61%	4,85%
		MANHÃ2	177	32,01%		5,57%
		MANHÃ3	65	11,75%		2,05%
	TARDE	TARDE1	16	2,89%	17,18%	0,50%
		TARDE2	30	5,42%		0,94%
		TARDE3	49	8,86%		1,54%
	NOITE	NOITE1	39	7,05%	11,21%	1,23%
		NOITE2	23	4,16%		0,72%
	SÁBADO	MANHÃ	MANHÃ1	96	21,67%	81,49%
MANHÃ2			189	42,66%	5,95%	
MANHÃ3			76	17,16%	2,39%	
TARDE		TARDE1	17	3,84%	14,90%	0,54%
		TARDE2	15	3,39%		0,47%
		TARDE3	34	7,67%		1,07%
NOITE		NOITE1	11	2,48%	3,61%	0,35%
		NOITE2	5	1,13%		0,16%
DOMINGO		MANHÃ	MANHÃ1	12	3,24%	74,59%
	MANHÃ2		177	47,84%	5,57%	
	MANHÃ3		87	23,51%	2,74%	
	TARDE	TARDE1	11	2,97%	12,70%	0,35%
		TARDE2	15	4,05%		0,47%
		TARDE3	21	5,68%		0,66%
	NOITE	NOITE1	35	9,46%	12,70%	1,10%
		NOITE2	12	3,24%		0,38%

Associação Salgado Pastel ⇒ Refrigerante

(Continua)

SALGADO PASTEL >> REFRIGERANTE						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA-FEIRA	MANHÃ	MANHÃ1	67	7,92%	26,00%	1,23%
		MANHÃ2	76	8,98%		1,40%
		MANHÃ3	77	9,10%		1,42%
	TARDE	TARDE1	152	17,97%	52,60%	2,80%
		TARDE2	111	13,12%		2,04%
		TARDE3	182	21,51%		3,35%
	NOITE	NOITE1	131	15,48%	21,39%	2,41%
		NOITE2	50	5,91%		0,92%
TERÇA-FEIRA	MANHÃ	MANHÃ1	43	5,24%	23,51%	0,79%
		MANHÃ2	71	8,65%		1,31%
		MANHÃ3	79	9,62%		1,45%
	TARDE	TARDE1	150	18,27%	53,47%	2,76%
		TARDE2	126	15,35%		2,32%
		TARDE3	163	19,85%		3,00%
	NOITE	NOITE1	140	17,05%	23,02%	2,58%
		NOITE2	49	5,97%		0,90%
QUARTA-FEIRA	MANHÃ	MANHÃ1	37	4,47%	25,63%	0,68%
		MANHÃ2	85	10,28%		1,56%
		MANHÃ3	90	10,88%		1,66%
	TARDE	TARDE1	180	21,77%	54,29%	3,31%
		TARDE2	109	13,18%		2,01%
		TARDE3	160	19,35%		2,94%
	NOITE	NOITE1	120	14,51%	20,07%	2,21%
		NOITE2	46	5,56%		0,85%
QUINTA-FEIRA	MANHÃ	MANHÃ1	43	5,17%	22,98%	0,79%
		MANHÃ2	81	9,75%		1,49%
		MANHÃ3	67	8,06%		1,23%
	TARDE	TARDE1	156	18,77%	54,27%	2,87%
		TARDE2	129	15,52%		2,37%
		TARDE3	166	19,98%		3,05%
	NOITE	NOITE1	150	18,05%	22,74%	2,76%
		NOITE2	39	4,69%		0,72%

(Conclusão)

SALGADO PASTEL >> REFRIGERANTE						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEXTA-FEIRA	MANHÃ	MANHÃ1	65	6,27%	25,19%	1,20%
		MANHÃ2	111	10,71%		2,04%
		MANHÃ3	85	8,20%		1,56%
	TARDE	TARDE1	175	16,89%	51,25%	3,22%
		TARDE2	141	13,61%		2,59%
		TARDE3	215	20,75%		3,96%
	NOITE	NOITE1	183	17,66%	23,55%	3,37%
		NOITE2	61	5,89%		1,12%
	SÁBADO	MANHÃ	MANHÃ1	61	9,56%	42,16%
MANHÃ2			127	19,91%	2,34%	
MANHÃ3			81	12,70%	1,49%	
TARDE		TARDE1	124	19,44%	44,04%	2,28%
		TARDE2	75	11,76%		1,38%
		TARDE3	82	12,85%		1,51%
NOITE		NOITE1	61	9,56%	13,79%	1,12%
		NOITE2	27	4,23%		0,50%
DOMINGO		MANHÃ	MANHÃ1	23	5,26%	37,30%
	MANHÃ2		83	18,99%	1,53%	
	MANHÃ3		57	13,04%	1,05%	
	TARDE	TARDE1	63	14,42%	44,39%	1,16%
		TARDE2	72	16,48%		1,32%
		TARDE3	59	13,50%		1,09%
	NOITE	NOITE1	61	13,96%	18,31%	1,12%
		NOITE2	19	4,35%		0,35%

Associação Salgado Enroladinho ⇒ Refrigerante

(Continua)

SALGADO ENROLADINHO >> REFRIGERANTE						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA-FEIRA	MANHÃ	MANHÃ1	31	7,51%	27,12%	1,17%
		MANHÃ2	45	10,90%		1,69%
		MANHÃ3	36	8,72%		1,35%
	TARDE	TARDE1	54	13,08%	49,39%	2,03%
		TARDE2	60	14,53%		2,26%
		TARDE3	90	21,79%		3,38%
	NOITE	NOITE1	59	14,29%	23,49%	2,22%
		NOITE2	38	9,20%		1,43%
	TERÇA-FEIRA	MANHÃ	MANHÃ1	31	7,64%	27,59%
MANHÃ2			49	12,07%	1,84%	
MANHÃ3			32	7,88%	1,20%	
TARDE		TARDE1	57	14,04%	48,28%	2,14%
		TARDE2	61	15,02%		2,29%
		TARDE3	78	19,21%		2,93%
NOITE		NOITE1	69	17,00%	24,14%	2,59%
		NOITE2	29	7,14%		1,09%
QUARTA-FEIRA	MANHÃ	MANHÃ1	25	6,67%	26,93%	0,94%
		MANHÃ2	46	12,27%		1,73%
		MANHÃ3	30	8,00%		1,13%
	TARDE	TARDE1	54	14,40%	48,00%	2,03%
		TARDE2	42	11,20%		1,58%
		TARDE3	84	22,40%		3,16%
	NOITE	NOITE1	62	16,53%	25,07%	2,33%
		NOITE2	32	8,53%		1,20%
QUINTA-FEIRA	MANHÃ	MANHÃ1	36	8,16%	26,76%	1,35%
		MANHÃ2	37	8,39%		1,39%
		MANHÃ3	45	10,20%		1,69%
	TARDE	TARDE1	53	12,02%	45,58%	1,99%
		TARDE2	49	11,11%		1,84%
		TARDE3	99	22,45%		3,72%
	NOITE	NOITE1	86	19,50%	27,66%	3,23%
		NOITE2	36	8,16%		1,35%

(Conclusão)

SALGADO ENROLADINHO >> REFRIGERANTE						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEXTA-FEIRA	MANHÃ	MANHÃ1	41	8,70%	30,15%	1,54%
		MANHÃ2	54	11,46%		2,03%
		MANHÃ3	47	9,98%		1,77%
	TARDE	TARDE1	60	12,74%	42,89%	2,26%
		TARDE2	67	14,23%		2,52%
		TARDE3	75	15,92%		2,82%
	NOITE	NOITE1	91	19,32%	26,96%	3,42%
NOITE2		36	7,64%	1,35%		
SÁBADO	MANHÃ	MANHÃ1	19	5,83%	40,80%	0,71%
		MANHÃ2	61	18,71%		2,29%
		MANHÃ3	53	16,26%		1,99%
	TARDE	TARDE1	62	19,02%	42,33%	2,33%
		TARDE2	35	10,74%		1,32%
		TARDE3	41	12,58%		1,54%
	NOITE	NOITE1	36	11,04%	16,87%	1,35%
NOITE2		19	5,83%	0,71%		
DOMINGO	MANHÃ	MANHÃ1	13	5,70%	39,04%	0,49%
		MANHÃ2	35	15,35%		1,32%
		MANHÃ3	41	17,98%		1,54%
	TARDE	TARDE1	30	13,16%	40,79%	1,13%
		TARDE2	26	11,40%		0,98%
		TARDE3	37	16,23%		1,39%
	NOITE	NOITE1	30	13,16%	20,18%	1,13%
NOITE2		16	7,02%	0,60%		

Associação Salgado Pastel ⇒ Café

(Continua)

SALGADO PASTEL >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA-FEIRA	MANHÃ	MANHÃ1	260	21,89%	60,35%	3,34%
		MANHÃ2	364	30,64%		4,68%
		MANHÃ3	93	7,83%		1,20%
	TARDE	TARDE1	41	3,45%	22,22%	0,53%
		TARDE2	74	6,23%		0,95%
		TARDE3	149	12,54%		1,92%
NOITE	NOITE1	93	7,83%	17,42%	1,20%	
	NOITE2	114	9,60%		1,47%	
TERÇA-FEIRA	MANHÃ	MANHÃ1	277	25,14%	63,88%	3,56%
		MANHÃ2	317	28,77%		4,08%
		MANHÃ3	110	9,98%		1,42%
	TARDE	TARDE1	44	3,99%	23,23%	0,57%
		TARDE2	58	5,26%		0,75%
		TARDE3	154	13,97%		1,98%
NOITE	NOITE1	106	9,62%	12,89%	1,36%	
	NOITE2	36	3,27%		0,46%	
QUARTA-FEIRA	MANHÃ	MANHÃ1	263	21,14%	57,88%	3,38%
		MANHÃ2	365	29,34%		4,70%
		MANHÃ3	92	7,40%		1,18%
	TARDE	TARDE1	53	4,26%	28,05%	0,68%
		TARDE2	164	13,18%		2,11%
		TARDE3	132	10,61%		1,70%
NOITE	NOITE1	124	9,97%	14,07%	1,60%	
	NOITE2	51	4,10%		0,66%	
QUINTA-FEIRA	MANHÃ	MANHÃ1	376	30,40%	70,57%	4,84%
		MANHÃ2	374	30,23%		4,81%
		MANHÃ3	123	9,94%		1,58%
	TARDE	TARDE1	45	3,64%	18,27%	0,58%
		TARDE2	49	3,96%		0,63%
		TARDE3	132	10,67%		1,70%
NOITE	NOITE1	96	7,76%	11,16%	1,24%	
	NOITE2	42	3,40%		0,54%	

(Conclusão)

SALGADO PASTEL >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEXTA-FEIRA	MANHÃ	MANHÃ1	270	22,84%	66,67%	3,47%
		MANHÃ2	390	32,99%		5,02%
		MANHÃ3	128	10,83%		1,65%
	TARDE	TARDE1	55	4,65%	25,04%	0,71%
		TARDE2	81	6,85%		1,04%
		TARDE3	160	13,54%		2,06%
	NOITE	NOITE1	43	3,64%	8,29%	0,55%
		NOITE2	55	4,65%		0,71%
	SÁBADO	MANHÃ	MANHÃ1	232	22,46%	74,25%
MANHÃ2			348	33,69%	4,48%	
MANHÃ3			187	18,10%	2,41%	
TARDE		TARDE1	47	4,55%	17,23%	0,60%
		TARDE2	52	5,03%		0,67%
		TARDE3	79	7,65%		1,02%
NOITE		NOITE1	69	6,68%	8,52%	0,89%
		NOITE2	19	1,84%		0,24%
DOMINGO		MANHÃ	MANHÃ1	63	8,01%	70,01%
	MANHÃ2		334	42,44%	4,30%	
	MANHÃ3		154	19,57%	1,98%	
	TARDE	TARDE1	41	5,21%	16,65%	0,53%
		TARDE2	38	4,83%		0,49%
		TARDE3	52	6,61%		0,67%
	NOITE	NOITE1	85	10,80%	13,34%	1,09%
		NOITE2	20	2,54%		0,26%

Associação Salgado Enroladinho ⇒ Café

(Continua)

SALGADO ENROLADINHO >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEGUNDA-FEIRA	MANHÃ	MANHÃ1	83	18,40%	59,20%	2,56%
		MANHÃ2	135	29,93%		4,17%
		MANHÃ3	49	10,86%		1,51%
	TARDE	TARDE1	17	3,77%	25,28%	0,53%
		TARDE2	29	6,43%		0,90%
		TARDE3	68	15,08%		2,10%
	NOITE	NOITE1	48	10,64%	15,52%	1,48%
		NOITE2	22	4,88%		0,68%
TERÇA-FEIRA	MANHÃ	MANHÃ1	92	21,60%	56,10%	2,84%
		MANHÃ2	98	23,00%		3,03%
		MANHÃ3	49	11,50%		1,51%
	TARDE	TARDE1	16	3,76%	27,46%	0,49%
		TARDE2	31	7,28%		0,96%
		TARDE3	70	16,43%		2,16%
	NOITE	NOITE1	48	11,27%	16,43%	1,48%
		NOITE2	22	5,16%		0,68%
QUARTA-FEIRA	MANHÃ	MANHÃ1	88	19,64%	57,14%	2,72%
		MANHÃ2	116	25,89%		3,58%
		MANHÃ3	52	11,61%		1,61%
	TARDE	TARDE1	23	5,13%	24,55%	0,71%
		TARDE2	30	6,70%		0,93%
		TARDE3	57	12,72%		1,76%
	NOITE	NOITE1	58	12,95%	18,30%	1,79%
		NOITE2	24	5,36%		0,74%
QUINTA-FEIRA	MANHÃ	MANHÃ1	103	21,41%	65,07%	3,18%
		MANHÃ2	142	29,52%		4,39%
		MANHÃ3	68	14,14%		2,10%
	TARDE	TARDE1	15	3,12%	19,75%	0,46%
		TARDE2	22	4,57%		0,68%
		TARDE3	58	12,06%		1,79%
	NOITE	NOITE1	56	11,64%	15,18%	1,73%
		NOITE2	17	3,53%		0,53%

(Conclusão)

SALGADO ENROLADINHO >> CAFÉ						
DIA	TURNO	PERÍODO	QTD VENDIDA	TAXA DO PERÍODO	TAXA DO TURNO	TAXA SOBRE TOTAL
SEXTA-FEIRA	MANHÃ	MANHÃ1	97	17,64%	58,55%	3,00%
		MANHÃ2	168	30,55%		5,19%
		MANHÃ3	57	10,36%		1,76%
	TARDE	TARDE1	26	4,73%	23,64%	0,80%
		TARDE2	30	5,45%		0,93%
		TARDE3	74	13,45%		2,29%
	NOITE	NOITE1	66	12,00%	17,82%	2,04%
		NOITE2	32	5,82%		0,99%
	SÁBADO	MANHÃ	MANHÃ1	89	18,24%	70,49%
MANHÃ2			166	34,02%	5,13%	
MANHÃ3			89	18,24%	2,75%	
TARDE		TARDE1	34	6,97%	19,26%	1,05%
		TARDE2	22	4,51%		0,68%
		TARDE3	38	7,79%		1,17%
NOITE		NOITE1	36	7,38%	10,25%	1,11%
		NOITE2	14	2,87%		0,43%
DOMINGO		MANHÃ	MANHÃ1	42	10,71%	68,11%
	MANHÃ2		132	33,67%	4,08%	
	MANHÃ3		93	23,72%	2,87%	
	TARDE	TARDE1	16	4,08%	16,07%	0,49%
		TARDE2	16	4,08%		0,49%
		TARDE3	31	7,91%		0,96%
	NOITE	NOITE1	42	10,71%	15,82%	1,30%
		NOITE2	20	5,10%		0,62%