

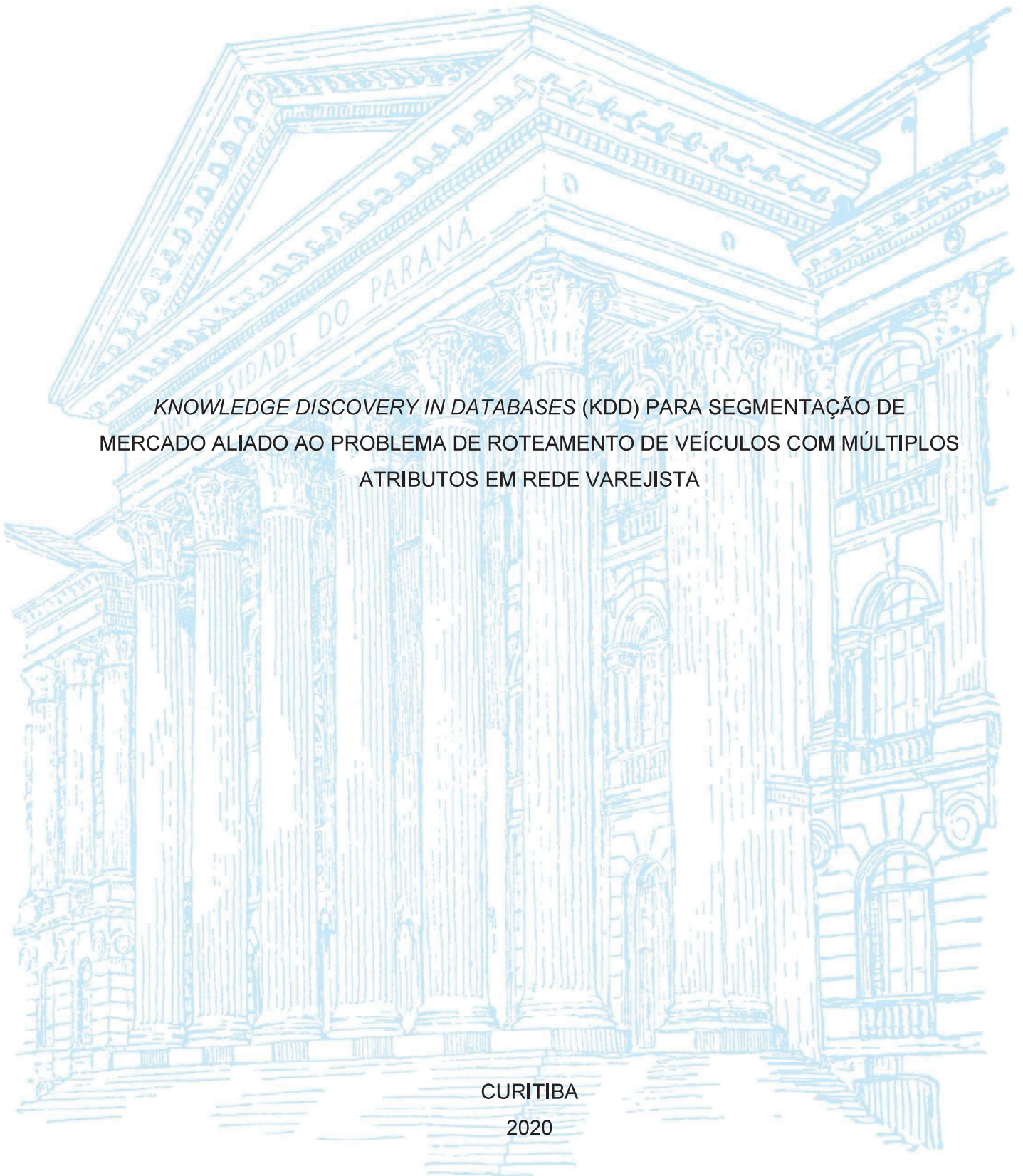
UNIVERSIDADE FEDERAL DO PARANÁ

KELLEN DAYELLE ENDLER

*KNOWLEDGE DISCOVERY IN DATABASES (KDD) PARA SEGMENTAÇÃO DE  
MERCADO ALIADO AO PROBLEMA DE ROTEAMENTO DE VEÍCULOS COM MÚLTIPLOS  
ATRIBUTOS EM REDE VAREJISTA*

CURITIBA

2020



KELLEN DAYELLE ENDLER

*KNOWLEDGE DISCOVERY IN DATABASES (KDD)* PARA SEGMENTAÇÃO DE  
MERCADO ALIADO AO PROBLEMA DE ROTEAMENTO DE VEÍCULOS COM MÚLTIPLOS  
ATRIBUTOS EM REDE VAREJISTA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Métodos Numéricos em Engenharia dos Setores de Ciências Exatas e de Tecnologia, da Universidade Federal do Paraná, área de concentração Programação Matemática, como requisito parcial para a obtenção do grau de Doutor.

Orientador: Prof. Dr. Cassius Tadeu Scarpin,  
Coorientadora: Prof. Dr. Maria Teresinha Arns Steiner.

CURITIBA

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR  
Biblioteca de Ciência e Tecnologia

E56k

Endler, Kellen Dayelle

*Knowledge Discovery in Databases (KDD)* para segmentação de mercado aliado ao problema de roteamento de veículos com múltiplos atributos em rede varejista [recurso eletrônico] / Kellen Dayelle Endler. – Curitiba, 2020.

Tese - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Métodos Numéricos em Engenharia, 2020.

Orientador: Cassius Tadeu Scarpin.

Coorientadora: Maria Teresinha Arns Steiner.

1. Mineração de dados (Computação). 2. Segmentação de mercado. 3. Comércio varejista. I. Universidade Federal do Paraná. II. Scarpin, Cassius Tadeu. III. Steiner, Maria Teresinha Arns. IV. Título.

CDD: 006

Bibliotecária: Vanusa Maciel CRB- 9/1928



## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **KELLEN DAYELLE ENDLER** intitulada: **KNOWLEDGE DISCOVERY IN DATABASES (KDD) PARA SEGMENTAÇÃO DE MERCADO ALIADO AO PROBLEMA DE ROTEAMENTO DE VEÍCULOS COM MÚLTIPLOS ATRIBUTOS EM REDE VAREJISTA**, sob orientação do Prof. Dr. CASSIUS TADEU SCARPIN, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 25 de Setembro de 2020.

Assinatura Eletrônica

25/09/2020 17:46:15.0

CASSIUS TADEU SCARPIN

Presidente da Banca Examinadora

Assinatura Eletrônica

25/09/2020 17:46:33.0

CLAUDIMAR PEREIRA DA VEIGA

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ - PPGOLD)

Assinatura Eletrônica

25/09/2020 17:48:33.0

JULIO CÉSAR NIEVOLA

Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ)

Assinatura Eletrônica

28/09/2020 15:20:49.0

EDUARDO ALVES PORTELA SANTOS

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

25/09/2020 18:04:53.0

JOSÉ EDUARDO PÉCORÁ JUNIOR

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Esse trabalho é dedicado à minha família.

## AGRADECIMENTOS

Em primeiro lugar, agradeço aos meus pais, Elton e Ana. Agradeço a Deus por ter me permitido ter pais como vocês, que sempre foram meu suporte em todos os sentidos, meus exemplos, minha inspiração. São eles os grandes responsáveis pela minha formação como pessoa e meus principais incentivadores. Agradeço a eles também pelo Kelvin, por me permitirem sentir como é grande o amor de irmão.

Agradeço ao meu marido Luiz Eduardo, por todo amor, carinho, dedicação e apoio durante esse período de mestrado e doutorado. Por ser meu suporte e minha motivação. Por me apoiar, acreditar em mim e compreender os momentos de restrições.

Agradeço ao meu orientador Dr. Cassius Tadeu Scarpin, a quem tenho uma profunda admiração pela sua trajetória e conquistas, pelas lições sobre fazer o que nos propomos com dedicação, comprometimento, energia e entusiasmo. Agradeço pelas oportunidades concedidas e pela confiança que sempre depositou em mim, pela liberdade que me concedeu para pesquisar e por dar o incentivo e suporte a esta pesquisa.

Agradeço à minha coorientadora Dra. Maria Teresinha Arns Steiner, a quem admiro não apenas pela disciplina com que conduz suas atividades sempre solícita em ajudar, mas também pela pessoa que é, que se dedica aos seus orientandos com todo o amor e carinho. Agradeço por ser minha incentivadora e por ser para mim uma inspiração como referência profissional.

Agradeço aos meus amigos do GTA0, em especial ao: Alexandre C. Choueiri, Cleder M. Schenekemberg, Deidson V. Kurpel, Nathália C. O. da Silva, Talita M. P. Schmidt, Tamires Sfeir, Thayse Duarte e Vinicius Martins Ton, pela generosidade em me ajudar tecnicamente e emocionalmente neste período.

Agradeço aos professores que com sua competência em compartilhar seus conhecimentos contribuíram para que essa fase fosse lembrada com orgulho. Em especial ao professor Dr. José Eduardo Pécora Júnior por me ajudar a encarar as dificuldades como uma evolução pessoal.

Agradeço também aos meus alunos, que ao participar do seu processo de desenvolvimento, me motivam a permanecer em evolução e em busca de conhecimento.

Meu agradecimento especial à minha família – tios, tias, avós e primos. Pela compreensão pelos momentos de ausência e pelo carinho dos momentos de encontro. Mesmo longe, vocês fazem toda a diferença.

Ao DAGA (Departamento de Administração Geral e Aplicada) da UFPR, por permitir a utilização do laboratório de computadores.

À CAPES, pelo apoio financeiro.

À todos que torceram por essa conquista.

“Se você pensa que pode  
ou se pensa que não pode,  
de qualquer forma você está certo”.

Henry Ford

## RESUMO

A possibilidade de prever atitudes e comportamentos de consumo, a partir do processamento adequado de informações armazenadas em bases de dados confere ao varejo uma vantagem competitiva em relação à concorrência. Da mesma forma, o planejamento eficiente de rotas para a entrega dos produtos ao cliente final é fundamental nesse segmento, visto que a experiência do consumidor está diretamente ligada ao recebimento da compra. Em geral, os trabalhos na literatura tratam do problema de roteirização de veículos ou do problema de segmentação de mercado, de forma independente. Raros são os trabalhos com a associação dos dois temas concomitantemente. Esta tese tem como objetivo desenvolver uma metodologia que possibilite considerar estratégias de segmentação de mercado, por meio do processo KDD, associadas a resolução do VRP para o planejamento de rotas para entregas de produtos no segmento varejista. Esta integração é cada vez mais realista em razão do aumento da comercialização em plataformas eletrônicas e, conseqüentemente, do volume de informações disponíveis dos clientes. Para tanto, o processo KDD é executado a partir de técnicas de: (i) agrupamento, utilizando-se Análise de Componentes Principais (ACP) seguida da Análise de Clusterização (AC), a fim de se criar grupos distintos de consumidores com base em critérios demográficos e comportamentais; e (ii) técnicas preditivas, utilizando-se de análise de Regressão Linear Múltipla (RLM) e de Redes Neurais Artificiais (RNAs), comparativamente, para prever o gasto gerado por clientes. Na sequência, desenvolveu-se um modelo matemático capaz de lidar com restrições complexas características do mercado varejista, denominado nesta tese por HMHDPCOMVRP (*Heterogeneous Multi-depot Periodic Close Open Mixed Vehicle Routing Problem* ou Problema de Roteamento de Veículos com Frota Heterogênea, Múltiplos Depósitos, Periódico, com Rotas Abertas e Fechadas). A pesquisa avança para uma versão estendida desse problema, considerando depósitos heterogêneos, assim como proposto por Calvet et al. (2016), a fim de considerar a minimização dos custos com deslocamento e a maximização do lucro estimado, por considerar as preferências dos clientes. Devido à complexidade do problema, um Algoritmo Híbrido com Particionamento de Conjuntos (HGA-SP – *Hybrid Genetic Algorithm with Set Partitioning*) foi proposto. A metodologia proposta foi ilustrada por meio da sua aplicação em caso real de um dos maiores grupos varejistas no segmento de hiper e supermercados da região sul do Brasil. Os resultados mostraram que é possível detectar relações entre os dados que permitiram prever o gasto dos clientes, considerando suas características e perfil de consumo. Quanto à solução do roteamento de veículos, os resultados do modelo de programação linear inteira mista são comparados com o algoritmo proposto. Os resultados do HGA-SP mostram uma melhoria percentual média acima de 75% em relação às soluções do modelo. Já para a versão estendida, esses valores são ainda melhores, superando o modelo em 87,5% considerando-se as melhores soluções e 81,25% considerando-se as soluções médias. Por fim, esta tese tem potencial de contribuir com a maneira de gerar abordagens inovadoras a fim de responder as principais questões de negócios varejistas relacionados a integração do processo de segmentação de mercado e planejamento de rotas.

Palavras-chave: Descoberta de conhecimento em dados, Problema de Roteirização de Veículos, Segmentação de Mercado, Múltiplos atributos, Varejo.



## ABSTRACT

The possibility of predicting consumer attitudes and behaviors, based on the appropriate processing of information stored in databases, gives retailers a competitive advantage over the competition. Likewise, efficient route planning for the delivery of products to the end customer is essential in this segment, since the consumer experience is directly linked to the receipt of the purchase. In general, works in the literature deal with the vehicle routing problem or the market segmentation problem, independently. Rare are the works with the association of the two themes concomitantly. This thesis aims to develop a methodology that makes it possible to consider market segmentation strategies, through the KDD process, associated with the VRP resolution for planning routes for product deliveries in the retail segment. This integration is increasingly realistic due to the increase in sales on electronic platforms and, consequently, the volume of information available to customers. For this, the KDD process is performed using techniques of: (i) grouping, using Principal Component Analysis (PCA) followed by Cluster Analysis (CA), in order to create different groups of consumers based on demographic and behavioral criteria; and (ii) predictive techniques, using Multiple Linear Regression analysis (MLR) and Artificial Neural Networks (ANNs), comparatively, to predict the expense generated by customers. In the sequence, a mathematical model capable of dealing with complex restrictions characteristic of the retail market was developed, named in this thesis by HMHDPCOMVRP (Heterogeneous Multi-depot Periodic Close Open Mixed Vehicle Routing Problem). The research advances to an extended version of this problem, considering heterogeneous deposits, as proposed by Calvet et al. (2016), to consider the minimization of travel costs and the maximization of estimated profit, as it considers customer preferences. Due to the complexity of the problem, a Hybrid Genetic Algorithm with Set Partitioning (HGA-SP) was proposed. The proposed methodology was illustrated through its application in the real case of one of the largest retail groups in the hyper and supermarket segment in the southern region of Brazil. The results showed that it is possible to detect relationships between the data that made it possible to predict customer spending, considering their characteristics and consumption profile. As for the vehicle routing solution, the results of the mixed integer linear programming model are compared with the proposed algorithm. The HGA-SP results show an average percentage improvement above 75% in relation to the model's solutions. For the extended version, these values are even better, surpassing the model by 87.5% considering the best solutions and 81.25% considering the average solutions. Finally, this thesis has the potential to contribute to the way of generating innovative approaches to answer the main questions of retail businesses related to the integration of the market segmentation process and route planning.

Keywords: Knowledge Discovery in Database. Vehicle Routing Problem. Market segmentation. Multiple attributes. Retailing.

## LISTA DE FIGURAS

FIGURA 1 - ETAPAS DO PROCESSO KDD .....	28
FIGURA 2 - REDE NEURAL <i>MULTY LAYER PERCEPTRON</i> .....	41
FIGURA 3 - PROCESSO DE SELEÇÃO DAS PUBLICAÇÕES.....	48
FIGURA 4 - TIPOS DE ESTUDOS EMPREGADOS .....	51
FIGURA 5 - DISTRIBUIÇÃO DE CARACTERÍSTICAS DE DADOS .....	52
FIGURA 6 - REDE DE COCITAÇÃO .....	62
FIGURA 7 - REPRESENTAÇÃO DO INDIVÍDUO E DA SOLUÇÃO .....	69
FIGURA 8 - REPRESENTAÇÃO DE UM CRUZAMENTO EM DOIS PONTOS.....	70
FIGURA 9 - CROSSOVER OX EM CROMOSSOMOS P1 E P2 (A) COM DELIMITADORES DE ROTA E (B) SEM DELIMITADORES DE ROTA. ....	70
FIGURA 10 - ESQUEMA DA METODOLOGIA ADOTADA.....	75
FIGURA 11 - PROCEDIMENTOS DO KDD .....	76
FIGURA 12 – DETALHAMENTO DAS TÉCNICAS DE AGRUPAMENTO (ACP E AC) .....	77
FIGURA 13 – DETALHAMENTO DAS TÉCNICAS DE PREDIÇÃO (RLM E RNA)...	79
FIGURA 14 –ABORDAGEM DE SOLUÇÃO PARA PROBLEMA DE ROTEIRIZAÇÃO .....	81
FIGURA 15 - FLUXOGRAMA DO HGA-SP.....	91
FIGURA 16 - ALGORITMO PARA SOLUÇÃO INICIAL .....	92
FIGURA 17 - EXEMPLO DO CÁLCULO DE CUSTO PARA O VFMP .....	98
FIGURA 18 - ILUSTRAÇÃO DE SOLUÇÕES FACTÍVEIS E INFACÍVEIS .....	98
FIGURA 19 - ALGORITMO PARA CALCULAR A CHAVE REFERENTE A UM VETOR $X$ .....	100
FIGURA 20 - EXEMPLO NUMÉRICO PARA O CÁLCULO DA CHAVE REFERENTE A UM VETOR $X$ .....	100
FIGURA 21 - ALGORITMO PARA O PROBLEMA DO CAMINHO MÍNIMO COM RESTRIÇÃO DE RECURSOS PARA HVRP .....	101
FIGURA 22 – EXEMPLO NUMÉRICO DO CÁLCULO DA CONTRIBUIÇÃO $\pi_k$ PARA A CHAVE ASSOCIADA AO ACRÉSCIMO DO VEÍCULO $k$ .....	102
FIGURA 23 - ALGORITMO PARA EXTRAIR AS ROTAS .....	103

FIGURA 24 - ALGORITMO PARA O PROCEDIMENTO <i>SPLIT</i> COM CARGA HETEROGÊNEA.....	104
FIGURA 25 - HISTOGRAMA RELACIONADO ÀS IDADES DOS CLIENTES .....	107
FIGURA 26 - APRESENTAÇÃO GRÁFICA DA MATRIZ DE CORRELAÇÃO. ....	109
FIGURA 27 - REPRESENTAÇÃO DA MATRIZ COM CARGAS FATORIAIS APÓS A ROTAÇÃO ORTOGONAL COM MÉTODO VARIMAX E COMUNALIDADES.....	112
FIGURA 28 - REPRESENTAÇÃO DOS ITENS RELACIONADOS AOS COMPONENTES .....	115
FIGURA 29 - DENDROGRAMA DOS CLUSTER FORMADOS.....	116
FIGURA 30 - REPRESENTAÇÃO DOS CLUSTER FORMADOS.....	117
FIGURA 31 - REPRESENTAÇÃO DAS 22 LOJAS LOCALIZADAS NO MUNICÍPIO DE CURITIBA .....	119
FIGURA 32 - REPRESENTAÇÃO GRÁFICA DOS RESÍDUOS VERSUS VALORES AJUSTADOS, QUANTIL QUANTIL DA NORMAL, RESÍDUOS VERSUS LEVERAGE .....	120
FIGURA 33 - HISTOGRAMA DA VARIÁVEL DEPENDENTE EM RELAÇÃO ÀS FUNÇÕES DE TRANSFORMAÇÃO .....	121
FIGURA 34 - REPRESENTAÇÃO GRÁFICA DOS RESÍDUOS APÓS TRANSFORMAÇÃO DE BOX-COX.....	124
FIGURA 35 - TEMPOS DE EXECUÇÃO.....	130
FIGURA 36 - VALORES OBSERVADOS E ESTIMADOS DE REDES .....	131
FIGURA 37 - REPRESENTAÇÃO DA REDE 15E12N22N1 – CLASSE 1 .....	132
FIGURA 38 - ILUSTRAÇÃO DO POSICIONAMENTO GEOGRÁFICO DOS CLIENTES E DAS LOJAS CONSIDERADAS .....	135
FIGURA 39 - ILUSTRAÇÃO DO RESULTADO DO EXEMPLO NUMÉRICO CONSIDERANDO A (I) MINIMIZAÇÃO DE DISTÂNCIA E A (II) MAXIMIZAÇÃO DE BENEFÍCIOS .....	139
FIGURA 40 - EXEMPLO DA INFLUÊNCIA DO FATOR DE DIVERSIDADE EM RELAÇÃO A ESCOLHA DE INDIVÍDUOS PARA A POPULAÇÃO. ....	141
FIGURA 41 - IMPACTO DO COEFICIENTE DE GRANULOMETRIA EM TERMOS DE QUALIDADE DE SOLUÇÃO E TEMPO COMPUTACIONAL.....	142
FIGURA 42 - TESTES PARA AVALIAÇÃO DO IMPACTO DO SP .....	142

FIGURA 43 - TESTES PARA AVALIAÇÃO NO NÚMERO DE INDIVÍDUOS DA POPULAÇÃO.....	143
FIGURA 44 - BOXPLOT COM DISTRIBUIÇÃO DE DISTÂNCIA E VALOR MONETÁRIO EM RELAÇÃO AS SOLUÇÕES POR MODELO .....	146
FIGURA 45 - BOXPLOT COM DISTRIBUIÇÃO DE BENEFÍCIOS OBTIDOS DAS SOLUÇÕES POR MODELO .....	147

## LISTA DE GRÁFICOS

GRÁFICO 1 - GRÁFICO SCREE DOS AUTOVALORES VERSUS O NÚMERO DE COMPONENTES, POR ORDEM DE EXTRAÇÃO .....	111
---	-----

## LISTA DE QUADROS

QUADRO 1 - TAXONOMIA DOS RVRP PROPOSTA POR (LAHYANI; KHEMAKHEM; SEMET, 2015).....	47
QUADRO 2 - TAXONOMIA DO VRP PROPOSTA POR (BRAEKERS; RAMAEKERS; VAN NIEUWENHUYSE, 2016).....	47
QUADRO 3 - RELAÇÃO DAS CARACTERÍSTICAS DOS ARTIGOS COM ALGORITMOS UNIFICADOS PARA MAVRP AVALIADOS.....	64
QUADRO 4 - VARIANTES FREQUENTEMENTE ENCONTRADAS NA LITERATURA.....	67

## LISTA DE TABELAS

TABELA 1 – SIGNIFICÂNCIA DAS CARGAS FATORIAIS BASEADO NO TAMANHO DA AMOSTRA.....	32
TABELA 2 – DISTRIBUIÇÃO DE PUBLICAÇÕES EM JORNAIS SELECIONADOS .....	49
TABELA 3 – CARACTERÍSTICAS DE CENÁRIOS .....	53
TABELA 4 – CARACTERÍSTICAS FÍSICAS DO PROBLEMA.....	57
TABELA 5 – RELAÇÃO DOS ATRIBUTOS RELACIONADOS AOS EVENTOS DE COMPRAS.....	78
TABELA 6 - NOTAÇÕES USADAS PARA O MODELO.....	84
TABELA 7 – NOTAÇÕES USADAS NO PARTICIONAMENTO DE CONJUNTOS	105
TABELA 8 - MEASURE OF SAMPLING ADEQUACY (MSA) .....	110
TABELA 9- VALORES DA VARIÂNCIA TOTAL EXPLICADA PELOS COMPONENTES IDENTIFICADOS NA ANÁLISE FATORIAL .....	110
TABELA 10 – DESCRIÇÃO DOS GRUPOS .....	116
TABELA 11 – TESTES DE NORMALIDADE PARA A VARIÁVEL DEPENDENTE	122
TABELA 12 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO <i>STEPWISE</i> PARA CLASSE 1 (DADOS TREINAMENTO).....	122
TABELA 13 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO <i>STEPWISE</i> PARA CLASSE 2.....	125
TABELA 14 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO <i>STEPWISE</i> PARA CLASSE 3.....	126
TABELA 15 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO <i>STEPWISE</i> PARA CLASSE 4.....	126
TABELA 16 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO <i>STEPWISE</i> PARA CLASSE 5.....	127
TABELA 17 – RESULTADO DAS ESTATÍSTICAS ANALISADAS DOS RESÍDUOS PARA CADA CLASSE .....	127
TABELA 18 – ANÁLISE DE VARIÂNCIA DO AJUSTE DOS MODELOS E MEDIDAS DE ERROS PARA CADA CLASSE.....	128
TABELA 19 – RESULTADO DAS CONFIGURAÇÕES DE REDES COM MELHORES PERFORMANCE .....	130

TABELA 20 – RESULTADOS DE TREINO E TESTE PARA AS RNAS.....	132
TABELA 21 – CARACTERÍSTICAS RELACIONADAS AOS TIPOS DE VEÍCULOS .....	134
TABELA 22 - CARACTERÍSTICAS DAS INSTANCIAS GERADAS.....	136
TABELA 23 – CARACTERÍSTICAS DO EXEMPLO NUMÉRICO .....	137
TABELA 24 – PARÂMETROS CONSIDERADOS PARA O EXEMPLO .....	138
TABELA 25 – RESULTADOS OBTIDOS COM MODELO MATEMÁTICO PLIM E COM O HGA-SP PARA O HMDPCOMVRP.....	144
TABELA 26 – RESULTADOS OBTIDOS COM MODELO MATEMÁTICO PLIM E COM O HGAS-SP PARA O HMDPCOMVRP COM VERSÃO ESTENDIDA.....	145



## LISTA DE SIGLAS

2E-RLP	- <i>Two-Echelon Time-Constrained Vehicle Routing Problem</i>
AC	- Análise de Clusterização
ACO	- <i>Ant Colony Optimization</i>
ACP	- Análise dos Componentes Principais
AD	- Anderson-Darling
AFC	- Análise dos Fatores Comuns
ALNS	- <i>Adaptive Large Neighborhood Search</i>
AS	- <i>Simulated Annealing</i>
AVRP	- <i>Asymmetric Vehicle Routing Problem</i> - Problema de Roteirização de Veículos Assimétrico
B2C	- Business-to-Consumer
BC	- <i>Branch and Cut</i>
BCP	- <i>Branch-and-Cut-and-Price</i>
BKS	- <i>Best Known Solution</i> ou Melhor Solução Encontrada
CCVRP	- <i>Cumulative Vehicle Routing Problem</i> ou Problema de Roteirização de Veículos Cumulativo
COMVRP	- <i>Close-Open Mixed Vehicle Routing Problem</i> - Problema de Roteirização de Veículos Misto Aberto e Fechado
CP	- Componentes Principais
CVRP	- <i>Capacitated Vehicle Routing Problem</i> ou Problema de Roteirização de Veículos Capacitados
DARP	- <i>Dial-a-Ride Problem</i> ou Problema <i>Dial-a-Ride</i>
DurVRP	- <i>Duration constraints</i> ou Restrições de duração da rota
DVRP	- <i>Dynamic Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos Dinâmico
FIFO	- <i>First-In-First-Out</i> ou Primeiro que Entra, Primeiro que Sai
FPVRP-IC	- <i>Flexible Periodic Vehicle Routing Problem with Inventory-Constraints</i> ou Problema de Roteamento de Veículos Periódico Flexível
FSM	- <i>Fleet Size and Mix</i> ou Problema de Roteamento de Veículos com Dimensionamento de Frota Heterogênea
GA	- <i>Genetic Algorithms</i> ou Algoritmo Genético
GVRP	- Generalized VRP
HEA	- Hybrid Evolutionary Algorithm
HFFVRP	- <i>Heterogeneous Fixed Fleet Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Frota Heterogênea Fixa
HFVRP	- <i>Heterogeneous Fleet Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Frota Heterogênea
HGA-SP	- <i>Hybrid Genetic Algorithm with Set Partitioning</i> ou Algoritmo Genético Híbrido com Particionamento de Conjuntos
HGSADC	- Hybrid Genetic Search with Advanced Diversity Control ou Busca Híbrida Genética com Controle Avançado de Diversidade
HMDPCOMVRP	- <i>Heterogeneous Multi-depot Periodic Close Open Mixed Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Frota Heterogênea, Múltiplos Depósitos, Periódico, com Rotas Abertas e Fechadas.

HVRP	- <i>Heterogeneous Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Frota Heterogênea
HVRPTW	- <i>Hard Time Windows Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Janela de Tempo Rígidas
ILS	- <i>Iterated Local Search</i>
IRP	- <i>Inventory Routing Problems</i> ou Problemas de Roteamento e Estoque
KDD	- <i>Knowledge Discovery in Databases</i> ou Descoberta de Conhecimento em Base de Dados.
LDVRP	- <i>Load-dependent Vehicle Routing Problem</i> ou Problema de Roteirização de Veículos dependente da carga
LIFO	- <i>Last-In-First-Out</i>
LNS	- <i>Large Neighborhood Search</i>
LoVRP	- <i>Load-Dependent Vehicle Routing Problem</i> ou Problema de Roteamento de Veículo Dependente da Carga
LRP	- <i>Location-Routing Problems</i> ou problemas de roteamento e localização
MAVRP	- <i>Multi-attribute Vehicle routing Problem</i> ou Problema de Roteamento de Veículos com Múltiplos Atributos
MDVRP	- <i>Multi-depot Vehicle routing Problem</i> ou Problema de Roteamento de Veículos com Múltiplos depósitos
MLP	- <i>Multi Layer Perceptron</i>
MSA	- <i>Measure of Sampling Adequacy</i> ou Medida de Adequação de Amostra
MTVRP	- <i>Multiple trips Vehicle Routing Problem</i> ou Problema de Roteirização com múltiplas viagens
OVRP	- <i>Open Vehicle Routing Problem</i> ou Problema de Roteirização de Veículos Aberto
OX	- <i>Order Crossover</i>
PLIM	- Programação Linear Inteira Mista
PO	- Pesquisa Operacional
PPC	- <i>Physical Problems Characteristics</i> ou características físicas do problema
PRP	- <i>Production-Routing Problem</i> ou Problema de Produção e Roteirização
PSO	- <i>Particle Swarm Optimization</i>
PVRP	- <i>Periodic Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos Periódico
RDVRP	- <i>Route Duration Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Duração de Rota
RLM	- Regressão Linear Múltipla
RNA	- Rede Neural Artificial
RRVRP	- Problema de Roteamento de Veículos <i>Rollon-Rolloff</i>
RVRP	- <i>Rich Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos Ricos
SC	- <i>Scenario Characteristics</i> ou Características do Cenário
SCM	- <i>Supply Chain Management</i> ou Gestão da Cadeia de Suprimentos
SDVRP	- <i>Vehicle Routing Problem with Site-dependent</i> ou Problema de Roteamento de Veículos com e Restrições de Acesso
SDVRP	- <i>Vehicle Routing Problem with Split Deliveries</i> ou Problema de Roteamento de Veículos com Entregas Fracionadas
SP	- <i>Set Partitioning</i> ou Problema de Particionamento de Conjuntos
TAC	- Transportador Autônomo de Cargas

TDVRP	- <i>Time-Dependent Vehicle Routing Problem</i> ou Problema de Roteamento Dependente do Tempo
TS	- <i>Tabu Search</i>
TSP	- <i>Travelling Salesman Problem</i>
UHGS	- <i>Unified Hybrid Genetic Search</i> ()
VDS	- <i>Vehicle and Driver Scheduling</i> ou Problema de Programação de veículos e Motoristas
VFMP	- <i>Vehicle Fleet Mix Problem</i> ou Problema de Roteamento de Veículos com Dimensionamento de Frota Heterogênea
VFMPWTW	- <i>Vehicle Fleet Mix with Time Windows Problem</i> ou Problema de Roteamento de Veículos com Dimensionamento de Frota Heterogênea com Janela de Tempo
VNS	- <i>Variable Neighborhood Search</i> ou Busca em Vizinhança Variável
VRP	- <i>Vehicle Routing Problem</i> ou Roteirização de Veículos
VRPB	- <i>Vehicle Routing Problem with Backhaul</i> ou Problema de Roteirização de Veículos com <i>Backhaul</i>
VRPFTT	- <i>Vehicle Routing Problem with Flexible Travel Time</i> ou Problema de Roteirização de Veículos com Tempo de Viagem Flexível
VRPGTW	- <i>General TW Vehicle Routing Problem</i>
VRPLB	- <i>Lunch breaks Vehicle routing Problem</i> ou Problema de Roteamento de Veículos com paradas para almoço
VRPMDP	- <i>Vehicle routing Problem with Mixed Pickup &amp; deliveries</i> ou Problema de Roteamento de Veículos com Coletas e Entregas Mistas
VRPMTW	- <i>Vehicle routing Problem with Multiple TW</i> ou Problema de Roteamento de Veículos com Múltiplas Janelas de Tempo
VRPPD	- <i>Vehicle Routing Problems with Pickups and Deliveries</i> ou Problemas de Roteamento de Veículos com Coletas e Entregas
VRPSD	- <i>Vehicle Routing Problem with Split deliveries</i> ou Problema de Roteamento de Veículos com entrega simultânea
VRPSPD	- <i>Vehicle Routing Problem with Simultaneous Pickup and Delivery</i> ou Problema de Roteamento de Veículos com coleta e entrega simultânea
VRPSTW	- <i>Soft Time Windows Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Janela de Tempo Flexíveis
VRPTW	- <i>Time Windows Vehicle Routing Problem</i> ou Problema de Roteamento de Veículos com Janela de Tempo
VRTDSP	- <i>Vehicle Routing And Truck Driver Scheduling Problem</i> ou Problemas de Programação e Roteamento de Caminhões
WCVRP	- <i>Waste Collection Vehicle Routing Problem</i> ou Roteamento de Veículos na Coleta de Resíduos

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>16</b>
1.1	DELIMITAÇÃO DO TEMA .....	16
1.2	OBJETIVOS .....	20
1.2.1	Objetivo Geral.....	20
1.2.2	Objetivos específicos.....	20
1.3	JUSTIFICATIVA E RELEVÂNCIA.....	21
1.4	CONTRIBUIÇÕES.....	24
1.5	ORGANIZAÇÃO DO TRABALHO.....	25
<b>2</b>	<b>REVISÃO DA LITERATURA .....</b>	<b>26</b>
2.1	SEGMENTAÇÃO DE MERCADO.....	26
2.1.1	Processo KDD .....	28
2.1.1.1	Técnicas de Agrupamento.....	29
2.1.1.2	Técnicas de Predição .....	34
2.2	CONCEITOS SOBRE O VRP .....	44
2.2.1	Revisão Sistemática do VRP .....	45
2.2.1.1	Procedimentos para revisão sistemática .....	47
2.2.1.2	Tipos de estudos e características dos dados.....	50
2.2.1.3	Características de Cenário .....	52
2.2.1.4	Características físicas do problema.....	57
2.2.1.5	Pilares teóricos e Oportunidades de Pesquisa .....	61
2.2.2	Revisão de características de algoritmos para o MAVRP .....	63
2.2.2.1	Descrição das variantes clássicas associadas ao problema estudado.....	65
2.2.2.2	Principais características dos algoritmos para MAVRP .....	67
2.3	CONSIDERAÇÕES FINAIS.....	73
<b>3</b>	<b>METODOLOGIA.....</b>	<b>75</b>
3.1	DESCRIÇÃO DO PROBLEMA DE ROTEAMENTO DE VEÍCULOS E FORMULAÇÃO MATEMÁTICA PROPOSTA.....	82
3.1.1	Definição do problema.....	82
3.1.2	Definições e notações .....	84
3.1.3	Modelo de Programação Inteira Mista para o HMDPCOMVRP.....	85
3.1.4	Modelo de Programação Inteira Mista HMDPCOMVRP com depósitos heterogêneos .....	87

3.1.4.1	Desigualdades válidas.....	88
3.2	ABORDAGEM DE SOLUÇÃO VIA ALGORITMO HEURÍSTICO HÍBRIDO COM PARTICIONAMENTO DE CONJUNTOS (HGA-SP).....	89
3.2.1	O Procedimento de Solução Inicial.....	92
3.2.2	O Procedimento de Melhoria de Rota.....	93
3.2.3	Espaço de busca .....	95
3.2.4	Gestão da População .....	96
3.2.5	Representação da Solução .....	97
3.2.6	Procedimento <i>Split</i> .....	97
3.2.7	Particionamento de Conjuntos.....	104
<b>4</b>	<b>RESULTADOS .....</b>	<b>107</b>
4.1	PROCESSO KDD.....	107
4.1.1	Técnicas de Agrupamento.....	108
4.1.1.1	Análise de Componentes Principais (ACP) .....	108
4.1.1.2	Análise de Clusterização (AC) .....	115
4.1.2	Técnicas de Predição .....	119
4.1.2.1	Regressão Linear Múltipla (RLM) .....	119
4.1.2.2	Redes Neurais Artificiais (RNA).....	128
4.2	RESULTADOS COMPUTACIONAIS PARA O VRP.....	132
4.2.1	Descrição das Instâncias.....	133
4.2.2	Exemplo Numérico .....	136
4.2.3	Avaliação dos Parâmetros.....	140
4.2.4	Resultados Testes Computacionais .....	143
<b>5</b>	<b>CONCLUSÕES .....</b>	<b>148</b>
	<b>REFERÊNCIAS.....</b>	<b>153</b>
	<b>ANEXO 1 – MEDIDAS DE POSIÇÃO REFERENTE À ETAPA DE ESTATÍSTICA DESCRITIVA DOS DADOS .....</b>	<b>167</b>
	<b>ANEXO 2 – CARGAS FATORIAIS.....</b>	<b>168</b>
	<b>ANEXO 3 – REPRESENTAÇÃO DAS REDES NEURAIS PARA AS CLASSES 2 A</b>	
<b>5</b>	<b>172</b>	

## 1 INTRODUÇÃO

Nesta introdução uma visão geral da temática é apresentada por meio da exposição da ideia da integração da segmentação de mercado com problema de roteirização, juntamente com suas variantes tratadas neste trabalho no contexto varejista. Além disso, apresentam-se os objetivos propostos para a tese a justificativa da associação do roteamento de veículos ao processo de segmentação de mercado utilizando-se Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases* - KDD). Por fim, são apresentadas as contribuições e a organização do presente trabalho.

### 1.1 DELIMITAÇÃO DO TEMA

O problema da distribuição física de mercadorias consiste em um elo da cadeia de suprimentos responsável por gerir o fluxo de produtos desde um depósito, armazém ou centros de distribuição, por meio de uma rede de transportes, para os clientes ou consumidores. A literatura aborda este problema como VRP, introduzido pela literatura por Dantzig e Ramser (1959), sendo um tema clássico da Pesquisa Operacional (PO). O VRP é um nome genérico para uma classe de problemas de otimização combinatória em que pontos de um grafo devem ser ordenados em uma ou mais sequências de forma mais econômica possível. Estas sequências são chamadas de rotas, as quais são percorridas por veículos, este conjunto chamado de frota, fazendo coleta e/ou entregas de itens e os pontos do grafo são os clientes dispersos geograficamente, os quais devem ser atendidos conforme necessidades específicas de cada problema. Os veículos deixam o depósito, servem os clientes na rede e retornam ao depósito após completar suas rotas e cada cliente é descrito por uma certa demanda. Uma revisão de literatura recente pode ser encontrada em Braekers, Ramaekers e Van Nieuwenhuyse (2016). Esta tese concentra seus esforços no problema logístico do planejamento de rotas para redes varejistas.

O varejo constitui um dos setores mais representativos na economia mundial (MOU; ROBB; DEHORATIUS, 2018) e tem assumido importância crescente no panorama empresarial de diversos países. Trata-se de uma das atividades empresariais com maior ritmo de transformações, respondendo às modificações do

ambiente tecnológico, econômico e social em que está inserido (BACH et al., 2020; PARENTE, 2010; REINARTZ; WIEGAND; IMSCHLOSS, 2019). No Brasil este fenômeno não é diferente. O setor varejista de bens e consumo brasileiro representa 22,3% dos empregos formais e uma participação de 20,42% do produto interno bruto do país em 2018 (SBVC, 2019). Especificamente, o varejo supermercadista brasileiro, objeto de estudo, registrou uma participação de 5,2% do PIB de acordo com a pesquisa Ranking ABRAS/SuperHiper (ABRAS, 2019).

O varejo supermercadista é um dos principais elos que as indústrias e fornecedores mantêm com seus consumidores finais (MADEIRA, 2019). À medida que a tecnologia avança, informações valiosas, tais como preferências dos consumidores e hábitos de consumo, podem ser coletadas dos bancos de dados de transações em uma loja (AGRAWAL; IMIELIŃSKI; SWAMI, 1993; KIANI, 2020; SRIKANT; AGRAWAL, 1997). Diversos varejistas no Brasil “não sabem o poder dos dados que possuem ou mesmo não conseguem transformá-los em uma estratégia eficiente de negócios, colocando a perder grandes oportunidades” (ABRAS, 2019). Em futuro próximo, é certo que o varejo supermercadista brasileiro vai passar por uma série de ajustes competitivos, pressionado por uma concorrência irreversível (BACH et al., 2020; SAAB; GIMENEZ, 2000; VAROTTO, 2018), o que levará as empresas a explorarem novas vantagens competitivas.

Neste contexto, a primeira parte deste trabalho consiste no estudo da exploração de informações armazenadas em bases de dados de redes varejistas. Estes estudos têm se mostrado um elemento fundamental para contribuir com as estratégias de *marketing* das organizações, de modo a direcionar os estímulos mais adequados a cada perfil de consumidor (SCHIFFMAN; KANUK, 1991). Afinal, técnicas de Mineração de Dados podem possibilitar a previsão de atitudes e comportamentos de consumo.

Associa-se à essa percepção do mercado, o planejamento de rotas para entregas dos produtos aos clientes, abordados de forma integrada. Esta tese tem o propósito de unir os conceitos de segmentação de mercado, sob a ótica de preferências de compras do cliente, com decisões de roteirização de veículos. Para isso, utiliza-se uma rede supermercadista brasileira que possui diversas lojas em uma grande capital do Sul do Brasil. As lojas desta rede possuem características, como tamanho físico, sortimento de produtos e perfil de clientes, diferentes. Esta

heterogeneidade em termos de ofertas comerciais, implica em uma vontade de consumo por parte dos clientes diferente de loja para loja, dependendo de como suas preferências são atendidas. Essas preferências, ao serem identificadas, podem ser utilizadas para a atração de clientes para compras em outras lojas da rede, nas quais normalmente, este cliente não iria adquirir seus itens de consumo, seja por desconhecimento ou por falta de estímulo.

Por sua vez, analisando-se o banco de dados relacionado a preferências de compras e características dos clientes, pode-se oportunizar a oferta de itens de outras lojas da rede, adequadas ao perfil do cliente. Uma vez que se identificou essa possibilidade, direcionou-se ofertas e efetivou-se essa venda, pouco provável anteriormente, a entrega desses itens, quando for o caso, sofrerá alterações em termos de roteiros de veículos. Essas alterações são foco deste estudo. Visa-se compreender a relação de ganho em alterar as rotas tradicionais de cada loja em função de estímulos de vendas para clientes não atingidos diretamente por cada loja, mas de conhecimento da Rede Supermercadista.

Esse impacto nas rotas tem características específicas. Existe, na literatura, uma gama extremamente ampla de variantes relacionada ao VRP de forma a compreender características e restrições reais adicionais. Nesta tese, o problema de Roteamento de Veículos com Frota Heterogênea, Múltiplos Depósitos, Periódico, com rotas Abertas e Fechadas (*Heterogeneous Multi-depot Periodic Close Open Mixed Vehicle Routing Problem - HMDPCOMVRP*) foi desenvolvido a partir da integração de características complementares, as quais foram previamente abordadas de forma independentes na literatura.

Para tanto, parte-se do problema clássico do Problema de Roteirização de Veículos Capacitados (*Capacitated Vehicle Routing Problem – CVRP*), no qual os veículos possuem limitação de capacidade, há somente um depósito, do qual a rota deve ser iniciada e terminada e a frota de veículos é homogênea. Este problema ao ser avaliado simultaneamente ao Problema de Roteirização de Veículos Aberto (*Open Vehicle Routing Problem - OVRP*), é conhecido como Problema de Roteirização de Veículos Misto Aberto e Fechado (*Close–Open Mixed Vehicle Routing Problem - COMVRP*). Liu e Jiang (2012) foram os primeiros a tratar este problema de modo a decidir as melhores rotas diante da possibilidade de utilizar frota terceirizada (relacionada à problemas abertos) e frota própria (relacionada à problemas fechados).



Os problemas fechados e abertos visam projetar rotas que atendam à demanda dos clientes minimizando o custo total da entrega, por meio de uma frota de veículos homogênea, localizada em um único depósito, para atender a um conjunto de clientes. A diferença entre eles é que no primeiro cada rota começa e termina exatamente em um depósito e no segundo não há necessidade desse retorno. O COMVRP é constituído por rotas abertas ou fechadas devido a utilização de frota própria ou contratada. A este problema são integradas características de múltiplos depósitos ou centros de distribuições, frota heterogênea e periodicidade.

A literatura apresenta diferentes abordagens, baseadas nas características, do VRP, buscando cada vez mais incorporar as complexidades da vida real. A seguir serão apresentados essas características adicionais ou restrições que são abordadas devido as especificidades do caso analisado.

O problema avaliado parte do CVRP. A esse problema, acrescenta-se a possibilidade de se utilizar frota própria e terceirizada. Essa característica é primordial em casos em que o sistema de transporte próprio possa não ser suficiente para atender às necessidades dos clientes, permitindo delegar uma parte das demandas dos clientes para os veículos terceirizados. Esta característica implica neste caso, em não precisar retornar ao depósito depois de visitar o último cliente, característica do OVRP, pois trata-se de veículo terceirizado. Fato este relacionado à prática de faturamento de contrato com as transportadoras. Neste trabalho, foram considerados o COMVRP, de modo a considerar frota própria e frota terceirizada simultaneamente.

O Problema de Roteirização de Veículos com Múltiplos depósitos (*Multiple Depot Vehicle Routing Problem* - MDVRP), também foi considerado, devido ao fato de existirem diversas lojas a serem avaliadas. Assim, cada veículo parte do depósito e após finalizada rota, dependendo do veículo próprio ou terceirizado, deve ou não retornar ao depósito. Abordou-se, também, uma dimensão de tempo ou periodicidade característica do Problema de Roteirização de Veículos Periódico (*Periodic Vehicle Routing Problem* - PVRP), pois o planejamento de rota deve ser executado em um horizonte de vários períodos. A atribuição de visitas a clientes está sujeita a restrições de compatibilidade com os padrões preestabelecidos, isto é, combinações aceitáveis de períodos de visita.

Por fim, outra característica abordada relaciona-se à heterogeneidade de veículos, relacionada à variante Problema de Roteamento de Veículos com Frota

Heterogênea (*Heterogeneous Fleet Vehicle Routing Problem - HFVRP*). Considerou-se, neste trabalho, que os tipos de veículos possuem atributos diferentes: capacidade, tempos máximos de rota, custos fixos e custos variáveis por unidades de distância. Adota-se um custo fixo para cada classe de veículos.

Quando o número de veículos não é restrito, tem-se o Problema de Roteamento de Veículos com Dimensionamento de Frota Heterogênea (*Fleet Size and Mix - FSM* ou *Vehicle Fleet Mix Problem - VFMP*). Neste caso, as instâncias da literatura consideram, que o número de veículos é igual ao número de clientes a serem atendidos. A variante do VRP com frota limitada é denominada Problema de Roteamento de Veículos com Frota Heterogênea Fixa (*Heterogeneous Vehicle Routing Problem - HVRP* ou *Heterogeneous Fixed Fleet Vehicle Routing Problem - HFFVRP*). Neste trabalho considerou-se tanto possibilidade de frota limitada e ilimitada para a frota própria e frota terceirizada, como veículos homogêneos ou heterogêneos em sua composição.

Este trabalho abordou a integração do problema de segmentação de mercado, partindo da observação, identificação e utilização das características dos clientes, e suas preferências de compras nas diversas lojas. Em seguida, foram compostas as rotas de entregas mais adequadas em termos de benefícios de compras e minimização de custos, de forma a responder positivamente a necessidade logística imposta pela abordagem comercial moderna.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Desenvolver uma metodologia que possibilite considerar estratégias de segmentação de mercado, por meio do processo KDD, associadas a uma variante do VRP com múltiplos atributos para o planejamento de rotas de entregas de produtos no segmento varejista.

### 1.2.2 Objetivos específicos

Os objetivos específicos a serem atingidos são os seguintes:

- a) Apresentar uma revisão sistemática que se utiliza da classificação de artigos do VRP segundo taxonomias propostas na literatura de forma a

destacar as suas principais características, apresentando para o problema com múltiplos atributos, as estratégias adotadas nas obtenções de suas soluções.

b) Apresentar uma metodologia que integra o (i) processo de descoberta de conhecimento de dados (KDD) para a segmentação de mercado com a (ii) variante do VRP abordada.

- Para a etapa (i), aplicar técnicas de agrupamento de dados a fim de se criar segmentos distintos de consumidores e lojas; e técnicas de predição para prever o gasto gerado pelos clientes.
- Para a etapa (ii):
  - i. Propor um modelo matemático para o problema de roteirização de veículos capaz de lidar com as restrições complexas características do mercado varejista considerando-se a abordagem clássica de minimizar custos de distribuição (estratégia de minimização).
  - ii. Apresentar uma variação deste modelo considerando a característica de depósitos heterogêneos, aliando-se as estratégias de segmentação de mercado por considerar as preferências dos clientes (estratégia de maximização de benefícios).
  - iii. Propor um Algoritmo Híbrido com Particionamento de Conjuntos (*Hybrid Genetic Algorithm with Set Partitioning - HGA-SP*) para resolução abordada.

c) Avaliar a metodologia proposta com dados reais de uma das maiores redes varejistas no segmento de hiper e supermercados da região sul do Brasil.

### 1.3 JUSTIFICATIVA

A primeira justificativa do trabalho baseia-se na sua importância prática do modelo matemático de roteirização desenvolvido. A aplicabilidade em problemas reais, em que o tomador de decisão passa a considerar lidar com transportadoras terceirizadas, é uma necessidade crescente na Gestão da Cadeia de Suprimentos (*Supply Chain Management – SCM*). Problemas que envolvem rotas abertas têm

aplicabilidade voltada às empresas de distribuição que não possuem frota de veículos própria e delegam esse processo a um contratado terceirizado (Transportadora ou Transportador Autônomo de Cargas – TAC) ou contratam alguns motoristas junto com seus próprios veículos (TAC Agregado).

Esse problema caracteriza o caso de muitos grupos empresariais que são compostos por vários depósitos ou lojas espalhadas geograficamente em uma região, e precisam atender às demandas dos clientes podendo contar com uma frota própria (neste caso, o veículo retorna ao depósito) ou terceirizada de veículos. Ou seja, na situação em que a demanda total de clientes é maior que a capacidade total de caminhões próprios, a empresa pode contratar veículos terceirizados para o atendimento dos clientes no prazo estipulado.

Em detrimento da sua importância pouca atenção tem sido dada à essa variante. O primeiro trabalho a tratar simultaneamente o VRP aberto e fechado é de Liu e Jiang (LIU; JIANG, 2012). No melhor do nosso conhecimento, esta é a primeira vez que o problema de roteamento de veículos com frota heterogênea, múltiplos depósitos, periódico, com rotas abertas e fechadas denominado HMDPCOMVRP é apresentado na literatura.

Outros quatro trabalhos relacionados à essa variante foram identificados. Um deles aborda o COMVRP com as janelas de tempo em um método VNS (*Variable Neighborhood Search*, ou Busca em Vizinhança Variável) (BRITO; EXPÓSITO; MORENO, 2016). Azadeh e Farrokhi-Asl (2019) apresentaram um COMVRP, com múltiplos depósitos e com frota heterogênea. Outro apresenta o COMVRP com múltiplos depósitos (*Multi-Depot VRP* ou MDVRP), assumindo que a frota de veículos seja heterogênea (TAVAKKOLI-MOGHADDAM et al., 2019). Este último trabalho, porém, considera uma função com dois objetivos e considera que cada veículo não precisa voltar ao depósito original. Por fim, Amalia; Suprajitno e Pratiwi (2020) apresentaram o Algoritmo de Morcego (*Bat Algorithm*) para solucionar o COMVRP. Diferentemente do trabalho de Azadeh e Farrokhi-Asl (2019), o presente trabalho acrescenta a possibilidade de planejamento das rotas ao longo de um horizonte de planejamento. Assim, permite-se a escolha ao cliente de um determinado dia para que ele seja atendido. Incluindo no modelo, a possibilidade de poder ser atendido mais de uma vez ao longo do horizonte de planejamento.

Ao longo das últimas décadas, os pesquisadores têm extensivamente explorado diferentes variantes do VRP com o propósito de reduzir custos logísticos – geralmente representados em função da distância percorrida pelos veículos. A minimização dos custos de distribuição tem um grande impacto competitivo na eficiência de qualquer companhia. No entanto, seguindo a tendência de se considerar problemas de roteamento de veículos mais realistas (CACERES-CRUZ *et al.*, 2015), deve-se observar que o lucro da empresa não se baseia apenas em minimização de custos. Mas também, na maximização de receitas com as vendas dos produtos.

Assim, inspirado pelo artigo de Calvet *et al.* (2016a), que combina aprendizagem estatística para segmentação de mercado com o problema de roteamento MDVRP, esta tese também considera a abordagem realista e inovadora de depósitos heterogêneos. Utiliza-se o adjetivo ‘inovador’ pois até então, pouca atenção foi dada a essa possibilidade de se considerar possíveis variações nas demandas decorrentes da designação de entregas para clientes provenientes de depósitos heterogêneos.

Nesta versão do problema, os depósitos são heterogêneos considerando sua composição na oferta comercial, sendo que os clientes mostram uma disposição diferente de consumir, dependendo de quão bem o depósito atribuído a ele se encaixa em suas preferências. Portanto, assume-se que diferentes atribuições de clientes levarão a níveis distintos de despesas do cliente. Como consequência, estratégias de segmentação de mercado precisam ser consideradas para aumentar as vendas e a receita total, enquanto contabilizam os custos de distribuição.

De fato, esta é uma característica peculiar de lojas ou centros de varejo pertencentes à um mesmo grupo. Seja pela oferta de diferentes produtos, políticas de crédito comercial e/ou serviços complementares que, geralmente, têm impacto não negligenciável na disposição do cliente para comprar. Consequentemente, sob existência de uma diversidade de lojas e ofertas comerciais, passa a ser imprescindível considerar não apenas os custos de distribuição, mas também vendas ou receita total esperadas mediante a atribuição do cliente para uma ou outra loja.

Tendo em vista a importância de se estimar os gastos de clientes, é prudente que seja feito o uso de uma metodologia baseada no processo KDD, de forma a considerar suas características e preferências. Tal metodologia apresentada neste

trabalho, por si só, pode contribuir com a maneira de gerar respostas para as principais questões de negócios varejistas, relacionadas a segmentação de mercado.

#### 1.4 CONTRIBUIÇÕES

Esta tese tem como objetivo contribuir para o avanço da integração de estudos que relacionem as estratégias de segmentação de mercado a tomadas de decisões logísticas, neste caso, relacionadas a roteamento de veículos para entregas de mercadorias. O desenvolvimento e as contribuições desta tese são assim organizados em três eixos complementares – teórico, metodológico e prático ou gerencial.

- a) A primeira contribuição, de cunho teórico, se dá na revisão sistemática da literatura do VRP, na qual são classificados artigos de acordo com taxonomias da literatura de modo a se identificar características de cenário e características físicas do problema, além dos tipos de estudo e características de dados, mais frequentemente abordadas. Além disso, são investigadas características dos algoritmos utilizados em VRPs com múltiplos atributos. Também são expostos os principais conceitos e pormenores relacionados às técnicas utilizadas para o processo KDD e suas suposições para serem aplicadas.
- b) Quanto a contribuição metodológica, tem-se a integração do KDD com o VRP, visto que é algo inovador e com grande potencial de aplicabilidade, em razão do aumento da comercialização em plataformas eletrônicas e conseqüentemente, do volume de informações disponíveis.
- c) Nesta perspectiva, a descrição do modelo HMHDPCOMVRP com todos os seus atributos tratados em conjunto, por si só, também consiste em uma contribuição, devido ao seu ineditismo.
- d) A pesquisa avança para uma versão estendida desse problema, com uma abordagem nova, considerando depósitos heterogêneos, assim como proposto por Calvet *et al.* (2016), a fim de voltar a atenção para maximizar o lucro estimado ao invés de apenas minimizar os custos com deslocamento. A proposta de análise dos resultados, em termos dos lucros esperados e dos custos de distribuição, é realizada a partir das

soluções encontradas para o problema estendido e para o problema clássico, contribuindo com a discussão da temática na literatura.

- e) Por fim, existe uma contribuição prática gerencial para auxílio a tomada de decisão dos varejistas durante o processo de segmentação de mercado e no planejamento de rotas. Além disso, a metodologia proposta foi baseada em um caso de *marketing* real baseado em um ambiente *Business-to-Consumer* (B2C) de uma rede varejista do segmento de super e hipermercados.

## 1.5 ORGANIZAÇÃO DO TRABALHO

O trabalho está organizado da seguinte forma: no Capítulo 2, os conceitos sobre o KDD e as técnicas utilizadas são apresentadas em detalhes. Também se expõe uma revisão sistemática contendo uma classificação de artigos de modo a sintetizar as principais características dos VRPs, e dos algoritmos para resolver problemas com múltiplos atributos. No Capítulo 3, apresenta-se a metodologia utilizada constituída pelo processo KDD, que visa considerar questões de segmentação de mercado associadas ao problema de planejamento de rotas. Destaca-se a descrição dos passos para a utilização das técnicas de agrupamento e de predição utilizadas. Seguida pela apresentação de como se realiza a integração destas técnicas com o VRP e suas abordagens de soluções, considerando todas as informações de dados necessárias. Descreve-se também o problema de roteamento de veículos a ser tratado, seus pressupostos e notações, sob uma perspectiva do problema clássico e de sua versão com depósitos heterogêneos. Bem como, descreve-se a abordagem de solução via Algoritmo Genético. No Capítulo 4 são apresentados e discutidos os resultados obtidos com as aplicações das técnicas do KDD associadas ao VRP para o planejamento de rotas. Apresentando-se, enfim, os resultados e discussões quanto a abordagem de solução exata e heurística. Por fim, no Capítulo 5 conclui-se o trabalho, apontando-se as principais contribuições e perspectivas de trabalhos futuros.

## 2 REVISÃO DA LITERATURA

Neste capítulo apresenta-se uma revisão de dois temas fundamentais para o desenvolvimento do presente trabalho: Segmentação de Mercado por meio de análise do processo KDD (*Knowledge Discovery in Databases*) e o VRP e suas variações.

### 2.1 SEGMENTAÇÃO DE MERCADO

O artigo seminal de Smith (1956) foi o que introduziu a teoria de segmentação de mercado na literatura de *marketing*. Segundo este autor, a estratégia de segmentação de mercado tem como pressuposto a existência de um mercado heterogêneo, caracterizado pela existência de demandas diferenciadas, que compõem distintos segmentos de mercado. Sua adoção implica a identificação de classes de consumidores que se constituem em possíveis segmentos-alvo da empresa e que norteiam os esforços de *marketing* no sentido de satisfazer a demanda dos segmentos escolhidos.

A segmentação de mercado visa à identificação dos consumidores de acordo com suas características e perfil de consumo (WEDEL; KAMAKURA, 2012). Sua adoção consiste na tarefa de agrupar e/ou classificar determinados consumidores conforme características de demografia, comportamentos, valores, entre outros (CARVALHO et al., 2016). A segmentação de mercado, desde então tem sido fundamental para a estratégia de *marketing* das organizações, sendo o seu sucesso vinculado à complexa tarefa de compreender os consumidores que possuem características diferentes entre si e que, portanto, podem reagir de maneiras diferentes a um determinado estímulo (SCHIFFMAN; KANUK, 1991). Para revisões sobre o tema ver (DÍAZ et al., 2018; HIZIROGLU, 2013; TKACZYNSKI; RUNDLE-THIELE, 2011).

Alguns critérios precisam ser levados em consideração na segmentação de mercado (KOTLER, 2000). Os segmentos, por meio de suas características, precisam ser: (i) identificáveis mediante algum método; (ii) de tamanho suficiente, ou seja, que possuam um número adequado de indivíduos, justificando, assim, economicamente uma ação de *marketing* específica; (iii) estáveis no seu comportamento por



determinado período de tempo, de modo que a ação de *marketing* empregada tenha o efeito planejado; e (iv) acessíveis à organização, em termos de comunicação e distribuição, a um custo viável.

Um processo de segmentação de mercado busca, então, agrupar os consumidores em uma ou mais dessas formas apresentadas. Pode ser a *priori*, em que os segmentos já são escolhidos previamente, ou *post hoc*, em que os segmentos são identificados após análise dos dados (BEANE; ENNIS, 1987). A segmentação por agrupamento envolve quatro etapas (JAIN; MURTY; FLYNN, 1999), a saber: (i) seleção de dados; (ii) definição de padrão de proximidade; (iii) escolha do método, técnica e/ou algoritmo de agrupamento e; (iv) interpretação e validação dos resultados.

Quanto aos critérios utilizados para classificar os métodos de segmentação, os mais comumente empregados são os critérios *post hoc versus a priori* e os descritivos *versus* preditivos (FOEDERMAYR; DIAMANTOPOULOS, 2008). Os métodos *a priori* são baseados em intuições e experiências anteriores para determinar o número de segmentos. Enquanto nos métodos *post hoc*, a análise de dados é o que leva aos segmentos. Nos métodos descritivos, o foco é explorar a relação entre as unidades de análise, não havendo distinção entre variáveis dependentes e independentes. Por outro lado, os métodos preditivos uma variável dependente é explicada por um conjunto de variáveis independentes e usam esse conjunto para segmentar (OLIVEIRA-BROCHADO; MARTINS, 2008).

Quanto às técnicas de segmentação comumente utilizadas, pode-se citar as estatísticas multivariadas ou técnicas de Mineração de Dados, como análise discriminante, regressão linear múltipla, regressão logística, análise de *clusters*, redes neurais artificiais, computação evolucionária, lógica *fuzzy*, meta-heurísticas, entre outras.

Exemplos de aplicações podem ser citados como a utilização de: combinação do algoritmo *fuzzy c-means* e algoritmo genético (ANSARI; RIASI, 2016); heurística gulosa (HUERTA-MUÑOZ; RÍOS-MERCADO; RUIZ, 2017); regressão logística e árvores de decisão (CAIGNY; COUSSEMENT; BOCK, 2018); *Self Organising Maps*, para estudar os vários fatores que podem ser usados para segmentar o mercado, *Nearest Neighbour* e *Support Vector Machines* para classificar quantitativamente o comportamento de compra com base em vários fatores (TIWARI et al., 2018);

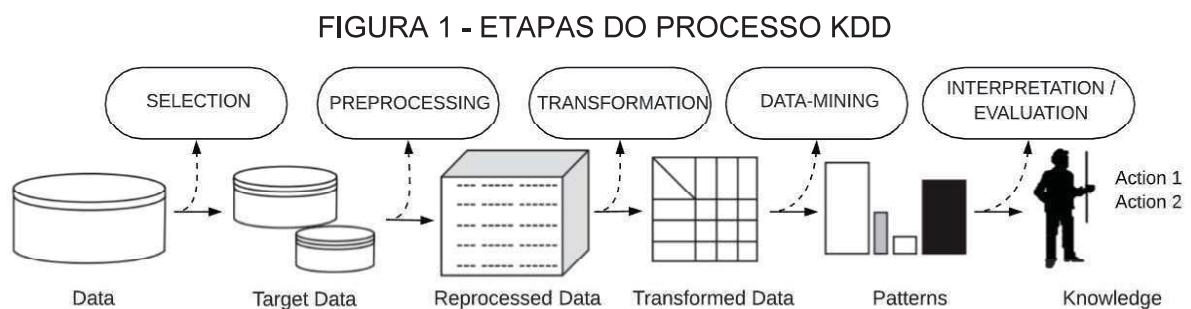
regressão linear múltipla para a influência das variáveis independentes na frequência de uso de quartos compartilhados e casas inteiras para o caso do Airbnb (LUTZ, C., & NEWLANDS; LUTZ; NEWLANDS, 2018) e sistemas não supervisionados e análise de agrupamento, abordando o problema de escolher a classificação mais apropriada de um determinado conjunto de classificações (SÁNCHEZ-HERNÁNDEZ et al., 2013).

Muitas são as referências internacionais citadas a fim de definir e caracterizar os pilares teóricos sobre o tema. Quanto às pesquisas nacionais, no melhor do conhecimento, o primeiro e único artigo bibliométrico específico sobre segmentação de mercado no Brasil foi realizado por Souza e Freitas (2016). Os autores fizeram um estudo bibliométrico sobre segmentação de mercado revisando 63 publicações dos principais periódicos nacionais e eventos acadêmicos tendo o *marketing* como área de concentração, publicados desde 1983. Suas principais conclusões foram que no Brasil, a segmentação de mercado não é usada como tema central das pesquisas, mas sim para analisar o perfil dos pesquisados, traçando-se poucas implicações gerenciais. Além disso, não há autores de referência no Brasil sobre o assunto, sendo um tema tratado por muitos pesquisadores, de forma descontínua.

Na próxima seção são apresentadas as principais características das técnicas utilizadas neste trabalho para a segmentação de mercado.

### 2.1.1 Processo KDD

O processo KDD apresentado por Fayyad (FAYYAD et al., 1996) é composto por diversas etapas, conforme mostrado na FIGURA 1 a seguir.



FONTE: Adaptado de FAYYAD et al. (1996)

As etapas de seleção, reprocessamento e transformação dos dados foram aqui obtidas por meio de técnicas de agrupamento utilizando-se da (i) Análise de Componentes Principais (ACP), que permite estudar a interrelação existente entre as variáveis, buscando a redução de dados por meio da combinação de variáveis para a formação de uma nova variável; (ii) Análise de Clusterização (AC), para o agrupamento das observações cujos valores das variáveis apresentarem maior similaridade.

Na etapa de mineração de dados utilizou-se técnicas de predição: (i) Regressão Linear Múltipla (RLM), e (ii) Redes Neurais Artificiais (RNA), comparativamente, para a criação de modelos de regressão que expliquem o impacto gerado por um grupo de variáveis sobre o comportamento de outra variável. Por fim, a avaliação e interpretação dos resultados foram utilizadas para a integração com o VRP proposto. Nas seções subsequentes, tem-se uma descrição mais detalhada das técnicas utilizadas nestas etapas.

#### 2.1.1.1 Técnicas de Agrupamento

Duas são as técnicas de agrupamentos apresentadas nesta seção: ACP e AC.

##### 2.1.1.1.1 Análise de Componentes Principais (ACP)

O objetivo principal da análise de componentes principais é o de explicar a estrutura da variância e covariância de um vetor aleatório, composto de  $p$ -variáveis aleatórias, por meio de combinações lineares das variáveis originais. Essas combinações lineares são chamadas de Componentes Principais (CP) e são não correlacionadas entre si. É importante que estas possuam o máximo de informação possível das variáveis originais, pois assim é possível garantir a qualidade das CPs (JOHNSON; WICHERN, 2002). Os princípios básicos desta técnica são descritos por vários autores, tais como: Johnson e Wichern (2002), Jolliffe (1986), Kendall (1975), Manly e Alberto (2016), Mardia, Kent e Bibby (1979) e Morrison (1976).

Para que se teste a adequação à utilização da ACP, avalia-se: a matriz de correlações, a estatística KMO, o teste de esfericidade e a matriz anti-imagem

(FÁVERO et al., 2009). A matriz de correlações mede a associação linear entre as variáveis, por meio do coeficiente de correlação de Pearson. De acordo com (HAIR et al., 2009) se a inspeção visual da matriz de correlações não revelar um número substancial de valores acima de 0.30, há fortes indícios de que a utilização da técnica não é apropriada.

O teste de *Bartlett Test of Sphericity* (BTS) serve para testar a hipótese nula de que a matriz de correlações pode ser a matriz identidade com determinante igual a 1. Isso significa que as inter-relações entre as variáveis são iguais a zero e neste caso deve-se reconsiderar a utilização da ACP (AAKER; KUMAR; DAY, 2001). Se a hipótese nula ( $H_0$ : a matriz de correlações é uma matriz identidade) não for rejeitada, isso significa que as variáveis não estão correlacionadas, e nesta situação, a técnica não é adequada. Por outro lado, se a hipótese nula for rejeitada, haverá indícios de que existem correlações significativas entre as variáveis originais.

O teste *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO) (KAISER, 1974) permite verificar a consistência dos dados originais, ao comparar as correlações simples com as correlações parciais, definido a partir da equação (1).

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2 = j}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad (1)$$

Onde:

$r_{ij}$ : coeficientes de correlação entre variáveis;

$a_{ij}$ : coeficientes de correlação parcial.

É desejável que o valor do índice KMO seja o mais próximo de 1. Isto quer dizer que o somatório dos coeficientes de correlação parcial entre as variáveis deve ser pequeno quando comparado ao somatório os coeficientes de correlação observados. Seus valores críticos são os seguintes (KAISER, 1974): valores acima de 0,90: adequação ótima dos dados; valores acima de 0,80 a 0,90: adequação boa dos dados; valores acima de 0,70 a 0,80: adequação razoável dos dados; valores acima de 0,60 a 0,70: adequação medíocre dos dados; valores abaixo 0,60: adequação imprópria dos dados.

Por fim, deve-se analisar se a matriz de correlações anti-imagem que contém os valores negativos das correlações parciais. Esta análise tem por objetivo avaliar a

necessidade de eliminação de alguma variável no modelo. A diagonal principal da matriz anti-imagem fornece o índice de Medida de Adequação de Amostra (*Measure of Sampling Adequacy* – MSA). Valores inferiores a 0,50 devem ser removidos da amostra e a análise fatorial deve ser realizada novamente (SCHAWB, 2015). Essa exclusão segundo os mesmos autores não é obrigatória visto que esta variável pode representar um fator isoladamente. Dessa forma, outro aspecto que deve ser observado diz respeito à avaliação das comunalidades, que representam a quantia de variância explicada pela solução fatorial de cada variável, onde comunalidades menores que 0,5 são consideradas como explicação insuficiente (HAIR et al., 2009).

Para decidir quantos componentes devem ser retidos, os critérios utilizados são: o critério da raiz latente (critério de Kaiser); critério *a priori*; critério de percentagem de variância; e critério do gráfico Scree. Pelo critério da raiz latente, escolhe-se a quantidade de componentes a reter, em função do número de autovalores (chamados de *Eigenvalues*, que representam a variância total explicada por cada fator) acima de 1. Já o critério *a priori* é o método mais simples, pois o pesquisador sabe quantos componentes extrair. O critério da porcentagem de variância consiste na escolha o percentual de variância explicada mínima. Por meio do método gráfico *Scree*, considera-se as componentes anteriores ao ponto de inflexão da curva, ou seja, o número ótimo de componentes que pode ser extraído antes que a quantia de variância única inicie o domínio sobre a estrutura de variância comum (CATTELL, 1966; PLA, 1986).

Com relação à rotação dos componentes principais, os métodos podem ser ortogonais e oblíquos. Os primeiros produzem componentes que não estão correlacionados entre si, cuja interpretação ocorre a partir de suas cargas (ou *loadings*). Já nos métodos oblíquos, estes estão correlacionados, tornando-se imprescindível a consideração simultânea das correlações e das cargas para a interpretação da solução. Destaca-se que o método Varimax, que visa minimizar o número de variáveis que apresentam altas cargas em um fator, facilitando a interpretação dos componentes, como o mais utilizado (FÁVERO et al., 2009). Após a rotação, não são alteradas a qualidade de ajuste do modelo fatorial, as comunalidades e o total de variância explicada pelos componentes, apenas o percentual de variância explicado em cada fator se altera.

A interpretação e nomeação dos componentes por meio das cargas fatoriais é a última etapa da técnica de análise fatorial, em que se decide quais cargas fatoriais devem ser consideradas. Conforme apontam Hair et al. (2005), as cargas fatoriais podem ser consideradas estatisticamente significantes, dependendo do número amostral. Com o objetivo de conseguir um nível de poder de 80%, o uso de um nível de significância de 0,05 a TABELA 1 contém os tamanhos de amostra necessários para que cada valor de carga fatorial seja considerado significativo.

TABELA 1 – SIGNIFICÂNCIA DAS CARGAS FATORIAIS BASEADO NO TAMANHO DA AMOSTRA

Carga fatorial	Tamanho Necessário da Amostra para significância*
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

FONTE: Adaptado de HAIR et al (2005). Nota: Significância se baseia em um nível de significância de 0,05, e um nível de poder de 80% e erros-padrões considerados como o dobro daqueles de coeficientes de correlação convencionais.

A significância prática na escolha das cargas fatoriais é utilizada geralmente para fazer inferências preliminares da matriz fatorial. Em síntese, considera-se que as cargas fatoriais acima de 0,3 atingem o nível mínimo, cargas fatoriais de 0,4 são mais importantes e maiores que 0,5 são consideradas estatisticamente significativas.

#### 2.1.1.1.2 Análise de Clusterização (AC)

A AC é uma técnica exploratória em que dada uma amostra de  $n$  indivíduos, cada um dos quais caracterizados por  $p$  variáveis, um critério deve ser criado para agrupar os indivíduos em *clusters*, de forma que os indivíduos que possuam características semelhantes estejam na mesma classe (HAIR et al., 2009).

As etapas para a AC segundo Malhotra (2001) são: a) formulação do problema; b) escolher uma medida de similaridade ou dissimilaridade; c) escolher o

algoritmo de Clusterização; d) decidir quanto ao número de *clusters*; e) interpretar e perfilar os agrupamentos; e f) avaliar o processo.

Dentre os métodos existentes para se mensurar a (dis)similaridade entre objetos, apresenta-se (POHLMANN, 2007):

- Medidas de distância: medida de separação entre dois pontos, utilizada para variáveis métricas.
- Medidas correlacionais: permitem trabalhar com variáveis categóricas.
- Medidas de associação: quando os indivíduos são agrupados com base nos coeficientes de correlação ou de outras medidas de associação.

Dentre as métricas de distâncias existentes, pode-se citar: distância euclidiana, distância city-block (Manhattan), distância de Mahalanobis (distância estatística) e Métrica de Minkowski. Dentre as quais, destaca-se a distância Euclidiana, a qual representa a distância geométrica no espaço multidimensional (POHLMANN, 2007). A distância entre os itens  $x$  e  $y$  é definida pela equação (2).

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (2)$$

Os algoritmos utilizados em AC podem ser classificados em duas categorias: hierárquicos e não hierárquicos. Os algoritmos não-hierárquicos, têm como objetivo encontrar diretamente uma partição de  $n$  indivíduos em  $g$  grupos, de modo que a partição atenda dois requisitos básicos: semelhança interna, também chamada de “coesão”, e isolamento dos clusters formados, também denominado como “separação” dos grupos (MINGOTI, 2007). É comum que os algoritmos computacionais utilizados nos métodos não-hierárquicos sejam do tipo iterativo e, comparando com os métodos hierárquicos, existe uma maior capacidade de análise de conjunto de dados com muitas observações. Exemplos desses métodos são o *k*-médias e o *Fuzzy c-Means* (JOHNSON; WICHERN, 2002).

Já os algoritmos hierárquicos criam uma estrutura encadeada entre os elementos usando uma medida de distância entre si do conjunto de dados. Essa estrutura forma uma representação gráfica chamada de dendrograma, onde os objetos semelhantes, segundo as variáveis estudadas, são agrupados entre si. Esses

métodos não exigem que já se tenha um número inicial de clusters. Eles podem ser classificados em dois tipos: Divisivos e Aglomerativos.

No método Divisivo todos os elementos começam juntos em um único cluster, e vão sendo separados um a um, até que cada elemento seja seu próprio cluster. No método Aglomerativo, cada ponto do espaço amostral é considerado um agrupamento e os dois pontos com mais alta similaridade são sucessivamente unidos até que um único cluster com todos os elementos seja formado. O número ideal de clusters é escolhido dentre todas as opções (KLEN, 2015).

O método Aglomerativo tem cinco algoritmos mais usados: ligação simples (considera a menor distância entre um grupo e outro), ligação completa (considera a maior distância entre um grupo e outro), ligação média (considera a distância média entre todos os indivíduos de um grupo para os do outro grupo), método de Ward (utiliza como distância a soma dos quadrados entre os dois agrupamentos, feita sobre todas as variáveis) e método centroide (considera a distância entre os centroides dos agrupamentos, onde os centroides são os centros (médias) dos agrupamentos) (JOHNSON; WICHERN, 2002).

#### 2.1.1.2 Técnicas de Predição

Duas são as técnicas de predição apresentadas nesta seção: RLM e RNAs.

##### 2.1.1.2.1 Regressão Linear Múltipla (RLM)

A finalidade da análise de regressão, segundo Hair et al., (2009), é estimar os valores de uma variável dependente com base nos valores conhecidos ou fixados das variáveis independentes (CUNHA; COELHO, 2009). Um modelo geral de regressão linear pode ser descrito segundo (FÁVERO et al., 2009) conforme a equação (3).

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \delta \quad (3)$$

Onde:

$Y$ : variável dependente métrica;

$\alpha$ : intercepto do eixo  $y$ ;

$\beta_k$ : coeficientes angulares de cada variável ( $k = 1, 2, \dots, n$ );

$X_k$ : são as variáveis explicativas (métricas ou *dummies*);



$\delta$ : termo de erro.

O parâmetro  $\alpha$  é calculado por meio do seu estimador  $a$ , conforme (4).

$$a = \bar{Y}_i - b_1\bar{X}_{1i} - \dots - b_n\bar{X}_{ni} \quad (4)$$

Onde:

$b_k$ : representa os estimadores dos betas para as  $n$  variáveis  $X$  utilizadas no modelo.

Já a estimativa dos betas segue-se o mesmo procedimento que para a regressão linear simples, mas é feito por etapa e por variáveis. A equação (5) apresenta a estimativa de  $\beta_1$ .

$$\beta_1 = \frac{cov(X_1, Y) \cdot Var(X_2) - cov(X_2, Y) \cdot cov(X_1, X_2)}{Var(X_1) \cdot Var(X_2) - [cov(X_1, X_2)]^2} \quad (5)$$

Enquanto uma regressão simples de duas variáveis resulta na equação de uma reta, um problema de três variáveis implica num plano, e um problema de  $k$  variáveis implica em um hiperplano. Apresenta-se a seguir alguns pressupostos básicos para uso da técnica RLM (CUNHA; COELHO, 2009):

- A variável  $Y$  é aleatória;
- A esperança matemática dos resíduos é nula, ou seja, a média dos resíduos é nula.
- A variância de  $\delta$  (termos de erro) é constante e igual a  $\sigma^2$  (condição de homoscedasticidade dos resíduos). A homoscedasticidade (variância constante dos resíduos) é uma propriedade essencial, que deve ser garantida, sob pena de invalidar toda a análise estatística.
- Os resíduos são independentes entre si. A verificação da autocorrelação pode ser feita pela análise do gráfico dos resíduos cotejados com os valores preditos, onde este deve apresentar pontos dispersos aleatoriamente, sem nenhum padrão definido ou pelo teste de Durbin-Watson.
- Os resíduos têm distribuição normal (distribuição de Gauss). A condição de normalidade dos resíduos é fundamental para a definição de intervalos de confiança e testes de significância. A verificação da normalidade pode ser feita pelos testes de aderência não paramétricos, como por exemplo, o de Kolmogorov-Sminorv.

Constatadas violações de uma ou mais suposições, algumas vezes é possível amenizar esse problema fazendo alguma transformação pertinente na variável resposta  $Y$ . As transformações normalmente usadas são as transformações: logarítmica na base  $e$  ( $\log(Y)$ ), raiz-quadrada ( $\sqrt{Y}$ ), recíproca ( $1/Y$ ), quadrática ( $Y^2$ ) e arcseno ( $\arcsen\sqrt{Y}$ ) (CHARNET et al., 1999). Box e Cox (1964) propuseram uma transformação que depende do parâmetro  $\lambda$ , conforme apresentado em (6).

$$Y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \log y, \lambda = 0 \end{cases} \quad (6)$$

Para se medir o quanto a variabilidade total dos dados é explicada pelo modelo de regressão, compara-se a Soma de Quadrados da Regressão com a Soma de Quadrados Total e tem-se o coeficiente de determinação ou de correlação múltipla ao quadrado  $R^2$  conforme apresenta-se em (7) segundo Stock e Watson (2004).

$$R^2 = \left( \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2}} \right)^2, 0 < R < 1 \quad (7)$$

Quando há o intuito de comparar o coeficiente de ajuste ( $R^2$ ) entre dois modelos ou entre um mesmo modelo com tamanhos de amostra diferentes, faz-se necessário o  $R^2$  ajustado, cuja expressão está descrita em (8).

$$R^2_{ajustado} = 1 - \frac{n-1}{n-k} (1 - R^2) \quad (8)$$

Onde:

$n$ : é o tamanho da amostra;

$k$ : número de parâmetros do modelo de regressão (número de variáveis explicativas mais o intercepto);

Algumas inferências adicionais podem ser feitas a partir do modelo de regressão múltipla. Uma das mais comuns é o teste F, que auxilia na avaliação da significância do modelo. Este teste baseia-se na comparação da soma dos quadrados dos erros do modelo de regressão múltipla original, com a soma dos quadrados dos erros de um modelo de regressão em que se supõe verdadeira a hipótese nula (FÁVERO, 2009). Assim, verifica-se a existência do modelo, uma vez que se os betas forem estatisticamente iguais a zero, o comportamento de alteração de cada uma das

variáveis explicativas não influenciará em absolutamente nada o comportamento da variável dependente. A estatística F pode ser escrita utilizando o  $R^2$  conforme em (9).

$$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{(n-k)}} \quad (9)$$

Para verificar quais variáveis explicativas consideradas são estatisticamente significantes para influenciar o comportamento da variável  $y$  utiliza-se a estatística  $t$  descrita por (10).

$$t = \frac{\hat{\beta}_i - \beta_i}{S(\hat{\beta}_i)} \quad (10)$$

Onde

S: significa o erro padrão de cada parâmetro em análise.

Especificando-se um nível de significância (p-valor), e por meio da estatística  $t$ , um conjunto de variáveis explicativas para o modelo de regressão é definido, a partir do método *Stepwise* (GUJARATI, 2006). A regressão *Stepwise* é uma abordagem sequencial em que a equação é estimada com um conjunto de variáveis independentes, que são seletivamente adicionadas (HAIR et al., 2009). Todas as variáveis independentes inicialmente são examinadas para inclusão na análise de regressão, e então, as variáveis independentes são acrescentadas uma de cada vez. Esse procedimento pode ser realizado por meio da seleção *forward*, ou eliminação *backward*. A seleção *forward* considera, inicialmente, um modelo simples usando a variável de maior coeficiente de correlação com a variável dependente, em seguida, uma variável por vez é incorporada até que não haja mais inclusões, e as variáveis selecionadas definem o modelo. A eliminação *backward* começa com o modelo contendo todas as variáveis independentes potenciais e elas são retiradas passo a passo em função da análise de significância.

#### 2.1.1.2.2 Redes Neurais Artificiais (RNA)

RNA são sistemas que possuem semelhanças com o cérebro humano, como a aprendizagem e a generalização de informações, devido ao processamento paralelo e a massiva conectividade dos “neurônios” – elementos de processamento – dispostos

em uma ou mais camadas (CARVALHO; BRAGA; LUDEMIR, 1998). Os problemas tratáveis por meio desta técnica podem ser divididos em três tipos: aproximação de função, classificador de padrões e agrupamento de dados (HAYKIN, 2007).

Uma grande vantagem das RNA é que são auto adaptáveis e não há necessidade de muitas suposições iniciais ao modelo, além de poderem ser utilizadas em séries não-lineares (KHASHEI; BIJARI, 2010). As RNAs têm a capacidade de aprendizagem, que ocorre nos processos iterativos dos ajustes. Os principais parâmetros que definem a arquitetura das RNA são: número de camadas, número de neurônios em cada camada, tipo de conexão entre os neurônios e topologia da rede (CARVALHO; BRAGA; LUDEMIR, 1998).

Os neurônios computam determinadas funções matemáticas, geralmente não-lineares, a partir de entradas recebidas. Estas unidades são dispostas paralelamente constituindo camadas e são interligadas com as unidades de camadas vizinhas por conexões geralmente associadas a pesos. Por analogia às interligações dos neurônios no cérebro humano, as conexões entre unidades de uma rede neural são chamadas sinapses, sendo os pesos denominados pesos sinápticos. Os pesos sinápticos ponderam as entradas recebidas por cada neurônio e são os responsáveis pelo armazenamento do conhecimento representado pela RNA. A solução de um problema com RNAs passa inicialmente por uma fase de aprendizagem, onde os pesos das conexões são ajustados até proporcionarem à rede a capacidade de representar o problema. Uma vez treinada, os pesos são fixados e a rede pode ser empregada como um modelo, estimando saídas a partir de um conjunto de dados de entrada.

O modelo aplicado no presente trabalho é o MLP (*Multy Layer Perceptron*). Algumas características básicas das redes MLP são as seguintes:

- A aprendizagem é supervisionada por correção de erro.
- A arquitetura possui mais de uma camada escondida, é acíclica (quando a saída de um neurônio não pode servir de entrada para algum neurônio anterior) e conectada (cada entrada é processada por todos os neurônios).
- A função de ativação de cada um dos neurônios da rede é uma função não-linear e precisa ser suave e diferenciável em todos os pontos, a

mais comumente utilizada é função não-linear sigmoidal, como a Logsig e Tansig.

- Uma rede MLP é uma rede *feedforward*, onde o fluxo de dados é sempre em uma única direção e ocorre pelo método supervisionado. Assim, a saídas dos neurônios de uma camada qualquer se conectam unicamente às entradas dos neurônios da camada seguinte. Portanto, o sinal de entrada se propaga por meio da rede em um só sentido, isto é, não existe realimentação.

Este tipo de rede tem sido aplicada com sucesso em diversas áreas, na solução de problemas complexos, desempenhando tarefas de classificação de padrões, reconhecimento de voz, reconhecimento de imagem, e controle. Uma vez que as redes perceptron só representam funções linearmente separáveis, um dos motivos do ressurgimento da área de redes neurais foi devido ao desenvolvimento do algoritmo *BackPropagation*.

Nas redes MLP, a função de ativação (não linear), aplicada a cada neurônio, utiliza como argumento o produto escalar do vetor de entrada ( $x$ ) e do vetor de peso ( $w$ ) associado a esse nodo, possui pelo menos uma camada oculta de processamento e alto grau de conectividade entre os neurônios, a qual é definida pelos pesos sinápticos. Utilizam a regra de aprendizagem por correção de erro do paradigma supervisionado. Sua arquitetura tem pelo menos duas camadas que são acíclicas e totalmente conectadas. A regra de propagação é dada pelo produto interno das entradas ponderadas pelos pesos com adição do termo bias, mas deve considerar que a saída da camada anterior é a entrada da camada atual (HAYKIN, 2007). A regra de propagação é descrita pelas equações (11) e (12).

$$u_j = \sum_{i=0}^n x_i w_{ij} + b_j \quad (11)$$

$$y_j = \varphi(u_j) \quad (12)$$

sendo:

$u_j$ : saídas do combinador linear do neurônio;

$x_i$  : entradas;

$w_{ij}$ : peso sinápticos para o erro do neurônio com a entrada.

$y_j$  : saída do neurônio;

$\varphi(u_j)$ : função de ativação do neurônio;

A função de ativação define a saída de um neurônio de acordo com o nível de atividade produzida pelas suas entradas. Funções de ativação comumente empregadas, de acordo Carvalho, Braga e Ludemir (1998) e Haykin (2007), estão descritas nas equações (13) a (16).

a) Função Limiar ou Degrau:

$$\varphi(u_j) = \begin{cases} +\gamma & \text{se } u_j \geq 0 \\ -\gamma & \text{se } u_j < 0 \end{cases} \quad (13)$$

b) Função Linear:

$$\varphi(u_j) = \alpha u_j \quad (14)$$

c) Funções Sigmóides:

1. Função Logística:

$$\varphi(u_j) = \frac{1}{1 + \exp(-\alpha u_j)} \quad (15)$$

2. Função Tangente Hiperbólica:

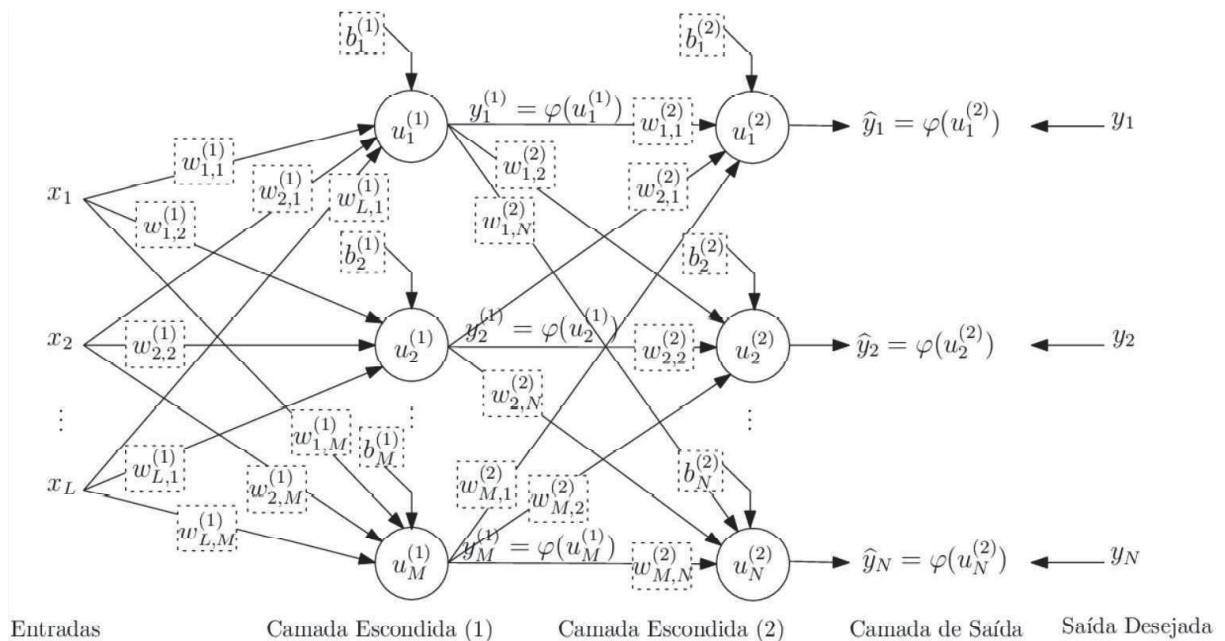
$$\varphi(u_j) = \operatorname{tgh}\left(\frac{u_j}{2}\right) = \frac{1 - \exp(-u_j)}{1 + \exp(-u_j)} \quad (16)$$

A função degrau é geralmente empregada com  $\gamma = 1$ , e pode ser redefinida para que as saídas situam-se dentro do intervalo de 0 a 1 (neurônio de McCulloch e Pitts). A função linear definida com  $\alpha = 1$  permite que a saída do neurônio seja igual ao nível de atividade interna. As funções logística e tangente hiperbólica são dois tipos de funções sigmoidais, produzindo como saídas valores entre 0 e 1, e entre -1 e 1, respectivamente. Por serem contínuas, crescentes, diferenciáveis e não-lineares, as funções sigmoidais são as mais comumente utilizadas nas camadas ocultas de RNAs tipo MLP.

Um exemplo da arquitetura de uma rede neural MLP, é mostrada na FIGURA 2. Ela tem  $L$  entradas, 2 camadas e diferentes quantidades de neurônios em cada camada, sendo, por exemplo,  $m$  neurônios na primeira camada e  $n$  neurônios na segunda camada. A quantidade de neurônios da camada de saída depende da quantidade de saídas da solução desejada. As camadas ocultas e a quantidade de

neurônios contidos em cada uma delas é um problema empírico, resolvido, em geral, por exaustivos testes. Para atualização dos pesos das camadas intermediárias utiliza-se o algoritmo de aprendizagem de retropropagação de erro (*backpropagation*). Basicamente, este algoritmo consiste em dois passos de computação: o processamento direto e o processamento reverso. No processamento direto, uma entrada é aplicada à rede neural e seu efeito é propagado pela rede, camada a camada. Durante o processamento direto, os pesos da rede permanecem fixos. No processamento reverso, um sinal de erro calculado na saída da rede é propagado no sentido reverso, camada a camada, e ao final deste processo os pesos são ajustados de acordo com uma regra de correção de erro (IYODA, 2000).

FIGURA 2 - REDE NEURAL MULTY LAYER PERCEPTRON



FONTE: Adaptado de Haykin (2007).

O cálculo para a variação dos pesos é descrito por (17).

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} + \Delta w_{ij}^{(n)} \quad (17)$$

Sendo

$w_{ij}$ : peso do neurônio  $j$  para a entrada  $i$ ;

$N$ : número de passos de ajustes dos pesos.

$\Delta w_{ij}$ : variação do peso para o neurônio  $j$  com a entrada  $i$ .

A Equação (18) apresenta a regra de atualização dos pesos.

$$\Delta w_{ij}^{(n)} = \eta \delta_j^{(n)} x_i^{(n)} \quad (18)$$

Sendo:

$\Delta w_{ij}$ : variação do peso para o neurônio  $j$  com entrada  $i$ ;

$\eta$ : taxa de aprendizagem;

$x_i$ : entrada  $i$ ;

$\delta_j$ : erro do neurônio  $j$ .

Os erros da camada de saída precisam ser propagados para as camadas escondidas, assim, os erros são calculados multiplicando pela derivada da função de ativação a diferença entre a saída esperada e a saída da rede. Os cálculos dos erros da camada de saída são descritos por (19).

$$\delta_j = (y_j - \hat{y}_j) \varphi'(u_j) \quad (19)$$

Sendo que:

$y_j$ : valor de saída esperado do neurônio  $j$ ;

$\hat{y}_j$ : saída do neurônio  $j$ ;

$\varphi'(u_j)$ : derivada da função de ativação de  $j$ .

Para o uso de RNA é necessário o uso de pares de entrada-saída já conhecidos. Em seguida, esses dados devem ser separados em dois conjuntos bem definidos, o conjunto de treinamento e o conjunto de teste, a fim de medir a performance da rede, considerando como ela responde a pares entrada-saída não apresentados durante o treinamento (o conjunto de teste).

Segundo Carvalho (CARVALHO; BRAGA; LUDEMIR, 1998), podemos definir os passos da seguinte maneira:

1. Inicializar a rede, os pesos ( $|w| < 0,1$ ), e parâmetros;
2. Fornecer dados de entrada à rede e calcular o valor da função de erro obtida, ao comparar com o valor de saída esperado.
3. Calcular as saídas da rede iniciando pela primeira camada escondida até a camada de saída (fase *forward*);
4. Calcular o erro de cada neurônio da camada de saída.
5. Atualizar pesos sinápticos.
6. Repete-se os passos 4 e 5 iniciando pela camada anterior a camada de saída até a primeira camada escondida (fase *backward*);



7. Se ainda tem padrões a aplicar, voltar ao passo 2;

8. Se o erro de algum dos padrões é maior que algum limiar preestabelecido, ou o número de iterações atingir um valor máximo, voltar ao passo 2 com primeiro padrão.

Não há na literatura métodos que pré-determinem a melhor arquitetura da rede. A combinação de camadas e neurônios em uma MLP deve ser escolhida por meio de testes de modo que garanta a melhor generalização da rede, que consiste na capacidade de um modelo de aprendizado de responder corretamente a padrões novos que lhe são apresentados. Segundo Tymvios et al. (2005) existem na bibliografia algumas relações empíricas para resolver este problema, mas o melhor método até o momento é a construção de vários modelos, selecionando aquele que atinge um treinamento satisfatório e melhor se ajusta para a aplicação desejada.

É importante mencionar que um treinamento prolongado demais gera um problema conhecido como *overtraining*, que pode levar a uma superespecialização da rede (principalmente quando se dispõe de poucos dados) e a uma perda da capacidade da rede de responder bem a dados nunca apresentados (perda da capacidade de generalização). É aceito que a melhor rede é aquela que fornece o menor erro no conjunto de para o conjunto de teste.

Da mesma forma, o incremento excessivo do número de camadas escondidas e neurônios dentro da RNA produzem uma rede muito grande que demanda um maior esforço computacional, e não responderá corretamente a padrões nunca vistos. Isto é, devido à rede extrair muitas características, ela pode tentar, de forma inapropriada, modelar também o ruído. E um número muito pequeno de camadas escondidas e neurônios produz uma RNA muito pequena, que não é capaz de armazenar todos os padrões necessários, a ponto de não modelar fielmente os dados. Então, se deve ter um compromisso entre convergência, a capacidade da rede de aprender todos os padrões do conjunto de treinamento e a generalização.

Para mensurar a generalização da rede utiliza-se medidas de erros. Este erro é calculado de duas formas distintas neste presente trabalho:

- d) Erro Médio Percentual (*Mean Absolute Error* - MAE): O MAE mede a magnitude média dos erros em um conjunto de previsões.

$$MAE = \left(\frac{1}{N}\right) \sum_{i=1}^N |y_i - \hat{y}_i| \quad (20)$$

e) Raiz do Erro Médio Quadrático (*Root Mean Square Error – RMSE*): RMSE é uma regra de pontuação quadrática que também mede a magnitude média do erro. É a raiz quadrada da média das diferenças quadráticas entre previsão e observação real.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (21)$$

Onde:

$y_i$  é o valor real da observação  $i$ ;

$\hat{y}_i$  é a previsão para a observação  $i$  e

$N$  é o número de observações no conjunto de validação.

## 2.2 CONCEITOS SOBRE O VRP

Esta seção descreve uma revisão sistemática da literatura realizada com o intuito de identificar as principais características abordadas pela literatura sobre o VRP. Foram classificados 229 artigos publicados entre 2015 e 2018. Os artigos são classificados de acordo com (i) Características de Cenários, (ii) Características Físicas de Problemas, (iii) Tipos de Estudo e das (iv) Características dos Dados utilizados. As principais questões abordadas nesta seção são:

Q1: Quais são as características de cenário e características físicas do problema mais frequentemente abordada em pesquisas sobre o VRP nos últimos três anos?

Q2: Que tipo de estudo e características de dados abordam?

Q3: Quais são os trabalhos mais citados que constituem os pilares teóricos do tema?

Os modelos recentes de VRP tendem, cada vez mais, a incorporar as complexidades da vida real. Os atributos do problema consistem nas características adicionais ou restrições que visam levar em conta as especificidades de aplicações.

Portanto, define-se que a nomenclatura utilizada para tratar de múltiplas características do VRP é o *Multi-attribute* VRP (MAVRP). As dimensões da maioria das instâncias de interesse bem como as decisões que combinam múltiplos atributos simultaneamente, (os chamados VRPs ricos – *rich* VRP, vide (LAHYANI; KHEMAKHEM; SEMET, 2015)), dificultam a aplicabilidade de métodos exatos, por isso a importância de soluções heurísticas. Mesmo uma pequena melhoria nas soluções pode ter enormes consequências econômicas e ambientais. Assim, a seguinte questão é abordada:

Q4: Quais são as estratégias consideradas “mais promissoras” utilizadas por algoritmos recentes selecionados por seu alto desempenho em MAVRP?

Considera-se este estudo um passo importante para o desenvolvimento de abordagens e/ou métodos mais complexos e robustos para resolver o máximo de variantes de roteamento de veículos possível, com essa a mesma abordagem/método proposto.

### 2.2.1 Revisão Sistemática do VRP

O aumento exponencial da velocidade de processamento e da capacidade de memória dos computadores também contribuem para o rápido crescimento do número de métodos de solução propostos para o VRP. A capacidade de resolver instâncias maiores estimula o progresso no domínio da investigação e do desenvolvimento de *softwares* comerciais para a VRP, utilizado por empresas, em uma grande variedade de setores industriais (PARTYKA J.; HALL, 2014; TOTH; VIGO, 2014).

Grande parte das pesquisas/revisões de literatura se concentram em variáveis específicas ou aspectos do VRP. Alguns exemplos disso são: o VRP heterogêneo (KOÇ et al., 2016), VRPs periódico (CAMPBELL; WILSON, 2014; CAMPBELL; O’KELLY, 2012), o VRP com janelas de tempo (GENDREAU; TARANTILIS, 2010; PSARAFTIS; WEN; KONTOVAS, 2016), VRPs dinâmico (PILLAC et al., 2013), VRPs estocástico (GENDREAU; JABALI; REI, 2014), VRP dinâmico e estocástico (RITZINGER; PUCHINGER; HARTL, 2016), veículo de roteamento com múltiplos depósitos (MONTROYA-TORRES et al., 2015), roteamento de veículos com entregas divididas (ARCHETTI; SPERANZA, 2012), roteamento de veículos *green* (LIN et al., 2014), VRP com limitações de carga (POLLARIS et al., 2015), VRP com múltiplas

viagens (CATTARUZZA; ABSI; FEILLET, 2016) e VRP para logística da cidade (*city logistics*) (CATTARUZZA et al., 2017).

Esforços anteriores de classificação e revisão da literatura sobre o tema VRP foram publicados por Lahyani, Khemakhem e Semet (2015), cobrindo publicações entre 2006 e 2013 e em Braekers, Ramaekers e Van Nieuwenhuyse (2016), avaliando artigos publicados entre 2009 e meados de 2015. As características do VRP avaliadas nesta revisão são baseadas nestas duas taxonomias.

A primeira, proposta por Braekers, Ramaekers e Van Nieuwenhuyse (2016), não inclui quaisquer problemas combinados, tais como problemas de roteamento e estoque (*inventory routing problems* - IRP), problemas de roteamento e localização (*location-routing problems* - LRP) ou problemas combinando decisões de programação (*scheduling*) relacionados com outras atividades, tais como a programação de máquinas ou planejamento de produção. A segunda, proposta por (LAHYANI; KHEMAKHEM; SEMET, 2015), trata extensivamente destes tipos de problemas, estudando diversas variantes do VRP com múltiplas restrições, os *multi-constrained VRP*, uma classe de problemas conhecidos como *Rich Vehicle Routing Problems* (RVRPs).

A taxonomia criada por Lahyani, Khemakhem e Semet (2015) foi projetada de acordo com conceitos centrais do roteamento relacionados a características que alteram significativamente a natureza do problema. Esta taxonomia visa construir um quadro para classificar qualquer estudo RVRP, apresentando as subclasses específicas. Os problemas são considerados de acordo com as Características do Cenário (*Scenario Characteristics* - SC) e as características físicas do problema (*Physical Problems Characteristics* - PPC). Apresenta-se, a seguir, os atributos mais importantes para as duas classes (QUADRO 1).

A taxonomia de Braekers, Ramaekers e Van Nieuwenhuyse (2016) é uma versão adaptada da taxonomia proposta por com Eksioglu, Vural e Reisman (2009). Estes autores identificaram 1021 artigos com VRP como tema principal, publicados entre 1959 e 2008. Neste trabalho, define-se cinco características principais: 1) Tipo de estudo; 2) SCs; 3) PPCs; 4) Características de Informação; e 5) Características dos Dados, cada uma delas com suas subcategorias detalhadas. A partir destas características, foram analisados ainda: os tipos de estudos e as características dos dados, cujos elementos de análise são apresentados no QUADRO 2.

QUADRO 1 - TAXONOMIA DOS RVRP PROPOSTA POR (LAHYANI; KHEMAKHEM; SEMET, 2015)

<b>1. Características do cenário</b>	<b>2. Problema características físicas</b>
1.1 Dados de entrada	2.1 Veículos
1.1.1 Estático	2.1.1 Tipo
1.1.2 Dinâmico	2.1.1.1 Homogêneo
1.1.3 Determinístico	2.1.1.2 Heterogêneo
1.1.4 Estocástico	2.1.2 Número
1.2 Componentes de gerenciamento de decisão	2.1.2.1 Fixo
1.2.1 Roteamento	2.1.2.2 Ilimitado
1.2.2 Estoque e roteamento	2.1.3 Estrutura
1.2.3 Localização e roteamento	2.1.3.1 Compartimentalizado
1.2.4 Roteamento e sequenciamento de motoristas	2.1.3.2 Não Compartimentalizado
1.2.5 Planejamento de produção e distribuição	2.1.4 Restrições de capacidade
1.3 Número de depósitos	2.1.5 Política de carregamento
1.3.1 Único	2.1.5.1 Ordem cronológica
1.3.2 Múltiplos	2.1.5.2 Nenhuma política
1.4 Tipo de operação	2.1.6 Regulamentos dos motoristas
1.4.1 Pickup or delivery	2.2 Restrições de tempo
1.4.2 Pickup and delivery	2.2.1 Restrição no cliente
1.4.3 Backhauls	2.2.2 Restrição no acesso rodoviário
1.4.4 Dial-a-ride	2.2.3 Restrição no depósito
1.5 Restrições de divisão da carga	2.2.4 Tempo de serviço
1.5.1 Separação permitida	2.2.5 Tempo de espera
1.5.2 Divisão não permitida	2.3 Estrutura da janela de tempo
1.6 Período de Planejamento	2.3.1 Janela de horário individual
1.6.1 Período único	2.3.2 Várias janelas de tempo
1.6.2 Múltiplos períodos	2.4 Restrições específicas
1.7 Uso múltiplo de veículos	2.5 Função objetivo
1.7.1 Single trip	2.6.1 Objetivo único
1.7.2 Multi-trip	2.6.2 Objetivos múltiplos

FONTE: Adaptado de Lahyani; Khemakhem e Semet (2015)

QUADRO 2 - TAXONOMIA DO VRP PROPOSTA POR (BRAEKERS; RAMAEKERS; VAN NIEUWENHUYSE, 2016)

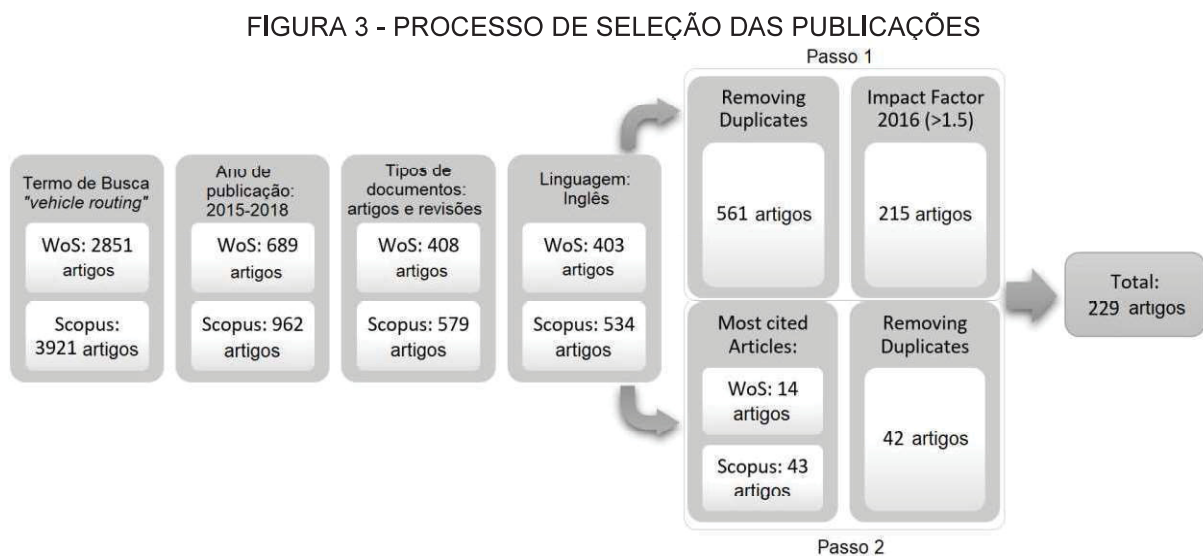
<b>3. Tipo de estudo</b>	<b>4. Características dos dados</b>
3.1 Teórico	4.1 Dados utilizados
3.2 Métodos aplicados	4.1.1 Dados em tempo real
3.2.1 Exato	4.1.2 Dados sintéticos
3.2.2 Heurística clássica	4.1.3 Dados reais e sintéticos
3.2.3 Simulação	4.2 Não foram utilizados dados
3.2.4 Métodos de solução em tempo real	
3.3 Documento de Implementação	
3.4 Pesquisa, revisão ou meta-pesquisa	

FONTE: Adaptado de Braekers, Ramaekers e Van Nieuwenhuyse (2016)

### 2.2.1.1 Procedimentos para revisão sistemática

Para a revisão sistemática realizada as seguintes fases foram definidas: estratégia de busca e seleção; e avaliação dos artigos. Utilizou-se para a execução da pesquisa os bancos de dados *ISI Web of Science* e *Scopus* usando o termo "*vehicle routing*" em "Título". Considerou-se apenas artigos publicados em revistas relevantes de língua inglesa, excluindo-se livros, artigos de congressos e dissertações.

Limitou-se a revisão da literatura a 2 passos. Primeiro, a pesquisa foi limitada a artigos publicados em revistas com um Fator de Impacto de pelo menos 1,5 nos domínios *Management Science* e *Operations Research or Transportation* (com base no Fator de Impacto de 2016 pela Thomson Reuters). O segundo passo foi ampliar esse conjunto com artigos altamente citados publicados em qualquer periódico classificado. Selecionou-se os artigos citados pelo menos cinco vezes por ano, em média, a partir do ano de sua publicação. Então, os seus resumos foram lidos para verificar a relevância do assunto. Esta estratégia de busca resultou em um conjunto de 229 artigos sobre o VRP (FIGURA 3).



FONTE: A autora (2020).

Uma visão geral de cada revista é mostrada na TABELA 2. A maioria (46,3%) dos artigos pesquisados foi publicada em: *European Journal of Operational Research*, *Computers & Operations Research* e *Expert Systems with Applications*.

TABELA 2 – DISTRIBUIÇÃO DE PUBLICAÇÕES EM JORNAIS SELECIONADOS

Título da Revista	Fator de Impacto	Número de Artigos Seleccionados	
		Passo 1	Passo 2
<i>IEEE Transactions on Vehicular Technology</i>	4.066	1	
<i>Omega – The International Journal of Management Science</i>	4.029	2	
<i>Expert Systems with Applications</i>	3.928	14	
<i>IEEE Systems Journal</i>	3.882	2	
<i>Transportation Research Part C: Emerging Technologies</i>	3.805	8	
<i>Transportation Research Part B: Methodological</i>	3.769	12	
<i>IEEE Transactions on Intelligent Transportation Systems</i>	3.724	4	
<i>International Journal of Production Economics</i>	3.493	4	
<i>Transport Reviews</i>	3.329	1	
<i>European Journal of Operational Research</i>	3.297	46	
<i>Transportation Science</i>	3.275	19	
<i>Transportation Research Part E-Logistics and Transportation Review</i>	2.974	12	
<i>Journal of Manufacturing Systems</i>	2.770	1	
<i>Networks &amp; Spatial Economics</i>	2.662	1	
<i>Transportation Research Part A-Policy and Practice</i>	2.609	1	
<i>Computers &amp; Operations Research</i>	2.600	46	
<i>Transportation Research Part D-Transport and Environment</i>	2.341	2	
<i>International Journal of Production Research</i>	2.325	6	
<i>Transportmetrica A-Transport Science</i>	2.033	1	
<i>Information Sciences</i>	1.910	3	
<i>Journal of Advanced Transportation</i>	1.813	2	
<i>Operations Research</i>	1.779	1	
<i>Journal of Intelligent Transportation Systems</i>	1.769	1	
<i>International Transactions in Operational Research</i>	1.745	4	
<i>Engineering Optimization</i>	1.728	2	
<i>Annals of Operations Research</i>	1.709	6	
<i>4OR-A Quarterly Journal of Operations Research</i>	1.559	1	
<i>OR Spectrum</i>	1.557	5	
<i>Computers and Industrial Engineering</i>	1.542	7	
<i>Computational Optimization and Applications</i>	1.520	2	
<i>Swarm and Evolutionary Computation</i>	1.411	1	
<i>Waste Management</i>	1.354		1
<i>Applied Soft Computing Journal</i>	1.308		2
<i>Networks</i>	1.213		1
<i>Applied Mathematical Modelling</i>	1.145		1
<i>Neurocomputing</i>	0.968		1
<i>Soft Computing</i>	0.750		1
<i>International Journal of Artificial Intelligence</i>	0.674		1
<i>International Journal of Bio-Inspired Computation</i>	0.622		1
<i>Journal of Sensors</i>	0.276		1
<i>Electronic Notes in Discrete Mathematics</i>	0.269		1
<i>Promet - Traffic – Traffico</i>	0.220		1

FONTE: A autora (2020).

O procedimento descrito permitiu que se chegasse às características e variantes do VRP mais abordadas na literatura. Apresentamos os resultados em

quatro partes: tipos de estudo e característica de dados, característica do cenário, características físicas do problema e base teórica.

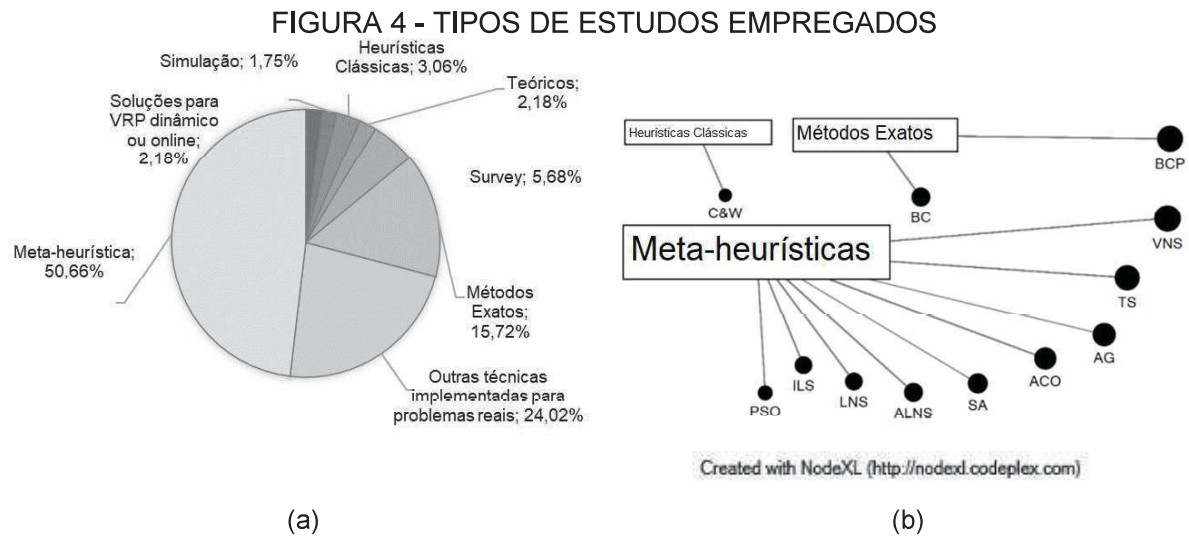
#### 2.2.1.2 Tipos de estudos e características dos dados

A respeito dos tipos de estudos de um artigo, apenas um número limitado de publicações tem um foco central em artigos teóricos (2.89%) e *surveys* ou meta-análise (5.37%). Os trabalhos de Eksioglu, Vural e Reisman (2009) e Lahyani, Khemakhem e Semet (2015) leva a observações semelhantes. Um importante artigo desta classe é o de Laporte (2016), que descreve várias classes de problemas de roteamento com *scheduling* e o de Han e Ponce-Cueto (2015) que tratam sobre o Roteamento de Veículos na Coleta de Resíduos *Waste Collection Vehicle Routing Problem* (WCVRP) na literatura.

Os artigos com métodos de solução em tempo real que tratam de VRPs dinâmicos ou *online* e artigos de simulação constituem uma porcentagem menor do número total de artigos analisados. Devido ao aumento da complexidade nas variações mais recentes do VRPs, o desenvolvimento de meta-heurísticas caracteriza-se como uma oportunidade para novas pesquisas.

A FIGURA 4(a) apresenta uma visão geral de cada tipo de artigo e métodos de solução. Os métodos mais utilizados são as meta-heurísticas (50,66%). Métodos exatos e heurísticas clássicas são aplicados com menos frequência (15,72% e 3,06%, respectivamente). Isso ocorre provavelmente devido ao fato de que métodos exatos são computacionalmente caros para instâncias complexas e grandes, enquanto soluções de heurísticas clássicas podem ficar presas em ótimos locais (LAHYANI; KHEMAKHEM; SEMET, 2015). Outras técnicas com implementações para problemas aplicados compreendem 24,02%. Observa-se que a soma de todos os itens ultrapassa o total de 100%, e isso acontece devido ao fato de alguns artigos apresentarem mais de uma forma de solução.





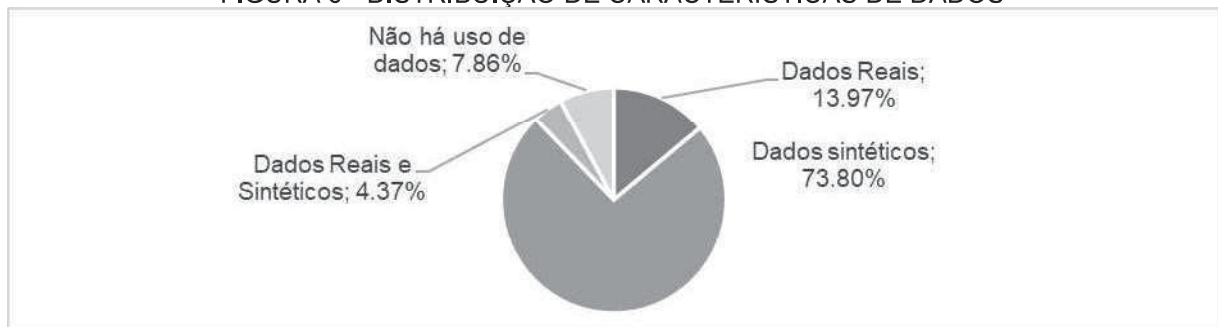
FONTE: A autora (2020).

A FIGURA 4(b) ilustra os métodos de resolução que aparecem em quatro ou mais artigos da amostra, exclusivamente ou em combinação com outros métodos de solução. Estes são representados por ordem de importância da esquerda para a direita. Assim, o método heurístico clássico proeminente foi o algoritmo de *savings* (CLARKE; WRIGHT, 1964). Os autores usam o algoritmo heurístico Clarke & Wright (C&W) modificado seguido por uma fase de busca local para resolver a formulação do problema de roteamento do veículo *Rollon - Rolloff* (RRVRP), problema de roteamento de nós com custo de arco assimétrico e um comprimento máximo de rota (LI et al., 2017). Li et al. (2016b) apresenta um VRP com dois elos com restrições de tempo (2E-TVRP) em um sistema de entregas considerando emissões de dióxido de carbono (CO<sub>2</sub>) utilizando para solução a heurística de C&W melhorada em uma fase de busca local.

Os métodos exatos que se destacam são o *Branch and Cut* (BC) e o *Branch-and-Cut-and-Price* (BCP). Por exemplo, Fukasawa, He e Song (2016) aplicaram o BCP para o VRP com minimização de energia, enquanto Hokama, Miyazawa e Xavier (2016) usaram uma abordagem BC para o VRP com restrições de carga. Cherkesly et al. (2016) aplicaram o BCP para o problema de coleta e entrega com janelas de tempo. Dentre os procedimentos classificados como meta-heurísticas, *Variable Neighborhood Search* (VNS), *Tabu Search* (TS), *Genetic Algorithms* (GA), *Ant Colony Optimization* (ACO), *Simulated Annealing* (SA), *Adaptive Large Neighborhood Search*

(ALNS), LNS, *Iterated Local Search* (ILS), e *Particle Swarm Optimization* (PSO) foram usados em quatro ou mais artigos diferentes, comumente aplicados em combinação com outros procedimentos. Meta-heurísticas têm sido estabelecidas como metodologias de última geração como abordagem de solução dos VRPs. Também foram evidenciados os artigos que apresentavam formulação matemática de programação linear inteira mista (PLIM). Em 143 artigos (62,44%), estas formulações são fornecidas juntamente com abordagens heurísticas para resolver o problema em questão.

FIGURA 5 - DISTRIBUIÇÃO DE CARACTERÍSTICAS DE DADOS



FONTE: A autora (2020).

No que se refere às informações e às características dos dados (FIGURA 5), 73,8% dos artigos utilizam dados artificiais (gerados aleatoriamente seguindo alguma distribuição de probabilidade) para testar seus métodos de solução (169 artigos). Instâncias reais são abordadas em 13,97% dos artigos, enquanto uma combinação de dados artificiais e reais são usados em 4,37% dos artigos. Os trabalhos teóricos representam 7,86% das publicações que não utilizavam nenhum tipo de dados.

### 2.2.1.3 Características de Cenário

A TABELA 3 apresenta a classificação relacionada às características de cenários do VRP. É importante enfatizar que dentre os artigos considerados para análises SC e PPC não estão incluídos aqueles exclusivamente teóricos (20 artigos), totalizando 209 artigos para esta análise.

Os fatores chave para a classificação de dados de entrada (1.1) são a incerteza e a variabilidade de dados durante o período de planejamento. As quatro

classes possíveis de dados de entradas são: determinístico, estocástico, estático e dinâmico. Psaraftis (1988) afirma que VRP pode ser estático ou dinâmico, e pode ser determinístico ou estocástico. Portanto, quatro combinações para essas características são possíveis: SD (estático e determinístico), SS (estático e estocástico), DD (dinâmico e determinístico) e DS (dinâmico e estocástico). Conseqüentemente, muitos estudos têm mais de uma característica de dados de entrada, portanto, a proporção deve ser maior que 100%. Nesta amostra, verificou-se, como esperado, que a maioria dos artigos (84,23%) trate das características de SD. Do restante, 6,31% são SS, 5,41% são DD e 4,05% são DS.

Um *survey* relevante sobre VRPs estocásticos é o trabalho de Gendreau; Jabali e Rei (2016), que examina as principais classes de VRP estocásticos com problemas com demandas, clientes e tempos de viagem ou serviço estocásticos.

TABELA 3 – CARACTERÍSTICAS DE CENÁRIOS

<b>1.1 Dados de entrada</b>	1.1.1 Estático	88,21%
	1.1.2 Dinâmico	10,04%
	1.1.3 Determinístico	87,77%
	1.1.4 Estocástico	11,35%
<b>1.2 Componentes de Gerenciamento de Decisão</b>	1.2.1 <i>Routing</i>	92,72%
	1.2.2 <i>Inventory and routing</i>	0,97%
	1.2.3 <i>Location and routing</i>	2,43%
	1.2.4 <i>Routing and driver scheduling</i>	1,94%
	1.2.5 <i>Production and distribution planning</i>	1,94%
<b>1.3 Número de Depósitos</b>	1.3.1 Único	90,78%
	1.3.2 Múltiplos	7,93%
<b>1.4 Tipo de Operação</b>	1.4.1 <i>Pickup or delivery</i>	90,78%
	1.4.2 <i>Pickup and delivery</i>	16,02%
	1.4.3 <i>Backhaul</i>	5,34%
	1.4.4 <i>Dial-a-ride</i>	0%
<b>1.5 Restrições de Divisão da Carga</b>	1.5.1 <i>Splitting allowed</i>	8,25%
	1.5.2 <i>Splitting not allowed</i>	91,75%
<b>1.6 Planejamento Período</b>	1.6.1 Período único	92,23%
	1.6.2 Multi-períodos	7,77%
<b>1.7 Uso Múltiplo de Veículos</b>	1.7.1 Viagem única	94,17%
	1.7.2 Múltiplas viagens	5,83%

FONTE: A autora (2020).

Estudos sobre o Problema de Roteamento de Veículos Dinâmico (DVRP) são apresentados na literatura com o trabalho seminal de Psaraftis (1988). Em um DVRP, também conhecido como VRP em tempo real ou *online*, parte dos dados de entrada são conhecidos durante a execução do plano. Os eventos dinâmicos mais comuns

são a chegada de novos pedidos de clientes, horários de serviço e horários de viagem. Ritzinger, Puchinger e Hartl (2016) fornecem uma pesquisa que resume a literatura do VRP dinâmico e estocástico. Psaraftis, Wen e Kontovas (2016) desenvolvem uma taxonomia de artigos DVRP de acordo com 11 critérios. Os problemas estocásticos e dinâmicos são bastante desafiadores e constituem quase 10,04% dos trabalhos selecionados.

Considerando as Componentes de Gerenciamento de Decisão (1.2), foram consideradas as seguintes integrações do VRP com outros problemas: Problema de Estoque e Roteirização (*Inventory Routing Problem - IRP*), Problema de Localização e Roteirização (*Location-Routing Problem - LRP*), Problema de Planejamento da Produção e Roteirização (*Production-Routing – PRP*), e *Problema de Programação de veículos e Motoristas (Vehicle and Driver Scheduling - VDS)* (LAHYANI; KHEMAKHEM; SEMET, 2015). O LRP visa determinar a localização de depósitos que atendem a clientes e as rotas respectivas, simultaneamente. Dentre os trabalhos que tratam sobre este tipo de problema tem-se o de Hof, Schneider e Goeke (2017) que apresenta um problema de localização e roteamento de estações de troca de bateria para veículos elétricos capacitados. Turan *et al.* (2017), por sua vez, propõem uma abordagem VNS para resolver o problema de estoque com a demanda incerta em dois elos, onde os produtos perecíveis são distribuídos de um armazém central para lojas de varejo. Outros artigos sobre o tema são Dong e Turnquist (2015) e Li e Tian (2016).

No IRP o fornecedor deve definir, além dos planos de roteamento, as quantidades a serem entregues, conhecendo os níveis de estoque dos clientes, para evitar a falta de estoque. Archetti, Fernández e Huerta-Muñoz (2017) apresentam o Problema de Roteamento do Veículos Periódico Flexível com restrições de estoque (*Flexible Periodic Vehicle Routing Problem with Inventory-Constraints - FPVRP-IC*), que considera níveis de estoques por períodos de tempo e os custos de estoque são considerados na Função Objetivo. Um exemplo recente com múltiplos depósitos é apresentado por Guimarães *et al.* (2019). O PRP, por sua vez, é abordado por (AHMADIZAR; ZEYNIVAND; ARKAT, 2015; FU; ALOULOU; TRIKI, 2017; HEIN; ALMEDER, 2016; MOONS *et al.*, 2017; SCHENEKEMBERG *et al.*, 2020a, 2020b).

No VRP clássico um único depósito é usado, o que pode ser bastante restritivo na prática. Assim, em aplicações da vida real, muitas vezes existem vários depósitos e as rotas podem ter diferentes locais iniciais e finais, logo pode ser necessário alocar

cada cliente ao depósito apropriado. Os problemas que envolvem questões relacionadas aos depósitos podem ter características diferentes em relação a quantidade de depósitos, decisão sobre seleção dos locais a serem instalados e as capacidades de cada depósito, afetando os custos gerais. Os múltiplos depósitos são considerados em 9.22% dos trabalhos. Montoya-torres *et al.* (2015) apresentam uma pesquisa com o estado da arte sobre o MDVRP.

Quanto ao tipo de operação dos problemas (1.4), podem distinguir-se quatro classes de problemas. Problemas de roteamento com *Pickup or deliveries* (1.4.1), em que os produtos são coletados ou recolhidos correspondente ao VRP clássico são abordados por 90.78% dos trabalhos. 16.02% dos trabalhos avaliados são dedicados a *pickup and delivery* (1.4.2), em que as mercadorias podem ser carregadas ou descarregadas em cada cliente. Esta classe é chamada de Problemas de Roteamento de Veículos com Coletas e Entregas (*Vehicle Routing Problems with Pickups and Deliveries* - VRPPD) e existem muitas extensões possíveis para o VRPPD. Uma extensão possível consiste na introdução de restrições de janelas de tempo, conforme apresenta o estudo de Mahmoudi e Zhou (2016). Os autores provaram fornecer soluções ótimas por meio de uma heurística com abordagem de programação dinâmica.

Uma variação é o VRP com Coleta e Entrega Simultânea (VRP *with Simultaneous Pickup and Delivery* - VRPSPD), onde operações de coletas e entregas são realizadas simultaneamente a fim de satisfazer o limite de capacidade do veículo com o objetivo de um total de minimizar a distância de viagens. Kalayci e Kaya (2016) e Avci e Topaloglu (2016) propõem meta-heurísticas para sua solução. Os primeiros desenvolveram um algoritmo híbrido de meta-heurística baseado em um ACO e VNS para sua solução. Avci e Topaloglu (2016) desenvolveram um algoritmo de busca local híbrido integrado com *Tabu Search*. Polat *et al.* (2015) propõem um modelo matemático de PLIM e um algoritmo que combina a clássica heurística de *savings*, o VNS e um mecanismo de perturbação para resolver o VRPSPD.

Tang, Yu e Li (2015) propuseram um algoritmo exato para o Problema de Roteamento e Sequenciamento de Veículos com Múltiplas Viagens de Coleta e Entrega de Clientes para o aeroporto. A característica de viagens múltiplas é encontrada em 12 artigos, algumas aplicações apresentam-se em Chu, Yan e Huang (2017), Francois *et al.* (2016), Rivera, Afsar e Prins (2015) e Wassan *et al.* (2017).

Problemas com a classe *Backhauls* (1.4.3) representam 5,34% da amostra, esta extensão inclui um conjunto de clientes a quem os produtos serão entregues, designados *linehauls* e um conjunto de fornecedores cujos produtos precisam ser transportados de volta para o centro de distribuição denominados *backhauls*. O Problema de Roteirização de Veículos com Backhaul ( *Vehicle Routing Problem with Backhaul* - VRPB) surge em aplicações como a indústria de alimentos de modo que os supermercados e lojas são considerados como os clientes *linehaul* e fornecedores de alimentos são os clientes de *backhaul* (TOTH; VIGO, 2002a). A pesquisa de Koç e Laporte (2018) aponta inúmeras oportunidades de pesquisa sobre as variantes do VRPB, apresentando extensões do problema: viagens múltiplas (WASSAN et al., 2017), carregamento bidimensional (DOMINGUEZ et al., 2016) e janelas de tempo (KÜÇÜKOĞLU et al., 2015).

Finalmente, a última classe (1.4.4) é referente ao problema Dial-a-Ride (*Dial-a-Ride Problems* - DARP), que compartilha várias características de problemas com coleta e entrega, porém, considera o transporte de pessoas. Assim, o nível do critério de serviço (a “qualidade” do serviço) se torna mais importante e complexo. Questões como pontualidade, redução do tempo de espera e duração das rotas são mais críticas nesse caso do que em outros problemas correlatos (CORDEAU; LAPORTE, 2007). Nenhum dos artigos da amostra aborda este tema. DARP surge frequentemente no transporte relacionado com a Área da Saúde, como o transporte de pacientes ou transporte de sangue (DETTI; PAPALINI; DE LARA, 2017; MASMOUDI et al., 2017; MOLENBRUCH et al., 2017).

Outras características, como períodos múltiplos (1.6.2) e entregas fracionadas (1.5.1) também aparecem em 7,77% e 8,25% dos artigos pesquisados, respectivamente. O VRP com Entregas Fracionadas (*Split Deliveries* - SDVRP), introduzido por Dror e Trudeau (1989), caracteriza-se pela possibilidade de visitas múltiplas ao mesmo cliente. O plano de distribuição pode ser calculado durante um único período ou vários períodos. No PVRP, introduzido por Beltrami e Bodin (1974), os pedidos dos clientes são conhecidos antecipadamente, bem como as possíveis combinações para os dias de visita.

Em relação aos usos múltiplos de veículos (1.7), Tang, Yu e Li (2015) propõem um algoritmo exato para o roteamento de veículos de múltiplas viagens e o problema de programação de coletas e entregas dos clientes no aeroporto. A característica de

viagens múltiplas (1.7.2) é encontrada em 12 artigos. Algumas aplicações são encontradas em: Chu, Yan e Huang (2017), Francois et al. (2016), Rivera et al. (2016) e Wassan et al. (2017).

#### 2.2.1.4 Características físicas do problema

As principais características físicas do problema são mostradas na TABELA 4, com os percentuais relacionados a todos os artigos avaliados.

TABELA 4 – CARACTERÍSTICAS FÍSICAS DO PROBLEMA

<b>2.1 Veículos</b>	2.1.1 Tipo	2.1.1.1 Homogêneo	89,69%
		2.1.1.2 Heterogêneo	10,31%
	2.1.2 Número	2.1.2.1 Fixo	9,39%
		2.1.2.2 Ilimitado	90,61%
	2.1.3 Estrutura	2.1.3.1 Compartilhado	2,87%
		2.1.3.2 Não compartimentado	97,13%
	2.1.4 Restrições de capacidade		91,27%
	2.1.5 Política de carregamento	2.1.5.1 Ordem cronológica	5,34%
		2.1.5.2 Nenhuma política	94,66%
	2.1.6 Regulamentação do condutor		2,91%
<b>2.2 Restrições de tempo</b>	2.2.4 Tempo de serviço		24,76%
	2.2.5 Tempo de espera		9,71%
<b>2.3 Estrutura da janela de tempo</b>	2.3.1 Janela de tempo individual		33,98%
	2.3.2 Várias janelas de tempo		0,97%
<b>2.4 Restrições de incompatibilidade</b>			15,05%
<b>2.5 Restrições específicas</b>			21,36%
<b>2.6 Função Objetivo</b>	2.6.1 Objetivo único		94,12%
	2.6.2 Múltiplos Objetivos		5,88%

FONTE: A autora (2020).

A principal característica que aparece nos artigos (91,27%) consiste em veículos capacitados (2.1.4). Os veículos podem diferir de acordo com suas capacidades e, inclusive, o mesmo veículo pode ter vários compartimentos com diferentes capacidades. Goodson (2015), Henke et al. (2015), Lahyani et al. (2015), Silvestrin e Ritt (2017) são responsáveis por 2,87% dos artigos que consideram VRPs com múltiplos compartimentos (2.1.3.1). O uso de múltiplos compartimentos é relevante quando vários produtos, que devem permanecer separados durante transporte, tem que ser carregados pelo mesmo veículo. Empresas de transporte de derivados de petróleo, com divisões devido ao tipo de produto, e as de alimento, com

divisões de temperaturas de acordo com o tipo do produto, são exemplos de empresas que utilizam veículos com mais de um compartimento.

Embora diversas variantes do VRP com Frota Heterogênea (HFVRP) (2.1.1.2) tenham sido desenvolvidas desde o trabalho seminal de Kirby (1959), a maioria dos artigos avaliados consideram veículos idênticos (89.69%). O artigo de Koç et al. (2016) consiste em uma pesquisa sobre este tema. Para outras aplicações, pode-se consultar: (AVCI; TOPALOGLU, 2016; BRUGLIERI et al., 2016; CALVET et al., 2016; DAYARIAN et al., 2015; DE ARMAS; MELIÁN-BATISTA, 2015; DOMINGUEZ et al., 2016; EHMKE; CAMPBELL; THOMAS, 2016; FU; ALOULOU; TRIKI, 2017; IORI; RIERA-LEDESMA, 2015; KOÇ et al., 2015; LAI et al., 2016; MANCINI, 2016; SCHYNS, 2015; WANG; LI; HU, 2015; XIAO; KONAK, 2016; YAO et al., 2016).

Os trabalhos sobre o VRP também podem ser classificados de acordo com o número de veículos disponíveis. Eksioglu, Vural e Reisman (2009) consideram três categorias com base no número de veículos usados: exatamente  $n$  veículos, até  $n$  veículos e uma quantidade ilimitada de veículos. Dentre os artigos pesquisados, 90,61% assume um número ilimitado de veículos de modo que o objetivo é geralmente obter uma solução que minimiza o número de veículos e/ou custo total de viagem. No entanto, problemas reais em empresas envolvem uma série de restrições de recursos, como uma frota fixa, por exemplo. Assim, a restrição de frota fixa foi detectada em 9,39% dos artigos.

A otimização combinada dos problemas de roteamento e carregamento leva a restrições de carregamento mais complexas do que uma restrição simples de peso ou volume, principalmente com o tratamento tridimensional das dimensões das caixas e do contêiner/carroceria do veículo (LAHYANI; KHEMAKHEM; SEMET, 2015). Uma abordagem simples encontrada na literatura é construir rotas de acordo com a ordem cronológica induzida pela política de carga. Os métodos de carregamento/descarga mais comuns são baseadas em duas regras: LIFO (*Last-In-First-Out*), na qual o último item a entrar é o primeiro a sair; e FIFO (*First-In-First-Out*), em que o primeiro item a entrar é o primeiro a sair. A regra LIFO define que se o veículo pegar o pedido  $i$  antes do pedido  $j$ , então deve entregar o pedido  $j$  antes de entregar o pedido  $i$ , respectivamente (CARRABS; CERULLI; CORDEAU, 2007). Cita-se referências dedicadas a diferentes variantes como o Problema de Roteamento de Veículo



Dependente da Carga (*Load-dependent vehicle routing problem – LoVRP*) (ZACHARIADIS; TARANTILIS; KIRANOUDIS, 2016; ZHANG; WEI; LIM, 2015), e abordando o problema com restrições de carregamento tridimensionais (BORTFELDT et al., 2015; MAENNEL; BORTFELDT, 2016).

Como oportunidades de pesquisas futuras Pollaris et al. (2015) apontam que um caminho interessante de pesquisa poderia ser incorporar restrições de distribuição da carga no veículo (KURPEL et al., 2018, 2020; MAENNEL; BORTFELDT, 2016; ZACHARIADIS; TARANTILIS; KIRANOUDIS, 2016). Esses modelos raramente incorporam outras restrições ricas e/ou outras restrições adicionais, como janelas de tempo, frota de veículos heterogênea, distância e duração máximos da rota e/ou regulamentações de motoristas, que podem refletir os problemas reais enfrentados pelos distribuidores.

Embora existam muitas obras na literatura dedicadas ao VRP, restrições relacionadas com a legislação sobre a condução e horários de trabalho não são frequentemente incluídos (LAHYANI; KHEMAKHEM; SEMET, 2015). Estas regras são impostas por regulamentações legais ou pelas próprias empresas. Alguns exemplos são: horas de trabalho diários, número e duração dos períodos de descanso diários e semanais, o número máximo de horas seguidas de condução, horas-extras e salários. Foram identificados seis artigos (2,91%) contendo tais restrições (BAI et al., 2015; BELHAIZA, 2016; GOEL; IRNICH, 2017; HUANG; LIN, 2015; KHODABANDEH et al., 2017; QIAN; EGGLESE, 2016).

Os primeiros estudos sobre a VRP com janela de tempos (VRPTW) foram conduzido por Knight e Hofer (1968), Pullen e Webb (1967) e Solomon (1987). Por sua vez, Solomon (1987) adicionou restrições de janelas de tempo para o VRP clássico e introduziu um conjunto clássico de instâncias, conhecidas como "instâncias de Solomon". Restrições de janela de tempo obrigam que o serviço inicie e termine dentro de uma determinada janela de tempo para cada cliente. Isto significa considerar os tempos de viagem entre as localizações dos clientes, os tempos de serviço nos consumidores e os tempos de carga/descarga no depósito.

A restrição de janela de tempo definida como *hard* refere-se ao caso em que não é permitido que o veículo chegue atrasado. Caso o veículo chegar antes da janela de tempo definida, este deve esperar até que o cliente esteja disponível. Restrições de janelas de tempo definidas como *soft* são aquelas que podem ser violadas

mediante o pagamento de penalidades. Dentre os artigos, 33,98% apresentam formulações e métodos de solução para o VRPTW e suas variantes. Dois artigos (BELHAIZA, 2016; SON; LOUATI, 2016) abordam restrições de múltiplas janelas de tempo, nas quais um dos intervalos de tempo associados a cada cliente deve ser selecionado.

Outra variação do VRP está relacionada as preocupações ambientais o Problema de Roteirização Verde (*Green Vehicle Routing Problems - Green VRP*). No *Green VRP* considera-se a distância mínima de viagem para cada veículo do depósito ao centro de distribuição e o total de emissões de dióxido de carbono do veículo. Tiwari e Chang (2015) tratam de uma abordagem baseada na distância para o cálculo das emissões de dióxido de carbono, considerando o carregamento do veículo como um fator que altera a quantidade de emissões. Xiao e Konak (2017) propõem um modelo de PLIM, que considera o Problema de Roteamento Dependente do Tempo (*Time-Dependent Vehicle Routing Problem - TDVRP*) com veículos heterogêneos, restrições de janela de tempo no cliente/veículo e o impacto das cargas de cada veículo em emissões de poluentes. Outros estudos que levam em conta a variante do *Green VRP* verde podem ser encontrados em: (ÇIMEN; SOYSAL, 2017; KOÇ; KARAOGLAN, 2016; KRAMER; SUBRAMANIAN; PENNA, 2016; LEGGIERI; HAOUARI, 2017; MONTOYA et al., 2016; SOYSAL et al., 2017; TURKENSTEEN, 2017).

Outro exemplo é a decisão que envolve restrições de *cross-docking*, que envolve um conceito de manuseio e distribuição de materiais no qual as mercadorias se movem diretamente do recebimento para o envio. Wang, Lan e Zhao (2017) propõem uma solução método para um VRP com *cross-docking* e entregas fracionadas. Outras aplicações relacionadas são encontradas em: Ahmadizar, Zeynivand e Arkat (2015), Enderer, Contardo e Contreras (2017), Grangier et al. (2017), Maknoon e Laporte (2017) e Yu, Jewpanya e Redi (2016).

O Problema de Roteamento Aberto de Veículos (*Open VRP - OVRP*), introduzido por Schrage (1981) é abordado por 21.36% dos artigos. Neste problema, o motorista não retorna ao depósito quando o último cliente da rota é atendido. Algumas aplicações pertinentes ao OVRP, podem ser encontradas em: Brito et al. (2015), Lalla-Ruiz et al. (2016), Soto et al. (2017) e Yu, Jewpanya e Redi (2016).

Uma grande variedade de funções objetivo é usada na literatura do VRP. Os objetivos mais comuns incluem a minimização de alguns ou todos estes critérios: a

distância total viajada, o tempo total, o custo total da rota, o tamanho da frota; e/ou a maximização da qualidade do serviço ou lucro arrecadado. A avaliação de múltiplos objetivos permite estender os problemas acadêmicos clássicos a fim de melhorar sua aplicação prática (JOZEFOWIEZ; SEMET; TALBI, 2008). O mesmo autor cita outros exemplos de objetivos que podem ser considerados: o balanceamento de distâncias percorridas por rota, visto que reclamações podem advir de horários de trabalhos muito desiguais entre os motoristas; satisfação dos consumidores quanto a datas de entregas; *Makespan*, referente a minimização da rota mais longa e inclusive, solicitações de *Marketing*, importantes para melhorar as vendas e a reputação da empresa. Quando vários objetivos identificados se encontram em conflito, algoritmos adequados têm de garantir algum *trade-off* entre eles. Em 5.88% dos artigos, funções multi-objetivos são estudadas. Exemplos de trabalhos com multiobjetivos podem ser encontrado em Kaiwartya et al., (2015) e Yalcin e Erginel (2015), com uma versão do VRP *Fuzzy* e dinâmico, respectivamente.

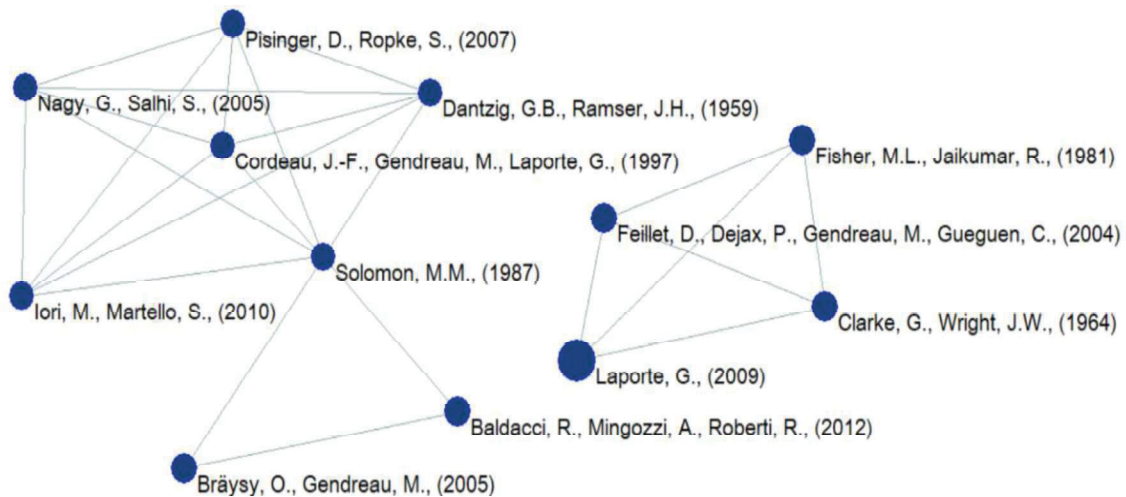
#### 2.2.1.5 Pilares teóricos e Oportunidades de Pesquisa

A fim de identificar os pilares teóricos da amostra, uma rede social foi construída entre os artigos e suas referências. A análise de cocitação mede a grau de conexão entre dois ou mais artigos pelos números de documentos que os referenciam. Esta análise se justifica, pois identifica grupos de conhecimento (ou frentes de pesquisas) estabelecidos por conjuntos de autores que são frequentemente referenciados. O resultado desta análise é mostrado na FIGURA 6, com todos os artigos da amostra compartilhados simultaneamente por pelo menos 14 artigos da rede, revelando certa influência nos artigos da área. A referência de destaque corresponde ao estudo teórico de 50 anos de o VRP por (LAPORTE, 2009). Os trabalhos seminais de Clarke e Wright (1964) e Dantzig e Ramser (1959) são os mais citados.

Outras referências destacam-se em diferentes assuntos: VRP com restrições de janela de tempo (BRÄYSY; GENDREAU, 2005; SOLOMON, 1987); heurística de busca tabu para problemas de roteamento de veículo periódicos, com vários depósitos (CORDEAU; GENDREAU; LAPORTE, 1997); soluções heurísticas (FISHER; JAIKUMAR, 1981; PISINGER; ROPKE, 2007); algoritmo exato para resolver o VRP (FEILLET et al., 2004); restrições de carregamento (IORI; MARTELLO, 2010);

algoritmos heurísticos para problemas de roteamento com coletas e entregas, com um único ou mais depósitos (NAGY; SALHI, 2005).

FIGURA 6 - REDE DE COCITAÇÃO



FONTE: A autora (2020).

Uma classe importante de problemas consiste nos VRP ricos (*Rich VRP*), associados à complexidade dos problemas de roteamento reais. O estudo recente de Lahyani, Khemakhem e Semet (2015) define o RVRP como problemas que estendem as variantes acadêmicas do VRP nos diferentes níveis de decisão, considerando pelo menos quatro aspectos estratégicos e táticos no sistema de distribuição e incluindo pelo menos seis diferentes restrições relacionadas às características físicas. Para problemas definidos exclusivamente por meio de aspectos estratégicos e táticos, pelo menos cinco características devem estar presentes em um problema para caracterizá-lo como um RVRP. Enquanto para problemas definidos exclusivamente por meio de características físicas, pelo menos nove delas devem estar presentes em um RVRP. Dez artigos da amostra consideram o problema em estudo como um RVRP, dentre eles estão: Calvet et al. (2016), De Armas e Melián-Batista (2015) e Mancini (2016).

Oportunidades de pesquisa são identificadas em decorrência da crescente demanda pela resolução de VRPs ricos ou com múltiplos atributos. Esta revisão propiciou a identificação de uma variante pouco explorada na literatura que é o COMVRP. A QUADRO 3 apresenta artigos sobre o COMPVRP, bem como os problemas em específico tratados e os métodos de solução.

QUADRO 3 - VISÃO GERAL DE VARIANTES ESPECÍFICAS DO VRP

	Artigos	Problema	Método
<b>COMVRP</b>	(LIU; JIANG, 2012)	COMVRP	Modelo matemático de programação inteira mista e algoritmo memético
	(BRITO; EXPÓSITO; MORENO, 2016)	COMVRP com janela de tempo	VNS
	(TAVAKKOLI-MOGHADDAM et al., 2019)	COMVRP com frota heterogênea e múltiplos depósitos	Método AHP-TOPSIS para avaliar à insatisfação dos motoristas quanto a distância média percorrida por rota.
	(AZADEH; FARROKHI-ASL, 2019)	COMVRP com múltiplos depósitos	Modelo matemático de programação inteira mista e uma meta-heurística híbrida
	(AMALIA; SUPRAJITNO; PRATIWI, 2020)	COMVRP	<i>Bat Algorithm (BA)</i>

FONTE: A autora (2020).

Como pode-se perceber o primeiro trabalho sobre o COMVRP é de 2012. Por se tratar de um domínio relativamente recente da pesquisa, várias oportunidades podem ser identificadas a partir deste problema. Importante mencionar que os trabalhos citados foram os únicos encontrados na literatura referente à variante COMVRP. Este problema não foi estudado incorporando-se outras restrições importantes, como periodicidade, janelas de tempo, restrições de comprimento máximo da rota e duração ou regulamentos do motorista que podem refletir os problemas reais enfrentados pelos distribuidores.

A continuação da revisão da literatura é realizada com o objetivo de abordar as variantes que foram associadas ao COMVRP neste trabalho, bem como as principais características de algoritmos que tratam do VRP com múltiplos atributos. A identificação destas estratégias promissoras utilizadas por estes algoritmos pode contribuir no desenvolvimento de novos métodos unificados eficientes.

### 2.2.2 Revisão de características de algoritmos para o MAVRP

Para as análises realizadas nesta seção, foram avaliados os principais artigos recentes que tratam do MAVRP, analisando-se artigos que trabalham com os atributos clássicos com instancias comparáveis. O QUADRO 4 apresenta os principais artigos analisados devido a sua notoriedade em tempos de solução para diferentes variantes do VRP. Apresenta-se para cada método proposto, quais e quantas foram as variantes

consideradas, sendo que estas podem ser testadas em conjunto. A relação dos resultados obtidos por cada um destes artigos com as Melhores Soluções Conhecidas BKS (*Best Known Solutions*) obtidas são apresentadas de acordo com as informações disponibilizadas por cada artigo. Os artigos que apresentam publicações de resultados parciais ou conceitos importantes e descrições que não constam no artigo avaliado também são analisados.

QUADRO 4 - RELAÇÃO DAS CARACTERÍSTICAS DOS ARTIGOS COM ALGORITMOS UNIFICADOS PARA MAVRP AVALIADOS

Artigo / Algoritmo Proposto	Variantes contempladas	No. variantes testadas / No. instâncias	Artigos anteriores relacionados
(VIDAL et al., 2014) / Unified Hybrid Genetic Search (UHGS)	<i>Multiple depots (MDVRP), Multiple periods (PVRP), Heterogeneous fleet (HVRP), Site-dependent (SDVRP), Split deliveries (VRPSD), Profits (TOP), Multiple trips (MTVRP), Pickup &amp; deliveries (VRPPD), Backhauls (VRPB), Asymmetric (AVRP), Open (OVRP), Cumulative (CCVRP), Load-dependent costs (LDVRP), Simultaneous P.&amp; D. (VRPSDP), Mixed Pickup &amp; deliveries (VRPMDP), Vehicle Fleet Mix (VFMP), Duration constraints (DurVRP), Hard TW (VRPTW), Soft TW (VRPSTW), Multiple TW (VRPMTW), General TW (VRPGTW), Time-dep. travel time (TDVRP), Flexible travel time (VRPFTT), Lunch breaks (VRPLB), Work hours regulations (VRTDSP), Site choice (Generalized VRP) (GVRP)</i>	29 variantes testadas / 1099 instâncias de referência testadas: 95% de melhores soluções conhecidas foram recuperadas ou melhoradas. 541 novos BKS.	(VIDAL et al., 2012, 2013a)
(KOÇ et al., 2015) / Unified Hybrid Evolutionary Algorithm (HEA)	<i>Vehicle Fleet Mix with Time Windows (VMFPTW), Heterogeneous fleet with Time Windows (HVRPTW), com restrição de tempo ou distância, máxima em rota;</i>	4 variantes / 360 instâncias de referência testadas: 49% de BKS foram recuperadas ou melhoradas. 75 novos BKS.	(DEMIR; BEKTAŞ; LAPORTE, 2012)
(PENNA et al., 2017) / Unified Hybrid Heuristic, HILS-RVRP	<i>Multiple depots (MDVRP), Heterogeneous fleet (HVRP), Multiple trips (MTVRP), Backhauls (VRPB), Mixed Backhauls (VRPB), Asymmetric (AVRP), Open (OVRP), Time Windows (VRPTW), Route Duration (RDVRP)</i>	12 variantes/643 instâncias de referência testadas: 71,70% de BKS são recuperadas ou melhoradas.	(PENNA; SUBRAMANIAN; OCHI, 2013; SUBRAMANIAN et al., 2012)

FONTE: O autor (2020).

Os primeiros estudos de VRP apresentavam soluções específicas para cada variante. Soluções unificadas, no entanto, surgiram mais recentemente. O trabalho recente de Vidal et al. (2014), aborda o desenvolvimento de um único algoritmo que é eficiente e aplicável a uma ampla família de variantes do VRP. Os autores demonstram através de testes computacionais em 29 variantes que a generalidade não vai necessariamente contra a eficiência para essas classes de problemas. Da mesma forma, Penna et al. (2017) propõem um método genérico para resolver vários problemas de roteamento de veículos com frota fixa por meio de um VNS.

Outro artigo relevante consiste no estudo de Erdoğan (2017) que apresentou um solver no *Microsoft Excel*<sup>®</sup>, desenvolvido usando o *Visual Basic for Applications*

(VBA). O *solver* foi disponibilizado pelo autor para *download* em um *site* acadêmico sem custo (ERDOGAN, 2013), e foi baixado mais de 12.855 vezes, até o momento desta pesquisa. O *solver* desenvolvido, no entanto, pode ser entendido e modificado por programadores de nível médio e traz uma interface amigável e acessível.

O VRP *Spreadsheet Solver* proposto pode resolver mais de 64 variantes do VRP, com base em recursos relacionados a visitas seletivas aos clientes, coletas e entregas simultâneas, janelas de tempo, composição da frota, restrição de distância e destino dos veículos. O *solver* foi capaz de resolver instancias de CVRP com até 200 clientes em 1h de CPU Time. Em termos de velocidade, é difícil qualquer comparação com os outros algoritmos de estado-da-arte, visto que o C++ pode ser mais de 2,7 vezes mais rápido que o VBA (o autor realizou um experimento simples para verificação - criou uma matriz de números inteiros com 10.000 elementos, preencheu essa matriz com  $[U [0, 1] * 1000]$  e repetiu esse processo 10.000 vezes). Além do fato de que o gerenciamento de memória de ponteiros do C++ é conhecido por ser muito eficiente.

#### 2.2.2.1 Descrição das variantes clássicas associadas ao problema estudado

A seguir apresenta-se uma breve descrição de variantes clássicas estudadas em MAVRPs e utilizadas neste trabalho. Uma revisão bibliográfica dos problemas tratados pode ser encontrada nos trabalhos: Braekers, Ramaekers e Van Nieuwenhuyse (2016), Cordeau et al., (2002, 2007); Gendreau et al. (2008), Golden, Raghavan e Wasil (2008) e Toth e Vigo (2014). Vidal et al. (2013b) apresentaram uma revisão sobre Problemas de Roteamento de Veículos com Múltiplos Atributos.

- OVRP: Neste problema o veículo não precisa retornar ao depósito depois de visitar o último cliente. Esta variante foi revisada por (LI; GOLDEN; WASIL, 2007). Zachariadis e Kiranoudis (2010) propuseram uma Busca Tabu, envolvendo memórias para armazenar avaliações de rotas, enquanto Fleszar, Osman e Hindi (2009) propuseram uma heurística eficaz com VNS. Um bom desempenho também foi alcançado pelo algoritmo híbrido de Repoussis et al. (2010). Brandão (2018) desenvolveu um algoritmo efetivo e eficiente para resolver o

OVRPTW, com um uso inovador e eficaz de cadeias de ejeção e soluções de elite.

- MDVRP: Considera-se mais de um depósito disponível, mas cada veículo deve iniciar e terminar no mesmo depósito em uma rota. O número de veículos por depósito é geralmente limitado. Revisões de literatura pode ser encontrada em: (KARAKATIČ et al., 2015; MONTOYA-TORRES et al., 2015). Ao considerar meta-heurísticas, soluções de alta qualidade foram produzidas pela UTS paralela (MAISCHBERGER; CORDEAU, 2011), ALNS (PISINGER; ROPKE, 2007), GA híbrido guiado por lógica *Fuzzy* (LAU et al., 2010) e Busca Híbrida Genética com Controle Avançado de Diversidade (HGSADC) (VIDAL et al., 2012) e a Busca Tabu híbrida granular (ESCOBAR et al., 2014).
- PVRP: Uma dimensão de tempo é introduzida, pois o planejamento de rota deve ser executado em um horizonte de vários períodos. Cada cliente requer um número total de serviços de acordo com algumas combinações aceitáveis de períodos de visita chamados padrões. A atribuição de visitas a clientes está, portanto, sujeita a restrições de compatibilidade com os padrões. O PVRP é revisado em (FRANCIS; SMILOWITZ; TZUR, 2008). Métodos exatos (BALDACCI et al., 2011) foram capazes de resolver alguns casos com até 100 clientes e 6 períodos de tempo.
- HVRP: Os clientes são atribuídos a tipos de veículos com características diferentes: capacidade, tempos máximos de rota, custos fixos e custos variáveis em termos de distância. Quando o número de veículos não é restrito, o problema é geralmente chamado de *Fleet Size and Mix* (FSM), caso contrário, a versão mais difícil é chamada de VRP Heterogêneo (HVRP) (ver Baldacci, Battarra e Vigo (2008), para uma revisão). As meta-heurísticas de última geração são baseadas hibridizações ILS e VNS (PENNA et al., 2013). Específico ao FSM, à Busca Tabu de Brandão (2011) e o híbrido GA de Liu, Huang e Ma (2009) também produzem soluções de boa qualidade.



- Restrições de duração da rota (DurVRP): existe um limite de distância percorrida ou tempo para as rotas.

### 2.2.2.2 Principais características dos algoritmos para MAVRP

Para analisar os conceitos de heurística do MAVRP, identificam-se três categorias de atributos, definidos em relação ao seu impacto na resolução heurística:

- ASSIGN: a atribuição de clientes e rotas para recursos (depósitos, dias, tipos de veículos);
- SEQ: que determinam a estrutura da rede e sequências de visitas e atributos
- EVAL: avaliação de sequências fixas

Essa classificação simples está intimamente conectada com as metodologias de resolução, pois lidar com esses três aspectos problemáticos leva a um método de solução completo. Observe que um atributo pode afetar vários aspectos do problema e, portanto, possivelmente aparecer em várias categorias. O QUADRO 5 apresenta exemplos de variantes clássicas abordadas pela literatura e sua classificação.

QUADRO 5 - VARIANTES FREQUENTEMENTE ENCONTRADAS NA LITERATURA

ASSIGN	SEQ	EVAL
<i>Multi-Depots</i> <i>Heterogeneous</i> <i>Multi-Periods</i> <i>Split Deliv.</i> <i>Prize Collect.</i> <i>Location</i> <i>Site</i> <i>Dependent</i> <i>Inventory</i> <i>Consistency</i>	<i>Backhauls</i> <i>Pick&amp;D.</i> <i>Multiple Trips</i> <i>Multi-Echelons Truck &amp; Trailer</i> <i>Generalized Graph specifics: Tree,</i> <i>Shoreline...</i>	<i>Open</i> <i>Time Windows</i> <i>Time dependent</i> <i>HOS Regulations</i> <i>2D-3D Loading Soft &amp; Multiple TW Duration</i> <i>Constr.</i> <i>Other time feat.</i> <i>Cumulative costs</i> <i>Simult Pick &amp; Deliv</i> <i>Pollution/Green</i> <i>Synchronization</i>

FONTE: Adaptado de Vidal et al. (2013b).

Vidal et al. (2013b) avalia 15 diferentes atributos clássicos do VRP em 64 meta-heurísticas diferentes que se destacavam pela qualidade de solução e/ou tempo computacional de resolução. Neste estudo, os autores concluem que os melhores métodos heurísticos não sustentam seu desempenho em uma única ideia. Por sua

vez, esses métodos apresentam um equilíbrio e uma complementaridade de conceitos.

Uma característica relevante dentre os algoritmos para o MAVRP estudados, refere-se ao espaço de busca e vem do potencial uso de soluções inactíveis (CORDEAU; LAPORTE; MERCIER, 2001; NAGATA; BRÄYSY; DULLAERT, 2010; VIDAL et al., 2012). Estudos realizados desde o trabalho de Glover (1986) relatam que uma exploração controlada de soluções inactíveis pode melhorar a busca, permitindo que ela faça a transição mais facilmente entre soluções viáveis estruturalmente diferentes. Além disso, o uso de soluções inactíveis pode contribuir para melhorar a robustez do método, que é menos dependente da disponibilidade de uma solução inicial viável (que é muitas vezes Problema NP-difícil). Um ponto importante a ser mencionado é que o artigo de Vidal et al. (2014) cita que as penalizações são ajustadas dinamicamente durante a execução do algoritmo para favorecer a geração de indivíduos factíveis. Vale mencionar que no artigo de Penna et al. (2017), apenas soluções factíveis são aceitas, com exceção apenas das variantes com janela de tempo.

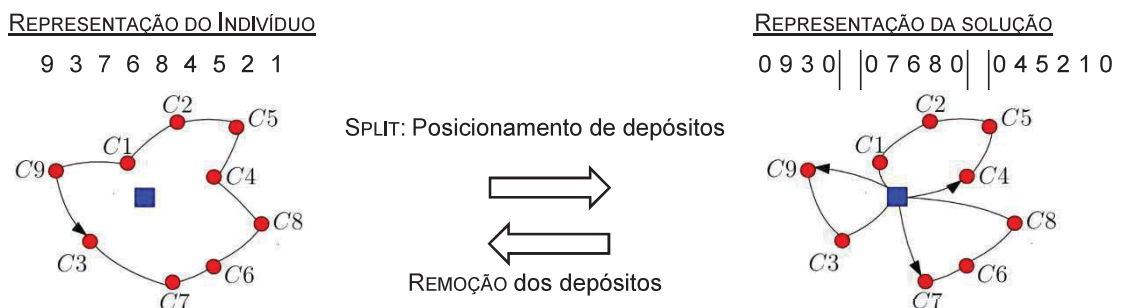
Outra característica refere-se à representação da solução. Para muitos MAVRPs, definir uma solução completa vai além da descrição da rota, já que diversas decisões adicionais relacionadas a diferentes atributos são especificadas. Muitas meta-heurísticas são então projetadas para explorar um espaço de busca feito de representações indiretas de soluções. Essas representações contêm apenas informações suficientes para que um algoritmo decodificador eficiente possa ser aplicado a fim de extrair uma ou várias soluções completas (VIDAL et al., 2013a).

O artigo seminal de Vidal et al. (2014), bem como em estudos anteriores (BOUDIA; PRINS; REGHIOUI, 2007; LIU; HUANG; MA, 2009; NGUEVEU; PRINS; CALVO, 2010; PRINS, 2009; VIDAL et al., 2012, 2013a) para aplicações em AG, e (DUHAMEL; LACOMME; TOUSSAINT, 2011) em GRASP + ELS), utilizam a representação de uma Grande-rota (*giant tour*) de Prins (2004) (FIGURA 7). Essa representação pode ser vista como uma rota do Problema do Caixeiro Viajante (*Travelling Salesman Problem* -TSP) para um veículo com capacidade infinita.

O fato de algoritmos baseados em AGs não terem obtido bons resultados nas tentativas iniciais pode ser explicado pelo modo de construção da representação da solução (ARAÚJO, 2008). As primeiras tentativas foram baseadas em soluções

(definidas por cromossomos no AG) com delimitadores de rotas, o que obrigava o algoritmo a executar procedimentos de reparo para transformar as soluções ineficazes obtidas nos crossovers em soluções viáveis. Para corrigir tais distorções dos cromossomos com delimitadores de rotas, é preciso utilizar algoritmos que corrigem os cromossomos gerados, o que eleva o custo computacional do crossover e prejudica a transmissão de bons trechos de rotas para as gerações seguintes, como sugere (PRINS, 2004). Este autor apresentou um algoritmo genético híbrido que se mostrou superior a todos os outros algoritmos genéticos até então publicados, comparando-se as instâncias de Christofides (1979) e Golden et al. (1998).

FIGURA 7 - REPRESENTAÇÃO DO INDIVÍDUO E DA SOLUÇÃO



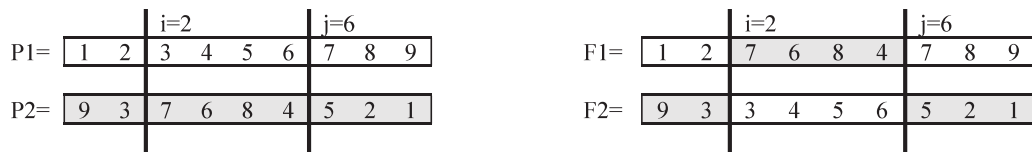
FONTE: A autora (2020).

Dois são os fatores fundamentais para seu sucesso:

- i. a codificação do cromossomo sem delimitadores de rotas, que permite a utilização do operador de crossover OX (*Order Crossover*) – método de cruzamento utilizado com mais sucesso nos algoritmos genéticos para VRP segundo Prins (2004).
- ii. formação das rotas pelo método de particionamento da Grande-rota, chamado de *split procedure*, o que garante que, para a sequência de clientes determinada pelo cromossomo, seja obtida a rota ótima.

Para ilustrar o fato de como os operadores de *crossover* comuns geram soluções ineficazes, apresenta-se o problema do caixeiro-viajante, em que o cromossomo de solução é representado por uma sequência ordenada de cidades a serem visitadas (FIGURA 8). Sejam P1 e P2 dois pais selecionados para cruzamento, e suponha que seja aplicado um operador de cruzamento em dois pontos, após a 2ª e 6ª posições.

FIGURA 8 - REPRESENTAÇÃO DE UM CRUZAMENTO EM DOIS PONTOS



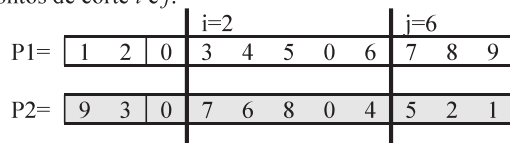
FONTE: A autora (2020).

As soluções F1 e F2 não representam roteiros válidos para o problema do caixeiro-viajante, uma vez que algumas cidades aparecem duas vezes, enquanto outras cidades não são visitadas. A fim de evitar esse tipo de problema, utilizam-se operadores específicos para tratar cromossomos que representam uma permutação. O operador OX, proposto por Davis (1985), é um exemplo de operador que permite tal tratamento (FIGURA 9). Dois pontos de corte são selecionados aleatoriamente determinando uma sub-rote; um pré-filho é então criado, copiando-se a sub-rote na posição correspondente do P1; os elementos da sub-rote são excluídos do P2; a partir do segundo ponto de corte, são preenchidas as posições vazias com os elementos não excluídos do pai 2; quando o fim do cromossomo foi atingido, continua-se a partir da primeira posição deste. Além deste, existe o operador *cycle crossover*, *partially mapped*, *position-based*, e *order-based* (KUMAR; GOPAL; KUMAR, 2013).

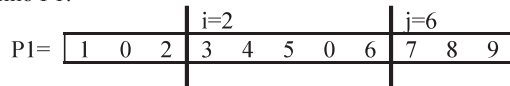
FIGURA 9 - CROSSOVER OX EM CROMOSSOMOS P1 E P2 (A) COM DELIMITADORES DE ROTA E (B) SEM DELIMITADORES DE ROTA.

CROSSOVER OX COM DELIMITADORES DE ROTA

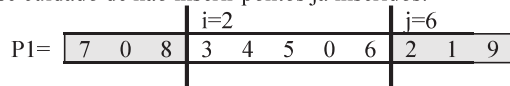
Seleciona-se os cromossomos P1 e P2 e escolhe-se dois pontos de corte  $i$  e  $j$ .



Características do P1 são passadas para o filho F1:



F1 é completado com os pontos a partir de  $j$ , tomando-se cuidado de não inserir pontos já inseridos:

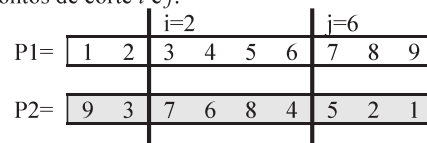


O mesmo procedimento é realizado para a obtenção de F2.

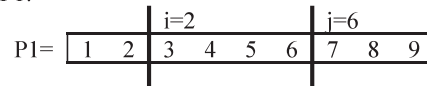
(a)

CROSSOVER OX SEM DELIMITADORES DE ROTA

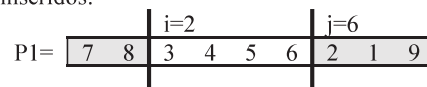
Seleciona-se os cromossomos P1 e P2 e escolhe-se dois pontos de corte  $i$  e  $j$ .



Características do P1 são passadas para o filho F1:



F1 é completado com os pontos a partir de  $j$ , tomando-se cuidado de não inserir pontos já inseridos:



O mesmo procedimento é realizado para a obtenção de F2.

(b)

FONTE: A autora (2020).

Em relação ao procedimento *Split*, este algoritmo reduz o problema de encontrar os delimitadores de rota para um problema de caminho mais curto em um gráfico acíclico auxiliar. Este algoritmo pode ser adaptado a cenários com custos penalizados e tamanho limitado da frota.

A busca em vizinhança de forma eficiente é outro tópico essencial para o desempenho dos MAVRPs, pois geralmente contribui para a maior parte do esforço computacional (VIDAL et al., 2014). Com exceção do artigo de Lau et al. (2010) que apresenta um algoritmo genético que utiliza uma mutação cruzada e aleatória, todos os algoritmos para MAVRP são baseadas em vizinhanças enumeráveis. Os algoritmos para MAVRP analisados na presente seção tem vizinhanças baseadas em um conjunto de movimentos. A partir de uma solução inicial  $s$ , uma heurística de busca local (*Local Search* – LS) explora uma vizinhança  $\mathcal{N}(s)$ , usando diferentes trocas de arco em  $s$  a fim de obter uma solução melhor  $s'$  substituindo  $s$  em uma nova iteração da heurística. Esta solução é um ótimo local do problema e da vizinhança usada.

A inclusão de componentes aleatórios nas várias escolhas de algoritmos é uma característica dominante das trajetórias de busca. Penna et al. (2017) realiza a busca local por meio do procedimento *Randomized Variable Neighborhood Descent* (RVND) de. No RVND ao invés de usar uma ordem pré-definida de vizinhanças para explorar o espaço de soluções, utiliza uma ordem aleatória a cada chamada. Mais especificamente, sempre que em uma determinada vizinhança não for possível melhorar a solução corrente, o RVND seleciona, aleatoriamente, outra vizinhança para continuar a busca pelo espaço de soluções. Esta abordagem, segundo Penna et al. (2013), produz, em média, resultados melhores do que a versão com ordem determinística das vizinhanças e tem vantagem de não necessitar de um estudo de qual a melhor ordem de vizinhanças.

Penna et al. (2017) aplica um conjunto  $\mathcal{N}$  de dez tipos de vizinhanças. Primeiramente, apresenta-se as estruturas de vizinhanças inter-rotas. Destes sete movimentos são válidos em todas as variantes. Destas estruturas, cinco são baseadas em movimentos  $\lambda$ -*interchanges* (OSMAN, 1993), considerada a troca de até  $\lambda=2$  clientes entre duas rotas. Essas trocas são melhores explicadas como duplas  $(\lambda_1, \lambda_2)$  (com  $\lambda_1 < \lambda$  e  $\lambda_2 < \lambda$ ), sendo que  $\lambda_1$  clientes são transferidos da rota  $r_1$  para a rota  $r_2$

e  $\lambda_2$  clientes da rota  $r_2$  para a rota  $r_1$ . Portanto, são avaliados os seguintes movimentos: Inserção (*Shift*) ((1; 0); (2; 0)); e Troca (*Swap*) ((1; 1); (2; 1); (2; 2)).

Uma sexta vizinhança considerada consiste em trocar dois segmentos de diferentes rotas (2Opt). E a última, *k-Shift*, consiste em mover  $k$  clientes consecutivos para de uma rota para o final de outra. Para as variantes com Múltiplos Depósitos duas vizinhanças, denominadas *ShiftDepot* e *SwapDepot*, foram implementadas. Por fim, duas vizinhanças para as variantes contendo entregas divididas, denominadas *RouteAddition* e *k-Split*, foram adotadas. Quanto as estruturas de vizinhança intra-rota, foram adotadas cinco estruturas de vizinhança intra-rotas bem conhecidas, a saber: Reinserção (similar ao movimento *Shift* (1,0)), *Or-Opt* (com dois e três clientes adjacentes), *2-Opt* e troca (similar ao *Swap* (1,1)). Essas vizinhanças também foram implementadas usando um procedimento RVND e são chamadas toda vez que uma solução é modificada durante a busca intra-rota.

Em (VIDAL et al., 2014), a melhoria das rotas é realizada selecionando-se aleatoriamente um vértice  $u$  e um vértice próximo  $v$  (segundo um limite de granularidade – conceito que será apresentado na sequência). Sendo  $x$  e  $y$  os vértices subsequentes a  $u$  e  $v$  na rota. A Busca Local repete, em ordem aleatória, sobre cada vértice  $u$  e cada um de seus vizinhos  $v$  avaliando os movimentos de trocas e remoção com um ou dois elementos, trocas de 2 elementos, e movimentos de 2-Opt entre rotas e intra-rotas. Os movimentos de troca e inserção são aplicados indiferentemente nas mesmas rotas ou em rotas diferentes. Os movimentos são examinados em ordem aleatória, sendo que o primeiro a produzir uma melhoria é implementado.

Koç et al. (2015) apresenta operadores destruição e operadores de reparo também selecionados aleatoriamente considerando seu desempenho passado e uma certa probabilidade. São nove operadores de remoção: aleatória; do pior em distância; do pior em tempo; de vizinhança; de Shaw; baseada em proximidade; baseada no tempo; baseada em demanda; e por custo médio por unidade. São três os operadores de inserção: Guloso; Guloso com função de ruído; e Guloso com o tempo em rota.

Devido ao grande esforço computacional que demanda esta etapa, técnicas que visam restringir a vizinhança, ou utilizá-la de forma mais eficiente. Como exemplo, cita-se a busca granular de Toth e Vigo (2002b) utilizadas por (VIDAL et al., 2012, 2013b, 2014) para reduzir o número de clientes avaliados nos movimentos. O

benefício de se utilizar um coeficiente de granulometria reside na maior eficiência da exploração do espaço da vizinhança, pois restringe a busca aos vizinhos com maior probabilidade de serem caminhos para as melhores soluções.

Outra característica importante é a gestão da diversidade populacional que se mostrou um fator chave de sucesso na obtenção de um bom desempenho para os MAVRPs (PRINS, 2004; GOEL e VIDAL, 2012 e VIDAL et al. 2012a). Afinal, manter a diversidade e o elitismo simultaneamente em uma população é difícil, visto que os procedimentos agressivos de melhoria local, usados na maioria das meta-heurísticas eficientes, tendem a direcionar fortemente a população para alguns ótimos locais, resultando em convergência prematura.

Por fim, os algoritmos avaliados operam procedimentos de gestão de diversidade, fazendo uso de uma métrica de distância entre indivíduos. Para os MAVRPs, essa métrica é usualmente baseada em diferenças entre as soluções (PRINS, 2004) ou similaridades nas sequências de rotas (PRINS, 2009; VIDAL et al., 2012, 2014)(PRINS, 2009; VIDAL et al., 2012, 2014), ou são projetadas especificamente para os atributos considerados (aqueles da categoria ASSIGN especialmente, por exemplo, (VIDAL et al., 2012). A diversidade pode então ser controlada por diferentes meios. Lau et al. (2010) baseiam-se na lógica *fuzzy* para adaptar parâmetros de busca relativamente à diversidade populacional e medidas de qualidade. Finalmente, o HGSADC (VIDAL et al., 2014) não considera a diversidade como uma restrição, mas como uma parte integral do objetivo que compete com a qualidade da solução durante avaliações individuais. Estudos empíricos realizados por este último autor mostram que o último mecanismo leva a uma maior diversidade e qualidade da solução.

### 2.3 CONSIDERAÇÕES FINAIS

Esta seção apresentou uma revisão sistemática da literatura, classificando artigos sobre o VRP em relação às taxonomias de Braekers, Ramaekers e Van Nieuwenhuysse (2016) e de Lahyani, Khemakhem e Semet (2015), que abordaram artigos até 2015. Os artigos deste trabalho foram, portanto, selecionados a partir de 2015, avaliando-se características físicas do problema, características dos dados

utilizados, características de cenário, os tipos de estudos realizados, bem como os pilares teóricos e oportunidades de pesquisas.

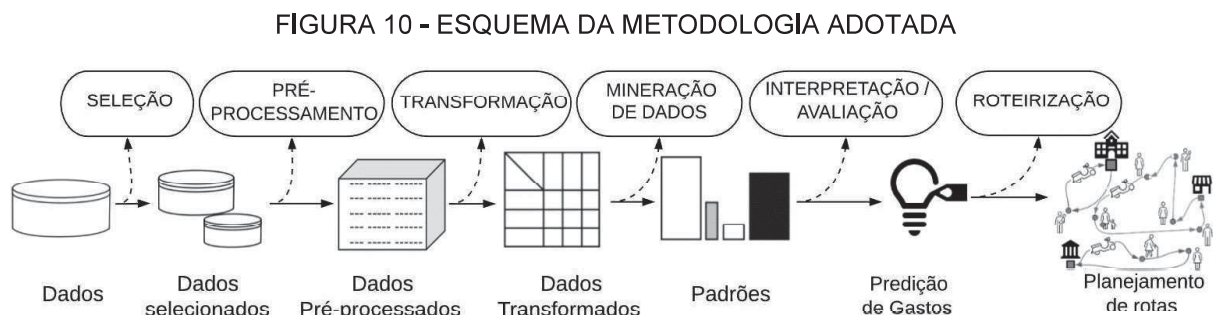
Embora uma ampla gama de variantes do VRP tenha sido identificada, observou-se uma tendência de estudos em direção às variantes do VRP com múltiplos atributos. Assim, identificou-se como uma variante com potencial a ser explorado o COMVRP, cujo primeiro artigo sobre o tema foi publicado em 2012 (LIU; JIANG, 2012). A essa variante é possível associar outras restrições e características de forma a tornar a solução mais aplicável no problema desta pesquisa. As variantes clássicas do VRP que são associadas ao COMVRP neste são também apresentadas.

A solução desse problema torna-se um desafio por conta da complexidade dos métodos de solução capazes de lidar com essas diversas características de forma unificada. E para contribuir com o desenvolvimento de novos métodos unificados investigou-se quais foram as estratégias utilizadas por algoritmos recentes sobre o MAVRP selecionados por seu alto desempenho. Identificou-se que os métodos bem-sucedidos não são determinados por um único fator, mas são o resultado de vários elementos da metodologia que envolve, por exemplo, o uso de diferentes espaços de busca, vizinhanças variáveis, *trade-off* entre a diversificação e intensificação, procedimentos para restringir a vizinhança.



### 3 METODOLOGIA

Este estudo foi realizado utilizando-se dados reais em um grupo de lojas de uma empresa do setor varejista. Um esquema da metodologia adotada para este estudo está apresentado na FIGURA 10. Inicialmente, realizou-se a seleção e limpeza dos dados a partir da base de dados disponibilizada pela empresa. No processo de limpeza, foram eliminados eventos de compras com dados faltantes ou informações incorretas. As informações coletadas na empresa em estudo são fornecidas pelos próprios clientes, que precisam registrar-se e fornecer dados pessoais antes de comprar. Nesta empresa, o fornecimento de dados pessoais está associado à algumas vantagens relacionadas à descontos que beneficiam o cliente.

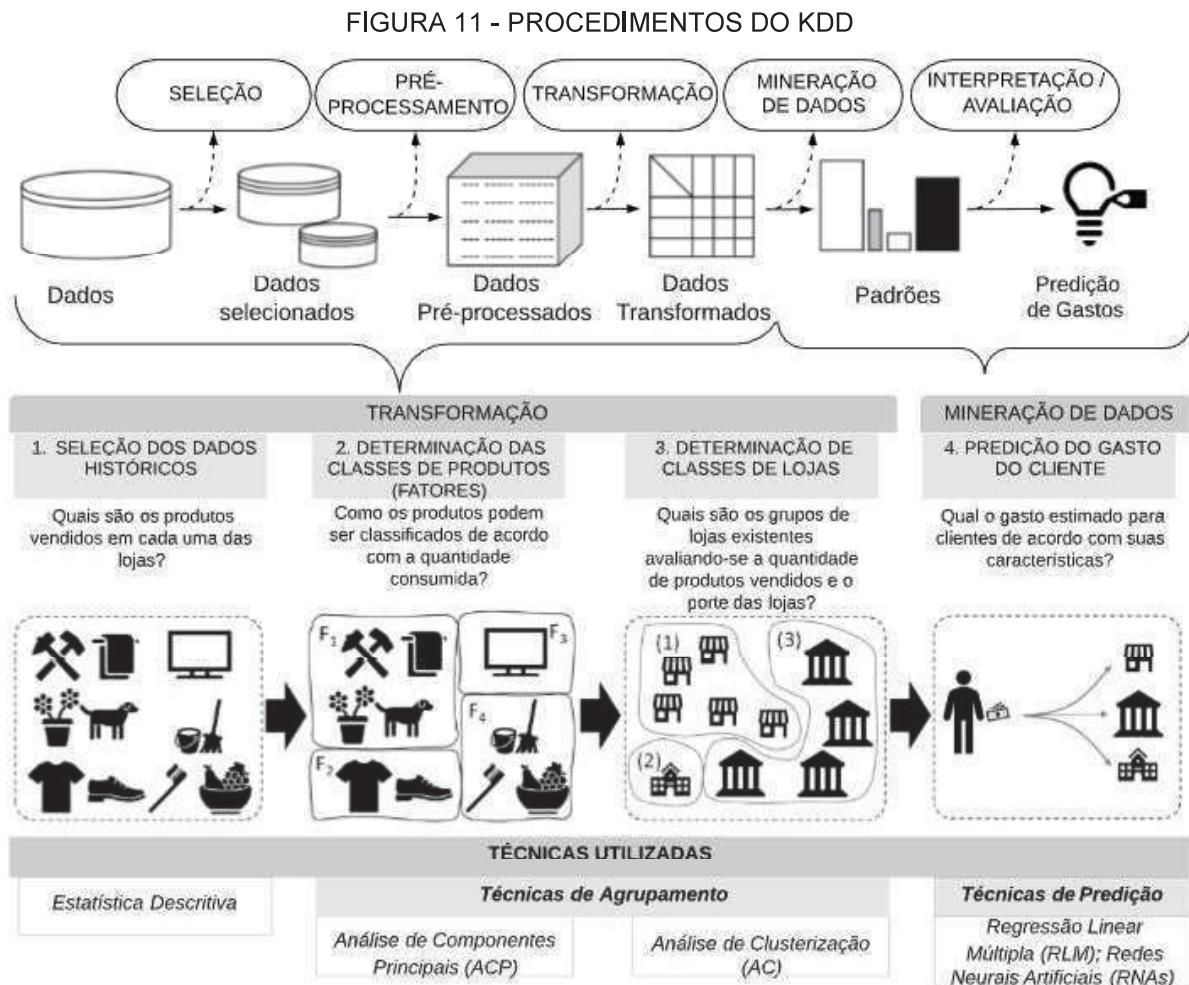


FONTE: A autora (2020).

A FIGURA 11 apresenta em detalhes as etapas de transformação e mineração de dados, ilustrando os procedimentos adotados para realização deste estudo, os questionamentos levantados ao longo da pesquisa e as técnicas utilizadas para a solução.

A base de dados utilizada nesta pesquisa foi extraída de um grupo de lojas varejistas situadas em Curitiba-PR. A base de dados cedida pelo grupo é composta por: histórico de vendas mensal com as especificações dos produtos adquiridos, além dos dados do cadastro do cliente e em qual loja foi efetuada a compra, compreendendo um período de análise de julho a setembro de 2019. Este período foi escolhido por não apresentar fenômenos estacionais e cíclicos, além de representar um conjunto de dados de tamanho significativo para que a análise quantitativa fosse

realizada. Esses eventos de compras contendo a identificação dos itens consumidos bem como suas quantidades são as variáveis de entrada da ACP.



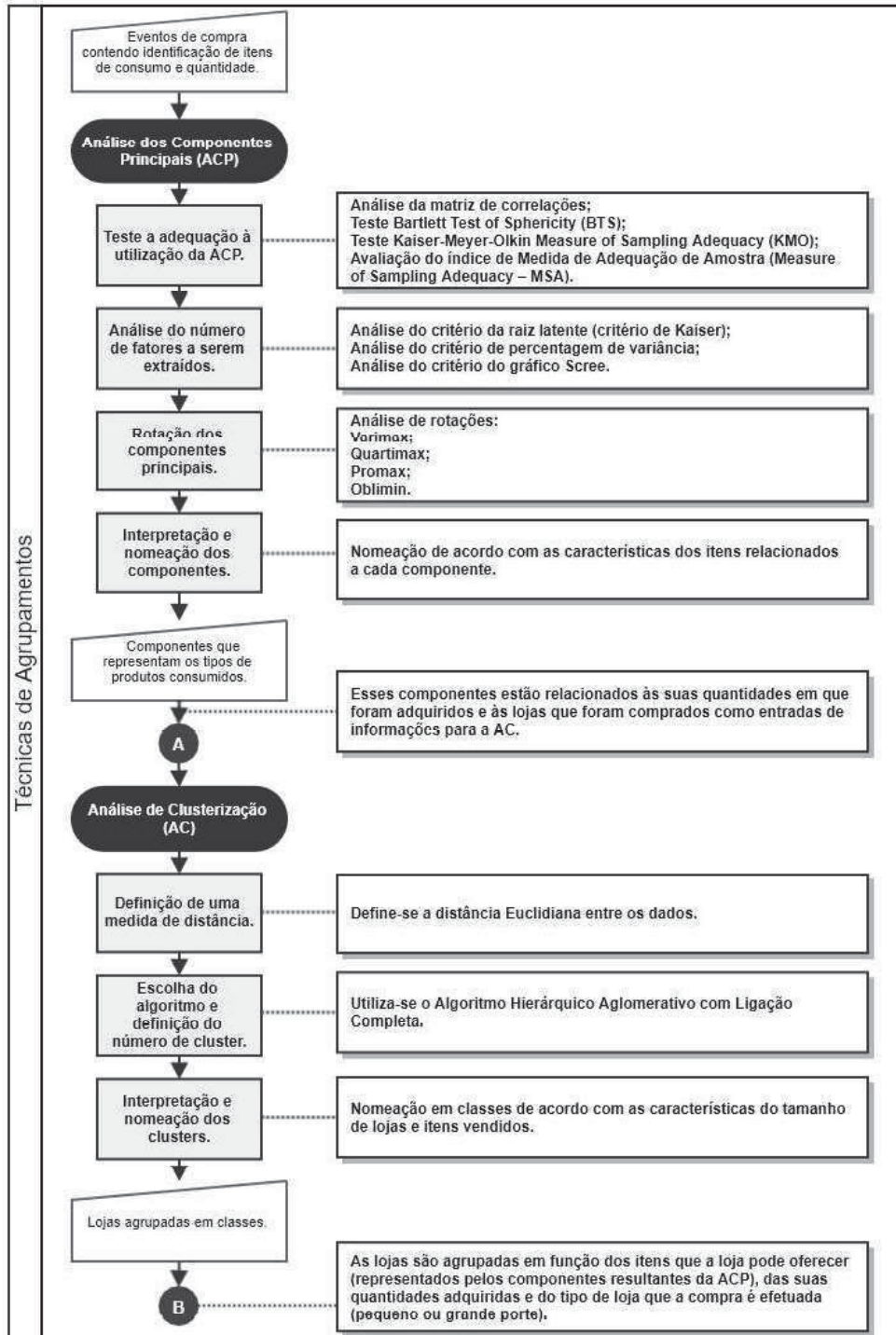
FONTE: A autora (2020).

Na ACP, realizou-se inicialmente testes para averiguação da adequação dos dados, seguido pela análise do número de fatores a serem extraídos avaliando-se os critérios de Kaiser, da porcentagem da variância e do gráfico de Scree. Em seguida, avalia-se a rotação dos componentes e realiza-se a sua interpretação de acordo com os tipos de produtos consumidos pelos clientes.

As variáveis de entradas para a AC são estes componentes relacionados às quantidades que foram adquiridos e lojas que foram comprados. Utilizou-se o algoritmo hierárquico aglomerativo com ligação completa, considerando o critério de distância euclidiana entre os dados. Como resultado desta análise tem-se os

agrupamentos de lojas em classes em função dos itens que a loja pode oferecer e do tipo de loja que a compra é efetuada (pequeno ou grande porte). O detalhamento da utilização das técnicas de ACP e AC são apresentados na FIGURA 12.

FIGURA 12 – DETALHAMENTO DAS TÉCNICAS DE AGRUPAMENTO (ACP E AC)



FONTE: A autora (2020).

Para a aplicação das técnicas preditivas de RLM e RNAs considerou-se como variável resposta (dependente) o gasto realizado pelo consumidor (EXPENDITURE) e como variáveis explicativas os fatores de consumo obtidos após a ACP, além de outras informações relacionadas às características dos clientes, como idade, sexo, endereço, renda, distância do cliente até a loja e relacionadas à loja que o cliente realizou a compra como o porte da loja (TABELA 5). As informações disponibilizadas pela empresa foram idade, gênero, endereço e itens preferidos dos clientes. A renda dos clientes foi inferida com base no seu endereço, por meio dos registros do (CENSO, 2015) com valores de renda de acordo para cada um dos 2.395 setores censitários existentes no município de Curitiba. As técnicas de predição são aplicadas para cada classe de loja resultante da AC, um esquema da utilização destas técnicas está apresentado na FIGURA 13.

TABELA 5 – RELAÇÃO DOS ATRIBUTOS RELACIONADOS AOS EVENTOS DE COMPRAS

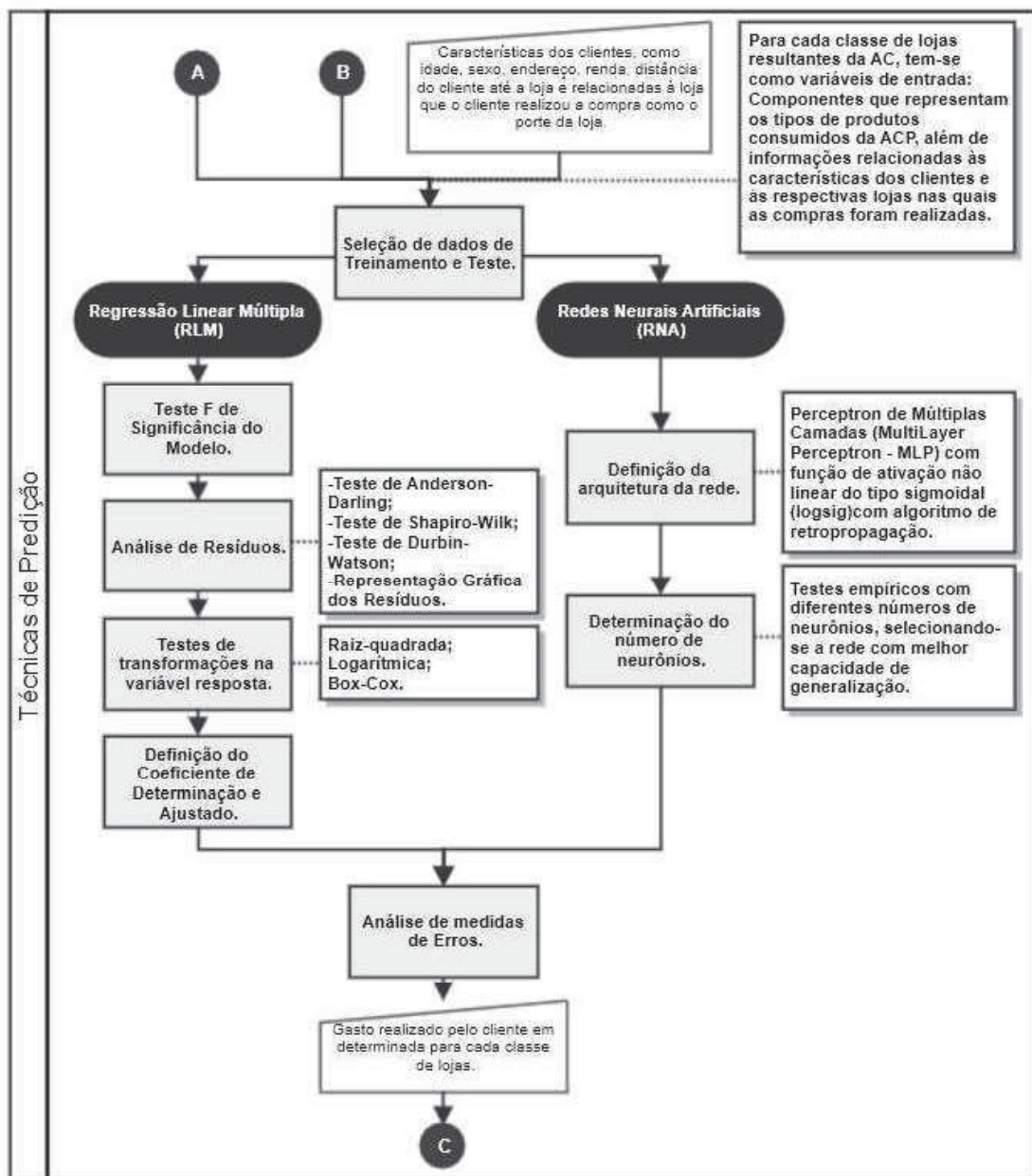
<b>Atributos</b>	<b>Categorias</b>
Idade (AGE)	Quantidade
Gênero (SEX)	0 = Feminino 1 = Masculino
Renda (INCOME)	1 = Sem rendimento nominal mensal 2 = Rendimento nominal mensal de até 1/2 salário-mínimo 3 = Rendimento nominal mensal de mais de 1/2 a 1 salário-mínimo 4 = Rendimento nominal mensal de mais de 1 a 2 salários-mínimos 5 = Rendimento nominal mensal de mais de 2 a 3 salários-mínimos 6 = Rendimento nominal mensal de mais de 3 a 5 salários-mínimos 7 = Rendimento nominal mensal de mais de 5 a 10 salários-mínimos 8 = Rendimento nominal mensal de mais de 10 a 15 salários-mínimos 9 = Rendimento nominal mensal de mais de 15 a 20 salários-mínimos 10 = Rendimento nominal mensal de mais de 20 salários-mínimos
Endereço (ADDRESS)	[1,2395]
Distância do Cliente até a loja (DISTANCE)	Quantidade
Artigo Preferido (F)	[1,11]

FONTE: A autora (2020).

Quanto à determinação do tamanho de amostra, (HAIR et al., 2009) afirma que a proporção mínima de observações por variáveis é 5:1, mas a proporção preferida é de 15:1 ou 20:1. O tamanho da amostra deve ser adequado para garantir poder estatístico e generalização, e quanto mais graus de liberdade melhor é a generalização. Os graus de liberdade podem ser calculados subtraindo-se do tamanho da amostra o número de parâmetros estimados. Visto que são consideradas 18 variáveis dependentes, considerando-se a proporção 20:1, a proporção mínima de

observações necessárias seria de 360. O número utilizado para análise é de 1000 observações para cada uma das classes de lojas. Além disso, a fim de permitir a comparação dos resultados, a amostra foi dividida aleatoriamente em 800 dados de treinamento e 200 dados de teste. Em relação ao pré-processamento de dados, devido à existência de variáveis binárias e variáveis contínuas, aplicou-se a normalização dos dados no intervalo entre 0-1.

FIGURA 13 – DETALHAMENTO DAS TÉCNICAS DE PREDIÇÃO (RLM E RNA)



FONTE: A autora (2020).

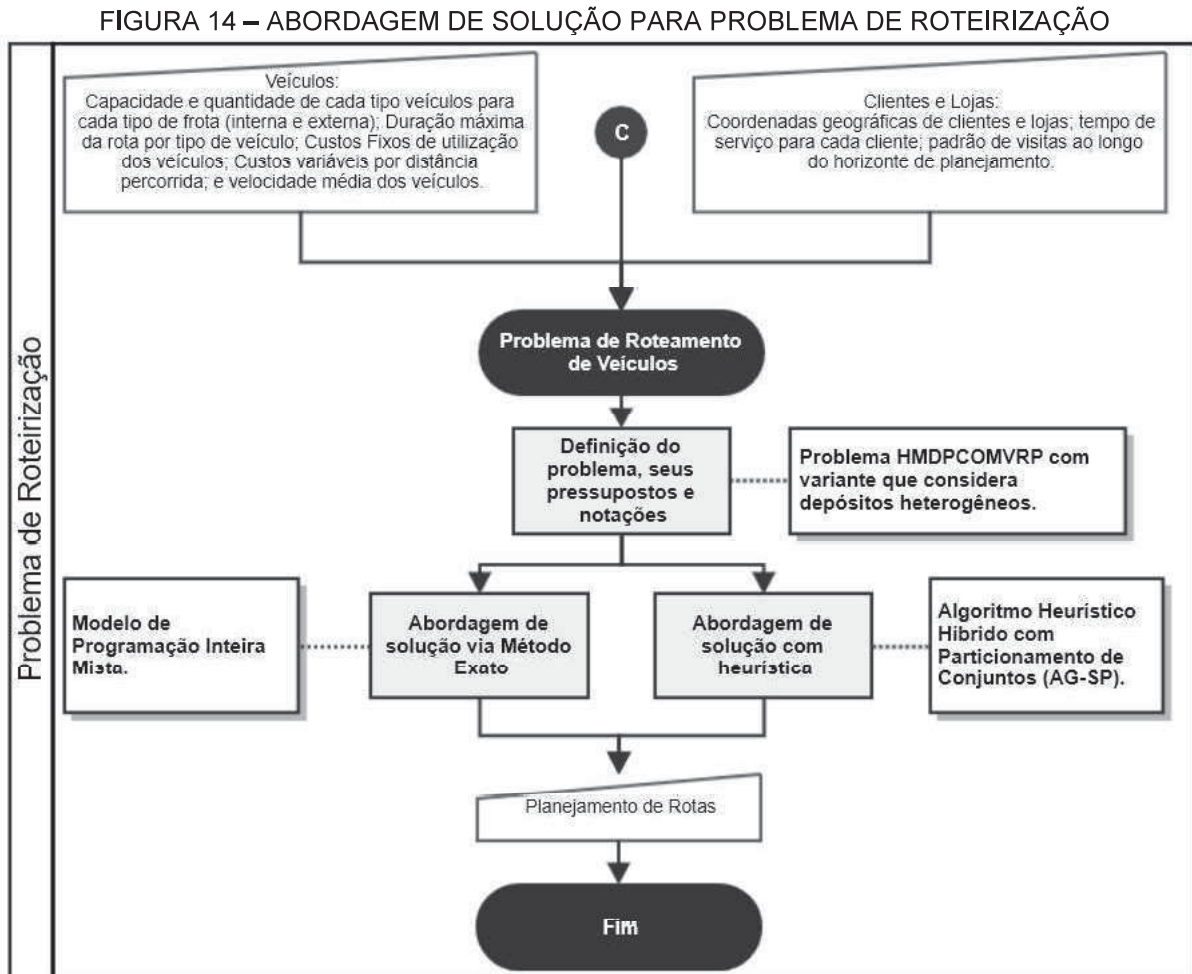
Para a RLM, realizam-se testes de significância do modelo e análise de resíduos. Constatadas violações de uma ou mais suposições, realiza-se testes de transformação da variável resposta, seguido pela identificação do coeficiente de determinação e coeficiente de determinação ajustado para os modelos gerados.

No caso da RNA, inicia-se pela determinação da arquitetura da rede, adotando-se Perceptron de Múltiplas Camadas com função de ativação não linear do tipo sigmoideal (*logsig*) com algoritmo de retropropagação. Testes empíricos com diferentes números de neurônios são realizados selecionando-se a rede com melhor capacidade de generalização. Para as duas técnicas são analisados o conjunto de treinamento e de testes e os erros são computados. Como resultado, tem-se as estimativas de valores gastos por cada perfil de cliente em cada loja da rede, de acordo com a sua classe.

Utilizou-se o *software* R (TEAM, 2013) versão 1.2.1335 para realizar a aplicação da metodologia. Este consiste em um *software* livre e de código aberto que permite a criação e o agrupamento de funções em arquivos executáveis dentro de seu ambiente, comumente chamados de ‘pacotes’. A seguir são apresentados os pacotes utilizados para cada etapa:

- Foram utilizados os pacotes *nortest* para o teste de Anderson-Darling (AD) e *forecast* para a transformação Box-Cox das variáveis, conforme descreve-se na Seção 4.
- Para as RNAs, o pacote utilizado foi o *Neuralnet*.
- Para a Análise Fatorial, utilizou-se o *ggfortify* e *psych*.
- O pacote *corrplot* foi utilizado para ilustrar a correlação entre variáveis.
- Para a análise de clusterização, utilizou-se *cluster* e *factoextra*.

Em relação a etapa de roteamento de veículos, parte-se de informações relacionadas a veículos, clientes e lojas, tal como descrito na FIGURA 14.



FONTE: A autora (2020).

As informações de demanda dos clientes relacionadas às estimativas de gastos são obtidas pelas técnicas preditivas RLM e RNAs que são utilizadas comparativamente, com o objetivo de escolher aquela que minimize os erros para cada classe de lojas avaliadas. Inicia-se o problema pela definição do problema, seus pressupostos e notações e apresenta-se duas abordagens de solução – exata e heurística.

Para a primeira abordagem, apresentou-se a proposta de um modelo matemático de Programação Inteira Mista para o problema, seguido da sua versão expandida, considerando-se depósitos heterogêneos. Para resolver o modelo proposto, utilizou-se o *solver Gurobi Optimization* versão 9.0.2. Devido à complexidade do problema, uma heurística híbrida, denominada de HGA-SP, foi proposta. As abordagens de solução aplicadas aos problemas foram implementadas na linguagem de programação C++ e os testes computacionais foram realizados em

um processador *AMD A10 PRO-7800B R7, 3,5GHz* com 8GB de memória principal com o sistema operacional *Windows 64bits*. A Seção 3.1 descreve o modelo matemático de PLIM com suas definições, pressupostos e notações, e a Seção 3.2 apresenta o HGA-SP proposto.

### 3.1 DESCRIÇÃO DO PROBLEMA DE ROTEAMENTO DE VEÍCULOS E FORMULAÇÃO MATEMÁTICA PROPOSTA

Nesta seção descreve-se o problema do planejamento de rotas tratado neste trabalho. Considera-se que um número de clientes, número de depósitos e demanda do cliente são predefinidos. Existe uma frota de veículos própria que pode, por vezes, ser insuficiente para o atendimento de toda a demanda. Assim, vindo a demandar a contratação de uma frota terceirizada. Tanto a frota própria quanto a terceirizada são compostas por veículos heterogêneos. Sendo que para o caso de frota própria ela tem necessariamente uma quantidade limitada de veículos. Enquanto para a frota terceirizada pode ser considerada ilimitada. Pode-se considerar um conjunto de motoristas disponíveis para os quais um custo fixo é destinado, no caso de frota própria. Ou mesmo, o custo fixo da contratação de um veículo no caso de frota terceirizada. Considera-se também o planejamento das entregas ao longo de um horizonte de planejamento. De modo que, cada cliente possa escolher o dia ou padrão de dias em que deseja receber o seu produto.

#### 3.1.1 Definição do problema

O HMDPCOMVRP é formalmente definido nesta seção. Seja  $G = (V, A)$  um grafo completo, onde  $A$  é o conjunto de arcos. O CVRP considera um único depósito, diferentemente do MDVRP. Neste se considera que o conjunto de vértices  $V = V^D \cup V^C$ , é constituído pelo conjunto de depósito de clientes ( $V^C$ ) e pelo conjunto de depósitos ( $V^D$ ), de onde os produtos são retirados para a entrega. Vértices  $v_i \in V^C$  representam clientes. Cada cliente  $i, i = 1, \dots, n$ , está associado a uma demanda determinística  $q_j$  a ser entregue (ao depósito atribui-se uma demanda  $q_0 = 0$ ). Uma distância não negativa  $c_{ij}$  é associada a cada arco  $(i, j) \in A$  e representa o custo da viagem do nó  $i$  ao nó  $j$ .



É válido mencionar que no clássico do MDVRP proposto por Cordeau et al. (1997), considera-se um número fixo de veículos disponíveis por depósitos. Neste trabalho, no entanto, existe um número total limitado de veículos de frota própria ( $s=0$ ) e frota de veículos terceirizados ( $s=1$ ). Estas características relacionam-se ao COMVRP e implicam a distinção do tipo de frota de modo que apenas os veículos próprios precisem retornar aos depósitos, enquanto os veículos terceirizados não (característica do OVRP).

Outra característica considerada, relaciona-se à heterogeneidade de veículos característica do HVRP. O que implica que cada frota de veículos, tanto própria quanto terceirizada pode ser composta por  $k$  tipos de veículos. O número de veículos de cada tipo  $k$  de frota  $s$  é limitado e representado por  $a_{sk}$ . Pode-se considerar que o número de veículos de frota terceirizada seja ilimitado ( $a_{sk} = +\infty$ ). Cada veículo de frota  $s$  do tipo  $k$  tem a capacidade  $Q_{sk}$ . Incorre-se em um custo por distância percorrida  $V_{sk}$ , e um custo fixo  $F_{sk}$ . É comum considerar o custo fixo apenas para os veículos terceirizados, assumindo-se assim,  $F_{0k} = 0$ . O custo operacional de um veículo é igual à soma de seu custo fixo e custo variável.

Uma dimensão de tempo é acrescentada ao problema, a fim de se planejar rotas ao longo de um horizonte de planejamento de  $|T|$  dias. No problema clássico do PVRP (CORDEAU; GENDREAU; LAPORTE, 1997) cada cliente  $i$  apresenta uma frequência  $f_i$ , que representam o número de visitas a serem realizadas ao longo do horizonte de planejamento. Embora essa característica seja incluída, para a aplicação realizada considera-se  $f_i = 1$ , ou seja, considera-se que cada cliente será atendido apenas uma vez ao longo do horizonte de planejamento. Além disso, cada cliente  $i$  apresenta uma lista  $L_i$  de possíveis combinações de visitas, chamadas de padrões  $p$ . A duração da viagem de um veículo por rota é calculada pela distância total da viagem dividido pela velocidade média dos veículos ( $\vartheta$ ) somada ao tempo de serviço em cada cliente  $i$  ( $\tau_i$ ), e é limitada a um valor máximo de  $H_{isk}$  para todo veículo do tipo  $k$  da frota  $s$ .

O objetivo do problema é determinar a combinação ideal dos veículos a serem utilizados, em conjunto com os padrões de dias de entregas para cada cliente, construindo as rotas correspondentes, de acordo com restrições de capacidade e duração de viagem. De tal modo a minimizar o custo operacional de todos os veículos,

ou seja, a soma dos custos fixos totais dos veículos e dos custos variáveis totais para todos os períodos.

Os principais pressupostos assumidos são:

- (1) Veículos próprios devem retornar aos depósitos de onde partiram, enquanto os veículos terceirizados não precisam fazê-lo.
- (2) Cada rota é atribuída a exatamente um veículo.
- (3) A duração total de viagem de uma rota não pode exceder a duração máxima da jornada de trabalho estabelecida.
- (4) Cada cliente deve ser atendido por um único veículo, mas um veículo pode atender a um conjunto de clientes, a menos que a soma das demandas exceda a capacidade do veículo.
- (5) O número de qualquer tipo de veículos próprios ou terceirizados é predeterminado.

### 3.1.2 Definições e notações

As definições e variáveis estão resumidas na TABELA 6.

TABELA 6 - NOTAÇÕES USADAS PARA O MODELO

Conjuntos e Índices:	
$V$	Todos os vértices; $V = V^D \cup V^C$
$V^D \in V$	Depósitos; $i, j, d \in \{1, \dots,  V^D \}$
$V^C \in V$	Clientes; $i, j \in \{1, \dots,  V^C \}$
$S$	Tipo de frota própria (0) ou terceirizada (1); $s \in \{0,1\}$
$K_s$	Tipos de Veículos; $k \in \{1, \dots,  K_s \}$
$M_{sk}$	Número de veículos disponíveis do tipo de frota $s$ com tipo de veículo $k$ ; $m \in \{1, \dots,  M_{sk} \}$
$L_i$	Lista de padrões de dias de entrega do cliente $i$ que possuem uma frequência de entrega definida $f_i$ ; $p \in \{1, \dots,  L_i \}$
$T$	Dias do horizonte de planejamento, $t \in \{1, \dots,  T \}$
Parâmetros:	
$S$	Veículos de frota própria ( $s=0$ ) e veículos de frota terceirizada ( $s=1$ )
$F_{sk}$	Custo Fixo do veículo $k$ (R\$/m)
$V_{sk}$	Custo Variável por distância percorrida do veículo $k$ com o tipo de frota $s$
$c_{ij}$	Distância percorrida do nó $i$ para $j$
$H_{sk}$	Duração máxima da rota do veículo $k$
$q_i$	Demanda do cliente $i$
$q_{id}$	Demanda do cliente $i$ no depósito $d$
$Q_{sk}$	Capacidade do veículo $k$ e frota $s$
$a_{sk}$	Número de veículos do tipo $k$ de frota $s$
$\tau_i$	Tempo de serviço no cliente $i$
$\alpha_{pt}$	1, se o dia $t$ pertence ao padrão de visitas $p$
$\vartheta$	Velocidade média dos veículos
Variáveis de Decisão:	

$x_{ijskmt}$	1, se o $k$ -ésimo veículo do tipo $s$ viaja do nó $i \in V$ para o nó $j \in V$ ; 0 caso contrário
$u_{isk}$	Variável contínua que representa o <i>upper bound</i> da carga de veículos ao deixar o nó $i$
$h_{isk}$	Variável contínua que representa o <i>upper bound</i> da distância de viagem, de veículos carregamento de veículos até deixar ao nó $i$
$y_{iskm}$	1, se o $m$ -ésimo veículo do tipo $k$ do tipo de frota $s$ é utilizado pelo nó $i \in V$ ; 0 caso contrário

FONTE: A autora (2020).

### 3.1.3 Modelo de Programação Inteira Mista para o HMDPCOMVRP

Usando a notação apresentada acima, um modelo PLIM pode ser formulado para minimizar o custo total das operações de entrega.

$$\text{Min } Z = \sum_{t \in T} \sum_{d \in V^D} \sum_{s \in S} \sum_{k \in K} F_{sk} \sum_{m \in M} \sum_{i \in V^D} \sum_{j \in V^C} x_{ijskmt} \quad (22)$$

$$+ \sum_{t \in T} \sum_{d \in V^D} \sum_{s \in S} \sum_{k \in K} V_{sk} \sum_{m \in M} \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ijskmt}$$

$$- \sum_t \sum_d \sum_{k \in K} V_{sk} \sum_{m \in M} \sum_{i \in V^C} \sum_{j \in V^D} c_{ij} x_{ij1kmt}$$

$$\sum_{p \in L_i} \sum_{d \in V^D} y_{ipd} = 1, \quad \forall i \in V^C \quad (23)$$

$$\sum_{j \in V} \sum_{s \in S} \sum_{k \in K_s} \sum_{m \in M_{sk}} x_{ijskmt} - \sum_{p \in L_i} \alpha_{pt} y_{ipd} = 0, \quad \forall i \in V^C, t \in T, d \in V^D, i \neq j \quad (24)$$

$$\sum_{j \in V} x_{ijskmt} \leq 1, \quad \forall i \in V^D, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j \quad (25)$$

$$\sum_{j \in V} x_{ijskmt} = 0, \quad \forall i \neq d, i \in V^D, d \in V^D, s \in S, k \in K_s, m \in M_{sk}, t \in T, i \neq j \quad (26)$$

$$\sum_{i \in V} x_{ijskmt} = \sum_{i \in V} x_{jiskmt}, \quad (27)$$

$$\forall j \in V^C, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j$$

$$\sum_{i \in V^C} x_{ij0kmt} = \sum_{i \in V^C} x_{ji0kmt}, \quad \forall j \in V^D, k \in K_0, m \in M_{0k}, t \in T, d \in V^D, i \neq j \quad (28)$$

$$h_{iskmt} = 0, \quad \forall i \in V^D, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j \quad (29)$$

$$h_{iskmt} + c_{ij}/\vartheta + \tau_i - M(1 - x_{ijskmt}) \leq h_{jskm}, \quad (30)$$

$$\forall i \in V, j \in V^C, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j$$

$$h_{i0kmt d} + \vartheta \cdot c_{ij} + \tau_i - M(1 - x_{ij0kmt d}) \leq H_{0k}, \quad (31)$$

$$\forall i \in V^C, j \in V^D, k \in K_0, m \in M_{0k}, t \in T, d \in V^D, i \neq j$$

$$h_{i1kmt d} + \vartheta \cdot c_{ij} + \tau_i - M(1 - x_{ij1kmt d}) \leq H_{1k}, \quad (32)$$

$$\forall i \in V^C, j \in V^C, k \in K_1, m \in M_{1k}, t \in T, d \in V^D, i \neq j$$

$$0 \leq h_{iskmt d} \leq \sum_{j \in V} x_{jiskmt d} \cdot H_{sk}, \quad (33)$$

$$\forall i \in V^C, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j$$

$$\sum_{j \in V} x_{jiskmt d} = y_{iskmt d}, \quad \forall i \in V^C, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j \quad (34)$$

$$\sum_{i \in V^C} q_i \cdot y_{iskmt d} \leq Q_{sk}, \quad \forall s \in S, k \in K_s, m \in M_{sk}, t \in T, i \neq j \quad (35)$$

$$y_{iskmt d} = 0, \quad \forall i \in V^C, \forall d \in V^D, s \in S, k \in K_s, m \in M_{sk}, t \in T, q_{id} > Q_{sk} \quad (36)$$

$$\sum_{j \in V^C} x_{ijskmt d} \leq a_{sk}, \quad \forall i \in V^D, s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j \quad (37)$$

$$x_{jiskmt d} = \{1,0\}, \quad \forall i, j \in V, s \in S, k \in K_s, m \in M_{sk} \quad (38)$$

$$y_{iskmyd} = \{1,0\}, \quad \forall i \in V, s \in S, k \in K_s, m \in M_{sk} \quad (39)$$

$$u_{iskmt d} \geq 0, \quad \forall i \in V, s \in S, k \in K_s, m \in M_{sk} \quad (40)$$

$$h_{iskmt d} \geq 0, \quad \forall i \in V, s \in S, k \in K_s, m \in M_{sk} \quad (41)$$

A Função Objetivo que minimiza custos fixos e variáveis está descrita em (22). As restrições (23) garantem que exatamente um depósito e uma combinação de visitas seja atribuída a cada cliente. As restrições (24) garantem que cada visita aos clientes aconteça apenas em períodos relacionados a escolha de combinação de visitas deste cliente, em um veículo do tipo  $k$  da frota  $s$ , vindo do depósito  $d$ . As restrições (25) garantem o uso de um único veículo por rota. As restrições (26) garantem a compatibilidade entre a designação do depósito  $d$  com os vértices de início das rotas. O conjunto de restrições (27) e (28) são restrições de conservação de fluxo. O primeiro define que cada veículo que chega em um cliente deve partir dele. O segundo indica que cada veículo de frota própria deve retornar ao mesmo depósito e essa limitação é executada apenas em veículos próprios porque não há necessidade de veículos terceirizados retornarem aos depósitos. Cada veículo apresenta

limitações quanto a duração da rota percorrida. As restrições de amplitude de duração bem como a eliminação de sub-rotas são dadas em (29) - (33), onde  $M$  representa um grande número positivo. As restrições (34) definem uma nova variável de decisão e especifica a relação entre as variáveis. As restrições (35) garante que a carga de cada rota não exceda a capacidade do veículo. As restrições (36) definem que o cliente  $i$  não será atendido pelo veículo de frota  $s$  com tipo  $k$  se esta carga do cliente for superior a capacidade deste veículo. As restrições (37) limitam o número de veículos próprios e terceirizados. As restrições (38) a (41) são restrições de domínio.

#### 3.1.4 Modelo de Programação Inteira Mista HMDPCOMVRP com depósitos heterogêneos

Como já mencionado, a versão de depósitos heterogêneos foi proposta por Calvet et al. (2016) e aplicada ao MDVRP. Para este problema, os autores assumiram que as lojas não são iguais entre si, o que permite considerar as preferências dos clientes.

Nesta tese, da mesma forma, assume-se que as demandas não são parâmetros fixos ( $q_i$ ), mas dependentes da atribuição dos clientes às lojas ou depósitos  $d$  ( $q_id$ ). Assume-se que as demandas não são conhecidas, mas podem ser previstas com base em um banco de dados histórico e informações sobre os clientes. O procedimento apresentado nesta tese leva em consideração a combinação de modelos preditivos de mineração de dados para prover as informações de demanda para o planejamento de rotas.

Assim, após a análise dos dados históricos, a mineração de dados captura a relação entre a disposição de cada cliente em gastar (variável resposta) em função de várias variáveis (preditores), que incluem o depósito designado e também as características dos clientes, como preferências, renda, sexo, idade etc. Em seguida, as rotas são construídas, associadas a atribuição de cliente para cada depósito.

O problema HMHDPCOMVRP, portanto, foi adaptado para considerar depósitos heterogêneos. Para tanto, a Função Objetivo descrita por (42) deve ser considerada, seguida pelas restrições (22) a (41) anteriormente apresentadas.

$$\begin{aligned}
Max Z = & \sum_{t \in T} \sum_{d \in V^D} \sum_{s \in S} \sum_{k \in K_s} \sum_{m \in M_k} \sum_{i \in V^C} q_{id} \cdot y_{iskmtd} \\
& - \left( \sum_{t \in T} \sum_{d \in V^D} \sum_{s \in S} \sum_{k \in K} F_{sk} \sum_{m \in M} \sum_{i \in V^D} \sum_{j \in V^C} x_{ijskmt} \right. \\
& + \sum_{t \in T} \sum_{d \in V^D} \sum_{s \in S} \sum_{k \in K} V_{sk} \sum_{m \in M} \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ijskmt} \\
& \left. - \sum_t \sum_d \sum_{k \in K} V_{sk} \sum_{m \in M} \sum_{i \in V^C} \sum_{j \in V^D} c_{ij} x_{ij1kmt} \right)
\end{aligned} \tag{42}$$

Esta Função Objetivo descreve a maximização de benefícios obtidos a partir das receitas com demandas  $q_{id}$  do cliente  $i$  atendido pela loja  $d$ , descontados os custos fixos e variáveis relacionados à utilização de veículos próprios ou terceirizados de determinados tipos em determinados dias de entrega ao longo em um horizonte de planejamento  $T$ .

Observe que as receitas não são consideradas no modelo para o HMDPCOMVRP, porque não dependem da atribuição de clientes aos depósitos e, conseqüentemente, são um valor constante. Portanto, neste modelo, as restrições (35) são removidas e as restrições (43) são acrescentadas.

$$\sum_{i \in V^C} q_{id} \cdot y_{iskmtd} \leq Q_{sk}, \quad \forall s \in S, k \in K_s, m \in M_{sk}, t \in T, d \in V^D, i \neq j \tag{43}$$

As restrições (43) garantem que a carga de cada rota que parte de cada loja  $d$  não exceda a capacidade do veículo.

#### 3.1.4.1 Desigualdades válidas

Algumas desigualdades válidas foram adicionadas ao modelo a fim de melhorar o desempenho computacional. Os seguintes conjuntos de desigualdades válidas (44) e (45) foram adaptadas de Liu e Jiang (2012), que as apresentam para o COMVRP.

$$\sum_{i \in V^c} \sum_{j \in V^D} \sum_{s \in S} \sum_{k \in K_s} \sum_{m \in M_{sk}} \sum_{t \in T} \sum_{d \in D} x_{jiskmtd} \geq \sum_{i \in V^c} \sum_{j \in V} \sum_{s \in S} \sum_{k \in K_s} \sum_{m \in M_{sk}} \sum_{t \in T} \sum_{d \in D} q_{id} \cdot \frac{x_{ijskmt d}}{Q_{sk}}, \quad i \neq j \quad (44)$$

$$\sum_{i \in V^c} \sum_{j \in V^D} \sum_{s \in S} \sum_{k \in K_s} \sum_{m \in M_{sk}} \sum_{t \in T} \sum_{d \in D} x_{jiskmtd} \geq \sum_{i \in V} \sum_{j \in V} \sum_{s \in S} \sum_{k \in K_s} \sum_{m \in M_{sk}} \sum_{t \in T} \sum_{d \in D} (\vartheta \cdot c_{ij} + \tau_i) \cdot \frac{x_{ijskmt d}}{H_{sk}}, \quad i \neq j \quad (45)$$

As desigualdades (44) limitam o número de veículos, isto é, veículos particulares/próprios e alugados/terceirizados, com relação a capacidade do veículo. As desigualdades (45) limitam o número de veículos próprios e terceirizados nos aspectos de duração da viagem.

### 3.2 ABORDAGEM DE SOLUÇÃO VIA ALGORITMO HEURÍSTICO HÍBRIDO COM PARTICIONAMENTO DE CONJUNTOS (HGA-SP)

O Algoritmo Genético desenvolvido por Holland e Holland (1975) é um método de otimização bio inspirado. Simplificadamente, as características físicas dos seres vivos são codificadas geneticamente pelos cromossomos formados pelas palavras (sequências de alfabetos) de 4 tipos de proteínas (alfabetos). A sequência destes códigos caracteriza a espécie e suas características físicas de um indivíduo. Estes cromossomos são consequências de cruzamentos dos cromossomos dos pais, acrescidas de uma mutação. Pelo decorrer das gerações, os indivíduos mais aptos para o ambiente têm mais chance de sobreviver.

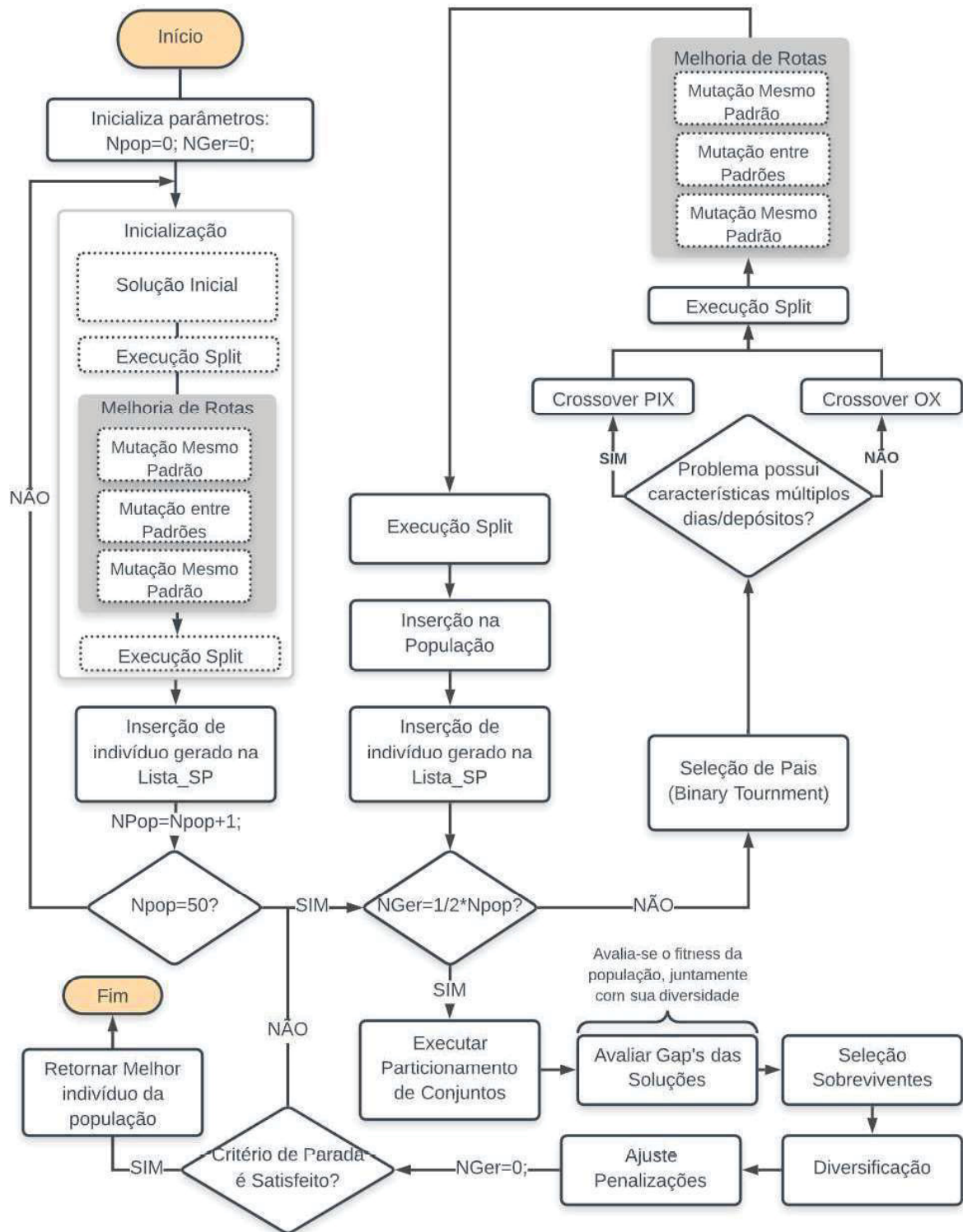
O atributo utilizado para qualificar os indivíduos e definir a sua permanência ou exclusão da população entre gerações sucessivas chama-se “Aptidão” (ou *Fitness*), normalmente associado à expressão matemática da Função Objetivo do modelo que representa o problema. Assim, esta simulação do processo evolutivo pretende produzir, à medida que as gerações forem se sucedendo, cromossomos cada vez mais bem adaptados, isto é, com melhor valor da função de adequação, de

maneira que, no final, obtém-se uma solução (cromossomo) com alto grau de adequação ao problema proposto.

O algoritmo heurístico híbrido proposto baseia-se na meta-heurística Algoritmo Genético e utiliza um método exato baseado no Problema de Particionamento de Conjuntos (*Set Partitioning - SP*), denominado por HGA-SP, do inglês, *Hybrid Genetic Algorithm with Set Partitioning*. O esquema geral deste algoritmo está exposto na FIGURA 15.



FIGURA 15 - FLUXOGRAMA DO HGA-SP



FONTE: A autora (2020).

Como se pode observar, a inicialização de uma população inicial contempla um procedimento de solução inicial seguido pelo procedimento *Split* e pelo

procedimento de melhoria de rotas para os indivíduos. Faz-se isso até que uma população com tamanho inicial de 50 indivíduos seja criada. Em seguida, inicia-se a fase de geração de soluções a partir da seleção de dois pais selecionados por *Binary Tournament*, que consiste em um método de seleção de indivíduos em que se escolhe dois indivíduos aleatórios e seleciona-se o de melhor *fitness*. Na sequência, inicia-se o Cruzamento desses pais para formação de um novo indivíduo que passa a ser reinserido na população. Todo esse procedimento é repetido até que o número de indivíduos gerados seja equivalente a  $\frac{1}{2} * Npop$ . Todas as rotas geradas por esses indivíduos são adicionadas a uma lista *Lista\_SP* para que então se inicie o procedimento de Particionamento de Conjuntos. Em seguida o cálculo dos *gap's* que avalia *fitness* e diversidade é realizado para todos os indivíduos. Para que enfim seja avaliado quais indivíduos serão sobreviventes. O processo de diversificação é realizado a cada 5 iterações sem melhoria de solução. Por fim, ajustam-se as penalizações, avalia-se o critério de parada e se não for satisfeito volta-se a etapa de geração de indivíduos, porém se satisfeito, o algoritmo retorna o melhor indivíduo da população. Cada uma dessas etapas será explicada a seguir.

### 3.2.1 O Procedimento de Solução Inicial

O procedimento utilizado para a Solução Inicial é descrito na FIGURA 16.

FIGURA 16 - ALGORITMO PARA SOLUÇÃO INICIAL

```

1. carga[p][depot]=0;
2. GrandeRota[p][depot]=0;
3. Escolha um cliente i aleatoriamente
4. while LC ≠ 0
5.     Escolha a loja d em que i adquira a maior carga  $q[i][d]$  sem ultrapassar a capacidade do maior veículo disponível ( $Q_{max}$ )
6.     Escolha o padrão de dias p aleatoriamente para alocar este cliente:
7.         GrandeRota[p][d] ← i
8.     Alocar este cliente ao veículo de maior capacidade disponível
9.     bool carga_completa = false;
10.    while (carga_completa == false e LC ≠ 0)
11.        Escolha cliente j de acordo com a proximidade de i,
12.        while (LC ≠ 0)
13.            if o cliente j pode ser adicionado ao mesmo padrão de dias
14.                if (carga[p][d] +  $q[i][d] < Q[|nveh|-1]$ )
15.                    carga[p][d] +=  $q[i][d]$ ;
16.                    GrandeRota[p][depot] ← j
17.                    Remova j de LC
18.                else
19.                    carga[p][d]=0

```

```

20.             carga_completa == true
21.         else
22.             Remova  $j$  de LC
23.              $LCN \leftarrow j$ 
24.          $i \leftarrow j$ 
25.     endwhile
26.      $LC \leftarrow LCN$ 
27. endwhile
28. end while

```

FONTE: A autora (2020).

Inicialmente, um cliente é escolhido aleatoriamente (linha 3). Forma-se uma lista de clientes LC a serem alocados. Até que todos os clientes não tenham sido alocados, escolhe-se uma loja  $d$  de acordo com a maior demanda do cliente, desde que não ultrapasse a disponibilidade de capacidade do maior veículo (linha 5). Em seguida, escolhe-se aleatoriamente um padrão de entrega dentre os padrões de entregas desejados pelo cliente (linha 6). O cliente enfim, é adicionado a Grande-Rota no padrão  $p$  e loja  $d$ , utilizando o veículo de maior carga disponível (linha 7 e 8).

Assim, enquanto a capacidade do maior veículo disponível não for ultrapassada, ou a lista de clientes a serem alocados não for esgotada, escolhe-se o cliente  $j$  de acordo com a proximidade de  $i$  (linhas 10 e 11). Enquanto LC ainda tiver clientes, eles serão testados para entrar na Grande-Rota no depósito  $d$  e padrão de dias  $p$ .

Para tanto avalia-se o padrão de dias em que o cliente  $j$  deseja ser atendido (linha 13) e a disponibilidade de carga para o uso do mesmo veículo (linha 14). Caso o veículo não comporte a carga daquele cliente, esse cliente é excluído da lista de clientes a serem alocados (LC) e alocado à uma segunda lista LCN para posterior alocação (linha 23), e  $i$  recebe o valor de  $j$  (linha 24). fecha-se o ciclo por completar a carga do veículo, então a lista LC passa a ser LCN (linha 26), e retorna-se ao início, linha 4.

### 3.2.2 O Procedimento de Melhoria de Rota

Este procedimento é adaptado de Vidal et al. (2014), e vai muito além dos conceitos clássicos do AG que propõem melhorias por meio de um processo de mutação aleatória. Afinal, são incluídos diversos algoritmos de buscas locais para o VRP.

Esta etapa foi realizada em duas fases: Melhoria de Rota e Melhoria de Padrões. Para a primeira fase, foram selecionados oito tipos de vizinhanças. Destas, seis são baseadas em movimentos  $\lambda$ -interchanges (OSMAN, 1993) considerada a troca de até  $\lambda=2$  clientes entre duas rotas. Essas trocas são melhores explicadas como duplas  $(\lambda_1, \lambda_2)$  (com  $\lambda_1 < \lambda$  e  $\lambda_2 < \lambda$ ), sendo que  $\lambda_1$  clientes são transferidos da rota  $r_1$  para a rota  $r_2$  e  $\lambda_2$  clientes da rota  $r_2$  para a rota  $r_1$ . Um vértice  $u$  selecionado aleatoriamente e um vértice próximo  $v$  (selecionado segundo um limite de granularidade ( $h$ )) são examinados em ordem aleatória. Sendo  $x$  e  $y$  os vértices subsequentes a  $u$  e  $v$  na rota. Os seguintes movimentos são utilizados:

- a) Se  $u$  é um cliente, remova-o e coloque-o depois de  $v$  (*Shift* (1; 0));
- b) Se  $u$  e  $x$  são clientes, remova-os e coloque  $u$  e  $x$  depois de  $v$  (*Shift* (2; 0));
- c) Se  $u$  e  $x$  são clientes, remova-os e, coloque  $x$  e  $y$  depois de  $v$  (*Shift* (2; 0));
- d) Se  $u$  e  $v$  são clientes, troque  $u$  e  $v$  (*Swap* (1; 1));
- e) Se  $u$ ,  $x$  e  $v$  são clientes, troque  $u$  e  $x$  por  $v$  (*Swap* (2; 1));
- f) Se  $u$ ,  $x$ ,  $v$  e  $y$  são clientes, troque  $u$  e  $x$  com  $v$  e  $y$  (*Swap* (2; 2));
- g) Se  $r(u) = r(v)$ , substitua  $(u, x)$  e  $(v, y)$  por  $(u, v)$  e  $(x, y)$  (*2-Opt* intra-rota);  
Se  $r(u) \neq r(v)$ , substitua  $(u, x)$  e  $(v, y)$  por  $(u, y)$  e  $(x, v)$  (*2-Opt* inter-rota).

Essas vizinhanças foram implementadas usando um procedimento *Randomized Variable Neighborhood Descent* (RVND) (SUBRAMANIAN et al., 2010). Assim, ao invés de usar uma ordem pré-definida de vizinhanças para explorar o espaço de soluções, utiliza-se uma ordem aleatória. Mais especificamente, sempre que em uma determinada vizinhança não for possível melhorar a solução corrente, o RVND seleciona, aleatoriamente, outra vizinhança para continuar a busca pelo espaço de soluções. Esta abordagem, segundo Penna; Subramanian; Ochi (2013) produz, em média, resultados melhores do que a versão com ordem determinística das vizinhanças e tem vantagem de não necessitar de um estudo de qual a melhor ordem de vizinhanças.

A fase de mutação em um mesmo padrão tem fim quando todos os movimentos possíveis são testados sucessivamente sem sucesso. Já para a fase de mutação entre padrões, avalia-se para cada cliente selecionado em ordem aleatória,

o custo mínimo para satisfazer os requisitos de visita  $i$  do depósito  $d$  de acordo com o padrão de visita  $pat$ . Se existir uma combinação de menor custo todas as visitas ao cliente  $i$  serão removidas e uma nova visita será inserida na melhor localização em cada sequência correspondente ao depósito e ao período. O procedimento para quando todos os clientes forem sucessivamente considerados sem nenhuma modificação.

### 3.2.3 Espaço de busca

Uma característica relevante do espaço de busca vem do potencial uso de soluções inactíveis (CORDEAU; LAPORTE; MERCIER, 2001; NAGATA; BRÄYSY; DULLAERT, 2010; VIDAL et al., 2012). Estudos realizados desde o trabalho de Glover (1986) relatam que uma exploração controlada de soluções inactíveis pode melhorar a busca, permitindo que ela faça a transição mais facilmente entre soluções viáveis estruturalmente diferentes. Além disso, o uso de soluções inactíveis pode contribuir para melhorar a robustez do método, que é menos dependente da disponibilidade de uma solução inicial viável (encontrar uma solução viável é muitas vezes um problema NP-difícil). O HGA-SP define para a avaliação de um indivíduo  $i$  o custo penalizado  $\phi^{COST}(i)$ , como sendo o somatório dos custos de cada rota ( $\phi(r)$ ). A equação (46) apresenta o custo de rota para o modelo de minimização e na equação (47) para a versão de maximização de benefícios.

$$\phi(r) = CF_k + CV * \varphi^{dist}(r) + \max\left\{\left(H - \varphi^{dur}(r)\right), 0\right\} * \omega^{Dur} + \max\left\{\left(Q - \varphi^q(r)\right), 0\right\} * \omega^Q \quad (46)$$

$$\phi(r) = \varphi^q(r) - \left( CF_k + CV * \varphi^{dist}(r) + \max\left\{\left(H - \varphi^{dur}(r)\right), 0\right\} * \omega^{Dur} + \max\left\{0, \left(Q - \varphi^q(r)\right)\right\} * \omega^Q \right) \quad (47)$$

A equação (48) compreende o custo fixo associado à utilização de um veículo  $k$  ( $CF_k$ ) e um custo variável ( $CV$ ) de acordo com a distância percorrida ( $\varphi^{dist}(r)$ ). Para o excesso de duração ( $H - \varphi^{dur}(r)$ ) existe uma penalização  $\omega^{Dur}$  e para o excesso de carga ( $Q - \varphi^q(r)$ ), existem uma penalização  $\omega^Q$ . Para finalizar a avaliação da solução  $sol$  ainda considera-se uma penalização por se utilizar um número de veículos maior do que o especificado.

$$\phi^{COST}(sol) = \sum \phi(r) + \max\{(|K| - n_{veh}), 0\} \quad (48)$$

O ajuste das penalizações é dinâmico durante a execução do algoritmo para favorecer a geração de indivíduos factíveis. Assim, seguindo a ideia apresentada em Vidal et al. (2012) para os problemas MDVRP e PVRP, fez-se aqui uma adaptação deste ajuste para o HGA-SP. Seja  $\xi^{REF}$  a proporção alvo de indivíduos factíveis,  $\xi^Q$  e  $\xi^D$  a proporção dos 10 últimos indivíduos factíveis em relação à capacidade do veículo e à duração da rota, respectivamente. O ajuste é descrito nas equações (49) e (50).

$$\begin{cases} \omega^D = \omega^D \times 1,2 & \text{se } \xi^D \leq \xi^{REF} - 0,05 \\ \omega^D = \omega^D \times 0,85 & \text{se } \xi^D \geq \xi^{REF} + 0,05 \end{cases} \quad (49)$$

$$\begin{cases} \omega^Q = \omega^Q \times 1,2 & \text{se } \xi^Q \leq \xi^{REF} - 0,05 \\ \omega^Q = \omega^Q \times 0,85 & \text{se } \xi^Q \geq \xi^{REF} + 0,05 \end{cases} \quad (50)$$

Os valores assumidos para  $\omega^D$  e  $\omega^Q$  iniciam com valor igual a 2, conforme sugere Vidal et al. (2012).

### 3.2.4 Gestão da População

Para aplicar o procedimento de seleção dos sobreviventes, as soluções são avaliadas não apenas em relação aos custos ou benefícios ( $\phi^{COST}(i)$ ) de seus indivíduos, mas também em relação à sua diversidade medida conforme propõem Vidal et al. (2014). Os mesmos autores foram os primeiros a utilizar a ideia de aliar *fitness* e diversidade em um AG. Neste trabalho, a proposta de aliar a diversidade é apresentada, porém, utilizando-se os *gaps* entre a solução obtida com a melhor solução encontrada. A normalização dos *gaps* é feita para que os valores fiquem no intervalo [0,1]. O escalonamento para o intervalo é realizado dividindo-se o *gap* de cada solução pelo seu valor máximo, conforme apresentam-se as equações (51) e (52).

$$GAP_{Fitness}(sol) = \frac{Fitness(sol)}{Fitness(sol) - Fitness(sol^*)} \quad (51)$$

$$Gap(sol) = \frac{GAP_{Fitness}(sol)}{\max_{sol_i \in S} (GAP_{Fitness}(sol_i))} + fd \frac{GAP_{Fitness}(sol)}{\max_{sol_i \in S} (GAP_{Fitness}(sol_i))} \quad (52)$$

O processo de diversificação é realizado de modo a eliminar 1/3 dos piores indivíduos, criando-se mais 50 indivíduos novos. Os sobreviventes são mantidos de

modo que a população mantenha seu tamanho inicial, e são avaliados conforme a distância de *Hamming* ou de *Broken\_pairs* – no caso de não haver características de periodicidade e múltiplos depósitos – conforme sugere Vidal (2014).

### 3.2.5 Representação da Solução

A representação de solução é composta apenas por clientes a serem atendidos em cada rota, não contendo delimitadores de rotas. Essa estratégia foi apresentada de forma eficiente por Prins (2004), que acabou inserindo os algoritmos genéticos no grupo das melhores heurísticas para solução de problemas de roteamento de veículos para as instâncias de Christofides (1979).

No presente trabalho utiliza-se a representação denominada de Grandes-rotas de Prins (2004). Essa solução é avaliada usando um procedimento *Split* que divide idealmente a sequência em rotas viáveis (Seção 3.2.6). A codificação do cromossomo sem delimitadores de rotas utilizada permite a utilização do operador de crossover OX (Seção 2.2.2.2 para uma descrição). Para este trabalho, além do OX, implementou-se o PIX de Vidal et al. (2013a), pois este considera a possibilidade de múltiplos depósitos e o periodicidade.

### 3.2.6 Procedimento *Split*

Apresenta-se o procedimento *Split* com a generalização utilizada para o HMDPCOMVRP. Inicialmente para o VFMP, no qual a frota é composta por  $a_k$  tipos de veículos, considerados de forma conjunta entre frota própria e terceirizada, tal que cada tipo de veículo  $k \in \{1, \dots, |K|\}$ , tal que  $|K| = |K_{s=0}| + |K_{s=1}|$  é definido por uma capacidade  $Q_k$ , um custo fixo  $CF_k$  e um custo por unidade de distância  $CV_k$ , chamado de custo variável. Um percurso de comprimento  $L$  feito por um veículo do tipo  $k$  tem um custo  $CF_k + L * CV_k$ . O objetivo é calcular um conjunto de viagens e atribuir veículos a viagens para minimizar o custo total. Um arco  $(i_1, i_j)$  é incluído em  $A$  para cada subsequência  $(S_i, S_{i+1}, \dots, S_j)$  de  $S$  se uma viagem servindo a esses clientes for viável, isto é, se a demanda total  $W_{ij}$  da subsequência (carga de viagem) não excede a capacidade do veículo como em (84).

$$W_{ij} = \sum_{k=i}^j q(S_k) \leq Q \quad (53)$$

A carga  $Q_{ij}$  e a duração  $L_{ij}$  de cada viagem pode ser determinada em  $O(1)$ . A viagem é factível se ela pode ser realizada pelo menos com o veículo de maior capacidade, ou seja,  $W_{ij} \leq Q_{max}$ . Seja  $Z_{ijk}$  o custo da viagem utilizando o veículo  $k$ . O custo  $Z_{ij}$  para o grafo auxiliar corresponde ao custo mais barato dentre os veículos compatíveis.

$$Z_{ij} = \min(Z_{ijk} = CF_k + CV_k L_{ij} : k \in K, W_{ij} \leq Q_{max}) \tag{54}$$

Utiliza-se um vetor  $B$  indexado de 1 até capacidade do maior veículo, de tal forma que  $B(W)$  fornece o melhor tipo de veículo para uma viagem com carga  $W$ . Para exemplificar numericamente, suponha uma viagem com demanda de 35 e duração de 200, utilizando-se frota própria. O cálculo do custo dessa rota assim como os parâmetros disponíveis estão na FIGURA 17.

FIGURA 17 - EXEMPLO DO CÁLCULO DE CUSTO PARA O VFMP

Tipo do veículo ( $k$ )	1	2	3	4	Tem-se que: $B(35)=3$ . Os veículos 1 e 2 teriam suas capacidades violadas. Assim, o custo dessa viagem seria: $Z = 250 + 30 * 200 = 6250$	W	B(W)
Disponibilidade ( $a_k$ )	$n$	$n$	$n$	$n$		1	1
Capacidade ( $Q_k$ )	20	30	40	50		2	1
Custo Fixo ( $CF_k$ )	100	200	250	300		:	:
Custo Variável ( $CV_k$ )	10	20	30	35		35	3
					:	:	
					50	4	

FONTE: A autora (2020).

No HVRP, o número de veículos é limitado e o problema pode ser infactível. Note que os dados precisam satisfazer (55), que afirma que o total de demanda dos clientes não deve ser maior que a capacidade total destes veículos. Assume-se que os tipos de veículos estão ordenados de acordo com suas capacidades.

$$\sum_{i=1}^n q_i \leq \sum_{k \in K} a_k Q_k \tag{55}$$

A FIGURA 18 mostra um caso de duas soluções com os mesmos clientes. Suponha uma frota com apenas dois veículos com capacidade de 10 e 14. Em  $S'$  não existe uma divisão de rota factível, embora a restrição (55) seja satisfeita. Segundo Prins (2009), esses casos são raros na prática, exceto nos casos em que a demanda total estiver muito próxima da capacidade da frota.

FIGURA 18 - ILUSTRAÇÃO DE SOLUÇÕES FACTÍVEIS E INFACÍVEIS

Seja $k = 2, a_1 = 1, Q_1 = 10, a_2 = 1, Q_2 = 14$	$S'$ : <span style="border: 1px solid black; padding: 2px;">1 4 3 5 2</span> $\rightarrow$ <i>infactível</i>
--	--



$S$ : <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">5</td></tr></table>	1	2	3	4	5	$\rightarrow$ <i>factível</i>	$q$ : <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="padding: 2px 5px;">5</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">7</td><td style="padding: 2px 5px;">4</td></tr></table>	5	2	4	7	4
1	2	3	4	5								
5	2	4	7	4								
$q$ : <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="padding: 2px 5px;">5</td><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">7</td></tr></table>	5	4	4	2	7							
5	4	4	2	7								

FONTE: A autora (2020).

O gráfico auxiliar  $\mathcal{H}$  para o HVRP é o mesmo que para o VFMP, com um arco para cada viagem  $(C_i, C_{i+1}, \dots, C_j)$ , tal que  $W_{ij} \leq Q_k$ . A única diferença é que não se sabe de antemão, devido a disponibilidades limitadas, se um veículo do tipo mais barato estará disponível. Portanto, o comprimento de viagem  $L_{ij}$  é usado como peso provisório para o arco em  $\mathcal{H}$ : o custo exato dependerá do tipo de veículo alocado para este arco.

Uma divisão ótima de rota corresponde ao caminho de menor custo do nó 0 para  $n$  em  $\mathcal{H}$ , com não mais do que  $a_k$  arcos para cada tipo de veículo  $k$ . Nesse problema do caminho mínimo com restrições de recursos cada tipo de veículo pode ser visto como um recurso disponível em unidades  $a_k$  e cada arco requer uma unidade de um tipo de veículo compatível. Em geral, tais problemas são NP-difícil, mas podem ser resolvidos rapidamente na prática, usando métodos de programação dinâmica. O número máximo de vetores distintos ( $\psi$ ), levando em conta o consumo nulo está apresentado em (56).

$$\psi = \prod_{k=1}^{|K|} (a_k + 1) \quad (56)$$

Os cálculos podem ser organizados usando uma matriz  $Y$ ,  $(n + 1) * \psi$ . Cada linha  $j$  de  $Y$  é indexada de 0 a  $\psi - 1$ . Na prática, um valor  $Y_{jh}$  para  $j$  representa um caminho mais curto do nó 0 para o nó  $j$  no grafo auxiliar.  $Y$  contém informações como o custo  $Y_{jh} \cdot Z$ , e o padrão de utilização parcial de frota para o caminho  $Y_{j,h} \cdot p$ . Ao contrário do VRP e do VFMP, vários padrões podem ser usados para o mesmo nó, com diferentes consumos de recursos.

Definindo-se uma chave  $h(x)$  para converter cada vetor possível  $X$ . Essa chave pode ser calculada pelo algoritmo da FIGURA 19 conforme sugere Prins (2009).

FIGURA 19 - ALGORITMO PARA CALCULAR A CHAVE REFERENTE A UM VETOR  $X$ 

01.	$h := 0$
02.	<b>for</b> $k:=1$ <b>to</b> $ K $ <b>do</b> $h := h(a_k + 1) + x_k$

FONTE: Adaptado de Prins (2009).

Na prática, o algoritmo pode ser usado (FIGURA 20), por exemplo, assumindo-se frota própria, se  $|K|=4$ , e  $a_k = (4,2,3,1)$ , tem-se  $\psi = 5 * 3 * 4 * 2 = 120$ .

FIGURA 20 - EXEMPLO NUMÉRICO PARA O CÁLCULO DA CHAVE REFERENTE A UM VETOR  $X$ 

Para calcular a chave de $X = (4,2,3,1)$ , faz-se: $h := h(a_k + 1) + x_k$ $h(0,0,0) = 0;$ $h := 0 * (4 + 1) + 4 = 4;$ $h := 4 * (2 + 1) + 2 = 14;$ $h := 14 * (3 + 1) + 3 = 59;$ $h := 59 * (1 + 1) + 1 = \mathbf{119};$  Para calcular a chave de $X = (1,0,0,0)$ , faz-se: $h := h(a_k + 1) + x_k$ $h(0,0,0) = 0;$ $h := 0 * (4 + 1) + 1 = 1;$ $h := 1 * (2 + 1) + 0 = 3;$ $h := 3 * (3 + 1) + 0 = 12;$ $h := 12 * (1 + 1) + 0 = \mathbf{24};$		$k$				$h(x)$
		1	2	3	4	
	$X$	0	0	0	0	0
		0	0	0	1	1
		0	0	1	0	2
		0	0	1	1	3
		0	0	2	0	4
		0	0	2	1	5
		0	0	3	0	6
		0	0	3	1	7
		⋮	⋮	⋮	⋮	⋮
		1	0	0	0	<b>24</b>
		⋮	⋮	⋮	⋮	⋮
		4	2	3	1	$\psi - 1 = \mathbf{119}$

FONTE: A autora (2020).

O algoritmo apresentado na FIGURA 21 fornece um algoritmo para implementação do método, utilizando uma matriz  $Y$  e um grafo auxiliar  $\mathcal{H} = (X, A, Z)$ , embora,  $\mathcal{H}$  não seja gerado explicitamente.  $X$  contém  $n + 1$  nós indexados de 0 a  $n$ .  $A$  contém um arco  $(i, j)$ ,  $i < j$ , se a sequência de clientes  $C_{i+1}$  para  $C_j$  apresenta-se dentro de limites aceitáveis em termos de carga e duração, verificado na linha 4. Basicamente, para cada nó  $j = 1, 2, \dots, n$ , armazenam-se os custos do caminho mínimo do nó 0 para o nó  $j$  em  $\mathcal{H}$ , e o predecessor de  $j$  neste caminho.

Todos os valores de  $Y$  são inicializados na linha 1: os custos na linha 0 são definidos como 0 e os outros custos como infinito. No final, a solução é dada pelo valor mais barato na linha  $n$  de  $Y$ . Se todos os valores da linha  $n$  ainda tiverem um custo infinito, o problema é infactível.

FIGURA 21 - ALGORITMO PARA O PROBLEMA DO CAMINHO MÍNIMO COM RESTRIÇÃO DE RECURSOS PARA HVRP

```

1.      Inicializar matriz Y
2.      for i:=1 to n do
3.           $W_{ij} := 0; L_{ij} := 0; j := i$ 
4.          while (j <= n &&  $W_{ij} < Q_{max} * \delta_Q$ )
5.               $W_{ij} := W_{ij} + q_{C_j}$  //cálculo da carga da rota
6.              if  $W_{ij} < Q * \delta_Q$ 
7.                  if  $i = j$  then  $L_{ij} := d_{0,C_j} + t_{C_j} + d_{C_j,0}$  //cálculo da duração da rota
8.                  else  $L_{ij} := L_{ij} - d_{C_{j-1},0} + d_{C_{j-1},C_j} + t_{C_j} + d_{C_j,0}$ 
9.                  if  $L_{ij} < L * \delta_D$  //  $C_i \dots C_j$  correspondem ao arco  $(i - 1, j)$  em  $\mathcal{H}$ 
10.                     for p:=0 to  $\psi - 1$  with  $Y_{ip}, Z = \omega$  do
11.                          $k = k | Q_k = Q_{max}$  //a avaliação dos veículos começa pelo tipo
12.                         de veículo de maior capacidade
13.                         capacity_exceeded:=false
14.                         while  $k > 0$  && capacity_exceeded==false
15.                             if  $Y_{p,k} + 1 \leq a_k$  && ( $W_{ij} < Q_k$ ) //verificação da disponibilidade de
16.                             veículos do tipo  $k$  e carga
17.                                  $Z_{ij} = CF_k + CV_k * L_{ij} + \max(0, Q_k - W_{ij}) * \omega_Q + \max(0, L_k - L_{ij}) * \omega_D$  // cálculo do custo rota considerando penalizações
18.                                  $h = 0$ , with  $i = 1$ 
19.                                  $h = p$ 
20.                                  $h_{new} = h + \pi_k$  //cálculo da nova chave para o consumo X
21.                                 if  $Z_{ij} + Y_{i-1,p} * Z < Y_{j,h_{new}} * Z$ 
22.                                      $Y_{j,h_{new}} * Z = Z + Y_{i-1,p} * Z$ 
23.                                      $Y_{j,h_{new}} * p = p$ 
24.                                      $Y_{j,h_{new}} * i = i - 1$ 
25.                                      $Y_{j,h_{new}} * k = k$ 
26.                                 endif
27.                             else capacity_exceeded=true;
28.                                 k=k-1
29.                             end while
30.                         end for
31.                     endif
32.                 j=j+1;
33.             endwhile
34.         endfor

```

FONTE: A autora (2020).

Na linha 3, inicializa-se valores de carga e distâncias. A partir de  $j = i$ , todos os clientes  $C_j$  são avaliados para compor a rota desde que a carga total da rota não ultrapasse a capacidade do maior veículo ( $Q_{max}$ ), conforme linha 4. Em seguida, calcula-se a nova carga da rota considerando-se o novo cliente  $j$  a ser adicionado e avalia-se a permissibilidade de carga (linhas 5 e 6). Se essa carga for menor que a permitida ( $Q * \delta_Q$ ), calcula-se a duração da rota  $L_{ij}$  (linhas 7 e 8). Se a duração dessa rota estiver dentro do limite permitido ( $L * \delta_D$ ), tem-se um arco  $(i - 1, j)$  em  $\mathcal{H}$ , sendo

que este arco possui os clientes  $C_i \dots C_j$ . Assim, para cada arco  $(i - 1, j)$  de  $\mathcal{H}$  são inspecionados cada valor de  $Y_{ip}$  (linha 10). Também se verifica para cada tipo de veículo a ser utilizado a sua disponibilidade  $(a_k)$ , e se excede ou não a capacidade do veículo  $k$ . Para isso, inicia-se a avaliação por  $k$  assumindo os valores do veículo de maior carga. Assim, a partir do primeiro veículo com capacidade inferior à carga necessária para a rota, os veículos com capacidades ainda menores nem chegam a ser testados.

Na linha 15, o custo da rota é calculado considerando-se as penalizações, conforme mencionado (Seção 3.2.3). Em seguida, realiza-se o cálculo da nova chave para o consumo  $X$ . De fato, a linha 17 pode ser computada em  $O(1)$ , se apenas adicionarmos a contribuição  $\pi_k$  em cada chave associada ao acréscimo do veículo  $k$  adicionado a  $x_k$ . Essa contribuição pode ser calculada como em (57).

$$\pi_k = \prod_{i=k+1}^{|K|} (a_i + 1), \quad k \neq |K| \quad \pi_{k=|K|} = 1 \quad (57)$$

Para exemplificar, apresenta-se a FIGURA 22, na qual supõe-se que 4 tipos de veículos ( $|K| = 4$ ) estão disponíveis. Partindo da chave inicial  $h = 0$ , com  $x = (0,0,0,0)$ , e supondo os seguintes acréscimos de um veículo de cada tipo  $k$ , com  $a_k = (4,2,3,1)$ .

FIGURA 22 – EXEMPLO NUMÉRICO DO CÁLCULO DA CONTRIBUIÇÃO  $\pi_k$  PARA A CHAVE ASSOCIADA AO ACRÉSCIMO DO VEÍCULO  $k$

$k = 1: x = (1,0,0,0)$ $\pi_1 = \prod_{i=2}^4 (a_i + 1) = (a_2 + 1) * (a_3 + 1) * (a_4 + 1) = 3 * 4 * 2 = 24;$ $h: = 0 * (4 + 1) + 1 = 1;$ $h: = 1 * (2 + 1) + 0 = 3;$ $h: = 3 * (3 + 1) + 0 = 12;$ $h: = 12 * (1 + 1) + 0 = 24;$ $h: = 0 + 24 = 24;$	$k = 3: x = (0,0,1,0)$ $\pi_3 = \prod_{i=4}^4 (a_i + 1) = 2;$ $h: = 0 * (4 + 1) + 0 = 0;$ $h: = 0 * (2 + 1) + 0 = 0;$ $h: = 0 * (3 + 1) + 1 = 1;$ $h: = 1 * (1 + 1) + 0 = 2;$ $h: = 0 + 2 = 2;$	<table border="1"> <thead> <tr> <th><b>k</b></th> <th><b><math>\pi_k</math></b></th> </tr> </thead> <tbody> <tr> <td><b>1</b></td> <td><b>24</b></td> </tr> <tr> <td><b>2</b></td> <td><b>8</b></td> </tr> <tr> <td><b>3</b></td> <td><b>2</b></td> </tr> <tr> <td><b>4</b></td> <td><b>1</b></td> </tr> </tbody> </table>	<b>k</b>	<b><math>\pi_k</math></b>	<b>1</b>	<b>24</b>	<b>2</b>	<b>8</b>	<b>3</b>	<b>2</b>	<b>4</b>	<b>1</b>
<b>k</b>	<b><math>\pi_k</math></b>											
<b>1</b>	<b>24</b>											
<b>2</b>	<b>8</b>											
<b>3</b>	<b>2</b>											
<b>4</b>	<b>1</b>											
$k = 2: x = (0,1,0,0)$ $\pi_2 = \prod_{i=3}^4 (a_i + 1) = 8;$ $h: = 0 * (4 + 1) + 0 = 0;$ $h: = 0 * (2 + 1) + 1 = 1;$ $h: = 1 * (3 + 1) + 0 = 4;$ $h: = 4 * (1 + 1) + 0 = 8;$ $h: = 0 + 8 = 8;$	$k = t = 4: x = (0,0,0,1):$ $h: = h(a_k + 1) + x_k$ $h: = 0 * (4 + 1) + 0 = 0;$ $h: = 0 * (2 + 1) + 0 = 0;$ $h: = 0 * (3 + 1) + 0 = 0;$ $h: = 0 * (1 + 1) + 1 = 1;$ $h: = 0 + 1 = 1;$	<p>Assim, ao se adicionar um veículo do tipo <math>k=1</math>, precisa-se adicionar à chave 24 unidades.</p>										

FONTE: A autora (2020).

Na linha 18, avalia-se o custo calculado  $Z_{ij}$  somado aos custos dos arcos anteriores a  $i$  em relação a  $Y_{j,h_{new}} \cdot Z$ . Os valores em  $Y$  são alterados em caso de melhoria no custo (linhas 19 a 22). A FIGURA 23 apresenta um algoritmo para extrair rotas de  $Y$ . O procedimento *enqueue* adiciona cada cliente no final da viagem.

FIGURA 23 - ALGORITMO PARA EXTRAIR AS ROTAS

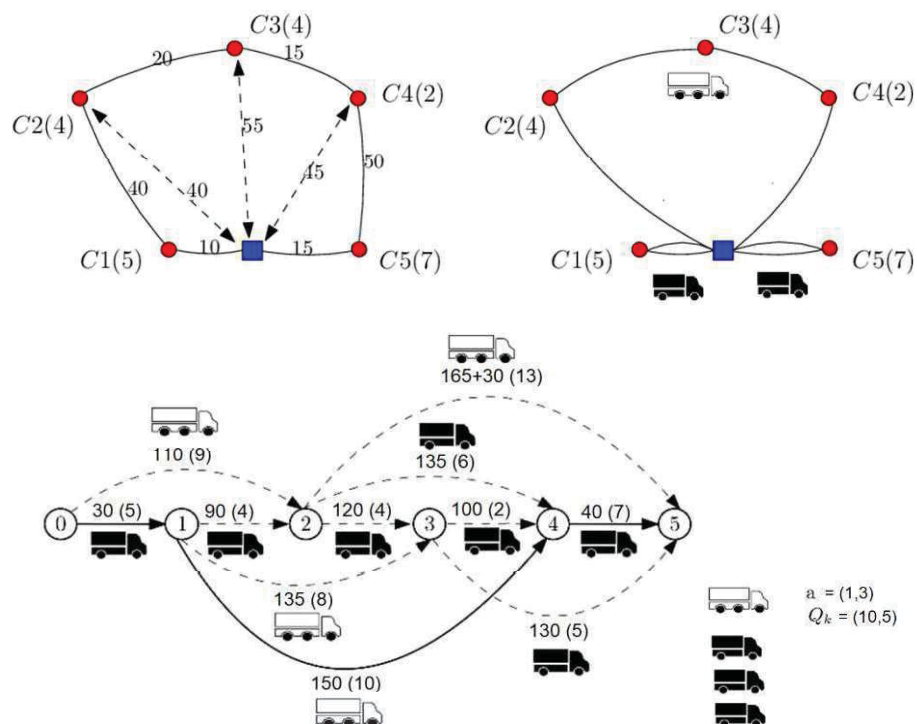
```

1.   $i = n$ 
2.  for  $p:=0$  to  $\psi - 1$  with  $Y_{ip} \cdot Z = \infty$  do
3.     $Y_{aux} := \min(Y_{ip}, Z)$ 
4.     $k = Y_{aux} \cdot k$ 
5.     $t := 0$ 
6.  Repeat
7.     $t:=t+1$ 
8.     $i = Y_{aux} \cdot i$ 
9.     $p = Y_{aux} \cdot p$ 
10.  for  $k:=i$  to  $j$  do enqueue(trip( $t$ ),  $S_k$ ) endfor
11.   $Y_{aux} = Y_{ip}$ 
12.   $j := i$ 
13. until  $i=0$ 

```

FONTE: A autora (2020).

A FIGURA 24 ilustra a aplicação algoritmo *Split* para uma sequência de 5 clientes  $S_1$  até  $S_5$ . O primeiro grafo mostra os custos dos arcos e as demandas de cada cliente em parênteses. Todos os clientes têm um tempo de serviço idêntico  $\tau_i=10$ , e os parâmetros de penalização são  $\omega_Q=10$ . Considerando-se  $L = \infty$ , e quatro veículos disponíveis, sendo que um possui carga de  $Q_1 = 10$  e três possuem carga de  $Q_2 = 5$ .

FIGURA 24 - ALGORITMO PARA O PROCEDIMENTO *SPLIT* COM CARGA HETEROGÊNEA

FONTE: A autora (2020).

Os arcos do grafo  $\mathcal{H}$  são apresentados e o custo da rota incluindo penalidades. Por exemplo, a rota atendendo clientes  $C_3$ ,  $C_4$  e  $C_5$  tem uso do veículo  $k = 1$ , um custo de  $165 + 30$ , a penalização de 30 correspondente ao excesso de carga de três unidades. A solução ideal do problema é 170 de distância e duração de 260 ( $30+40+150$ ) considerando-se os tempos de serviços. Ela é composta das três rotas seguintes: rota 1 visita  $C_1$ , rota 2 com as visitas  $C_2$ ,  $C_3$  e  $C_4$  e rota 3 visita  $C_5$ .

### 3.2.7 Particionamento de Conjuntos

O procedimento de Particionamento de conjuntos é executado utilizando-se a lista de rotas armazenadas em  $Lista\_SP$ . Esta lista é composta por todas as rotas geradas a cada iteração. Uma  $Lista\_SP\_auxiliar$  armazena as melhores rotas obtidas a cada iteração. Assim, se o procedimento de Particionamento de Conjuntos demora em determinada iteração um tempo superior a um minuto, a  $Lista\_SP$  é apagada e ela passa a assumir as rotas da  $Lista\_SP\_Auxiliar$ . A TABELA 7 apresenta as notações utilizadas.

TABELA 7 – NOTAÇÕES USADAS NO PARTICIONAMENTO DE CONJUNTOS

Conjuntos e Índices	
$V$	Todos os vértices; $V = V^D \cup V^C$
$V^D \in V$	Depósitos; $d \in \{1, \dots,  V^D \}$
$V^C \in V$	Clientes; $i \in \{1, \dots,  V^C \}$
$T$	Dias do horizonte de planejamento, $t \in \{1, \dots,  T \}$
$r \in R$	Rotas possíveis para cada dia e depósito; $r \in \{1, \dots,  R \}$
$R_i \in R$	Rotas que pertencem ao cliente $i$ ;
$p \in P$	Padrões de dias; $pat \in \{1, \dots,  P \}$
$S$	Tipo de frota própria (0) ou terceirizada (1); $s \in \{0,1\}$
$K_s$	Tipos de Veículos; $k \in \{1, \dots,  K_s \}$
Parâmetros	
$Z_r$	Custo associada à rota $r$
$T$	Número total de dias no horizonte de planejamento
$f_i$	Frequência de atendimento do cliente $i$ ao longo de $T$
$m$	Número de veículos disponíveis em cada dia, cada um com capacidade $Q_k$
$n$	Total de números de clientes para visitar
$C_i$	O conjunto de dias válidos para cada cliente $i \in \{1, \dots, n\}$
Variáveis	
$Y_r$	1, se a rota $r \in R$ estiver na solução; 0, caso contrário.
$p_{i,pat}$	1 se o cliente $i$ estiver designado ao padrão de dias $pat$ na solução; 0, caso contrário.
$\theta_{i,d}$	1, se o cliente $i$ estiver designado ao depósito $d$ na solução; 0, caso contrário.
$\Omega_{i,t,d}$	1, se o cliente $i$ estiver designado ao dia $t$ e depósito $d$ na solução; 0, caso contrário.

FONTE: A autora (2020).

A formulação do PPC para o problema é implementada como segue.

$$\text{Min} \sum_{r \in R} Y_r * Z_r \quad (58)$$

$$\sum_{r \in R_i} Y_r = \Omega_{i,t,d}, \quad \forall i \in V^C, t \in T, d \in V^D \quad (59)$$

$$\sum_{t \in T} \sum_{d \in V^D} \Omega_{i,t,d} \geq f_i * p_{i,pat}, \quad \forall i \in V^C, pat \in P \quad (60)$$

$$\sum_{t \in T} \sum_{d \in V^D} \Omega_{i,t,d} \leq f_i, \quad \forall i \in V^C, pat \in P \quad (61)$$

$$\sum_{pat \in P} p_{i,pat} = 1, \quad \forall i \in V^C \quad (62)$$

$$\sum_{t \in T} \Omega_{i,t,d} = f_i * \theta_{i,depot}, \quad \forall i \in V^C, d \in V^D \quad (63)$$

$$\sum_{d \in V^D} \theta_{i,d} = 1, \quad \forall i \in V^C \quad (64)$$

$$\sum_{r \in R} Y_r \leq m_{sk}, \quad \forall s \in S, k \in K_s \quad (65)$$

$$Y_r \in \{0,1\}, \quad r \in R \quad (66)$$

$$\Omega_{i,t,d} \in \{0,1\}, \quad i \in V^C, t \in T, d \in V^D \quad (67)$$

$$p_{i,pat} \in \{0,1\}, \quad i \in V^C, pat \in P \quad (68)$$

$$\theta_{i,t} \in \{0,1\}, \quad i \in V^C, d \in V^D \quad (69)$$

A Função Objetivo (58) minimiza a soma dos custos escolhendo a melhor combinação das rotas. As restrições (59) garantem que uma única rota do subconjunto  $R_i$ , esteja associada ao cliente  $i$ , caso o mesmo esteja designado ao dia  $t$  e depósito  $d$  na solução. As restrições (60) determinam para todo cliente  $i$  e para o padrão de dias  $pat$  que ele pertença, que as variáveis  $\Omega_{i,t,d}$  assumam o valor de 1 para todos os dias  $t$  do padrão de dias e para algum depósito  $d$  selecionados para  $i$ , no mínimo igual a sua frequência, caso o padrão seja selecionado. As restrições (61) garantem que a frequência de visitas não seja ultrapassada para todo cliente  $i$ . As restrições (62) garantem que cada cliente  $i$  seja atendido em apenas um padrão de dias  $pat$ . As restrições (63) determinam para cada cliente  $i$  e para cada depósito  $d$ , que as variáveis  $\Omega_{i,t,d}$  assumam o valor de 1 caso o depósito  $d$  seja selecionado para  $i$  garantindo que o cliente  $i$  seja atendido na frequência de dias desejada. As restrições (64) garantem que cada cliente  $i$  esteja associado a apenas um depósito  $d$ . As restrições (65) garantem que o número de veículos da frota  $s$  do tipo  $k$  seja menor que o número de veículos disponível. As restrições (66) a (69) definem o domínio das variáveis de decisão.



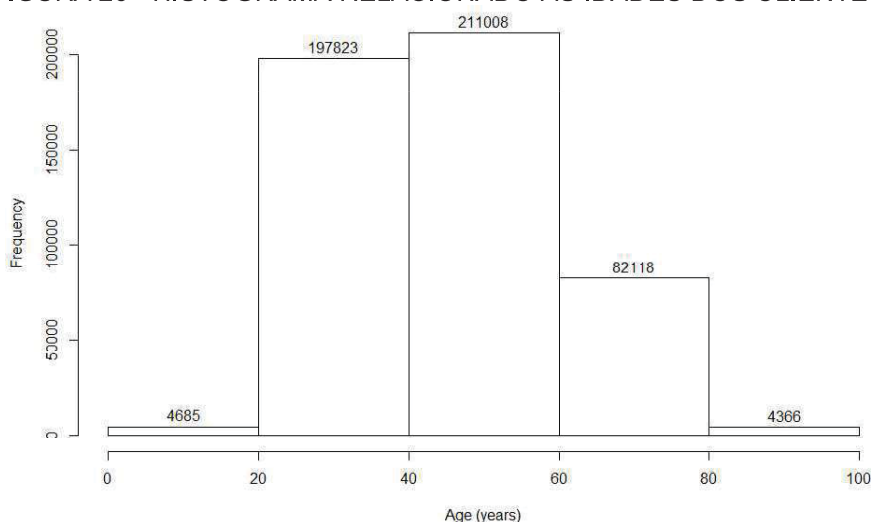
## 4 RESULTADOS

Este capítulo está dividido em duas seções. Na Seção 4.1 apresenta-se uma descrição da utilização de cada uma das técnicas do KDD, juntamente com a apresentação dos devidos testes estatísticos realizados, assim como, os resultados obtidos. A Seção 4.2 apresenta os resultados computacionais comparativos das abordagens de solução exata e heurística do problema de roteirização.

### 4.1 PROCESSO KDD

Para a análise, foram considerados inicialmente 500.000 eventos de compras, realizadas ao longo de três meses em 22 lojas do varejo, contemplando 322.000 clientes diferentes. A amostra é composta por eventos de compras realizados em sua maioria por mulheres (52,10%). Em relação às idades 39,56% dos clientes tem idade entre 20 e 40 anos, 42,20% apresentam idade entre 40 e 60 anos. Sendo que 17,297% dos clientes tem mais de 60 anos, conforme pode-se verificar no histograma da FIGURA 25.

FIGURA 25 - HISTOGRAMA RELACIONADO ÀS IDADES DOS CLIENTES



FONTE: A autora (2020).

Também são apresentadas medidas de posição e dispersão para cada uma das variáveis relacionadas às classes de itens de consumo, conforme apresentado no Anexo 1. A seguir, técnicas de agrupamento são apresentadas na Seção 4.1.1, por

meio da Análise Fatorial para a redução de dados por meio da combinação de variáveis, e da Análise de Clusterização, para o agrupamento das observações relacionadas ao perfil de lojas. Na Seção 4.1.2 apresentam-se as técnicas preditivas RLM e RNA, comparativamente, para a criação de modelos que expliquem o gasto gerado pelos clientes. Nas seções subsequentes, tem-se uma descrição mais detalhada das técnicas utilizadas nestas etapas.

#### 4.1.1 Técnicas de Agrupamento

Nesta seção são apresentados os resultados das aplicações das técnicas ACP e AC.

##### 4.1.1.1 Análise de Componentes Principais (ACP)

A ACP foi utilizada com o propósito de reduzir o número de variáveis relacionadas a 36 conjuntos de itens de consumo para cada evento de compra. Estes conjuntos referem-se a uma classificação prévia dos produtos das lojas. Primeiramente, avalia-se a matriz de correlação entre essas variáveis, conforme apresentado na FIGURA 26. Caso todas as correlações entre variáveis fossem inferiores a 0,30, esta técnica teria fortes indícios de não ser apropriada. O que se observou, no entanto, a partir desta análise, é que existem valores significativos para justificar a utilização da técnica.



maioria, muito próximos de 1, o que denotam que as variáveis são perfeitamente previstas sem erro pelas demais variáveis.

TABELA 8 - MEASURE OF SAMPLING ADEQUACY (MSA)

<i>Variável</i>	MSA	<i>Variável</i>	MSA	<i>Variável</i>	MSA	<i>Variável</i>	MSA	<i>Variável</i>	MSA	<i>Variável</i>	MSA
<i>C.1.1</i>	0,979	<i>C.1.7</i>	0,873	<i>C.3.1</i>	0,907	<i>C.4.3</i>	0,972	<i>C.5.1</i>	0,961	<i>C.5.7</i>	0,929
<i>C.1.2</i>	0,954	<i>C.2.1</i>	0,738	<i>C.3.2</i>	0,910	<i>C.4.4</i>	0,965	<i>C.5.2</i>	0,969	<i>C.5.8</i>	0,976
<i>C.1.3</i>	0,981	<i>C.2.2</i>	0,852	<i>C.3.3</i>	0,899	<i>C.4.5</i>	0,956	<i>C.5.3</i>	0,973	<i>C.5.9</i>	0,985
<i>C.1.4</i>	0,960	<i>C.2.3</i>	0,729	<i>C.3.4</i>	0,926	<i>C.4.6</i>	0,954	<i>C.5.4</i>	0,974	<i>C.5.10</i>	0,981
<i>C.1.5</i>	0,969	<i>C.2.4</i>	0,645	<i>C.4.1</i>	0,965	<i>C.4.7</i>	0,973	<i>C.5.5</i>	0,937	<i>C.5.11</i>	0,925
<i>C.1.6</i>	0,952	<i>C.2.5</i>	0,868	<i>C.4.2</i>	0,978	<i>C.4.8</i>	0,957	<i>C.5.6</i>	0,968	<i>C.5.12</i>	0,968

FONTE: A autora (2020).

Para se decidir quantos componentes devem ser retidos foram avaliados os critérios de porcentagem da variância, da raiz latente (critério de Kaiser) e o critério do gráfico *Scree*. Observou-se, portanto, que 11 componentes em conjunto explicam 60,29% da variância total das variáveis utilizadas. Estes resultados podem ser observados na TABELA 9.

TABELA 9- VALORES DA VARIÂNCIA TOTAL EXPLICADA PELOS COMPONENTES IDENTIFICADOS NA ANÁLISE FATORIAL

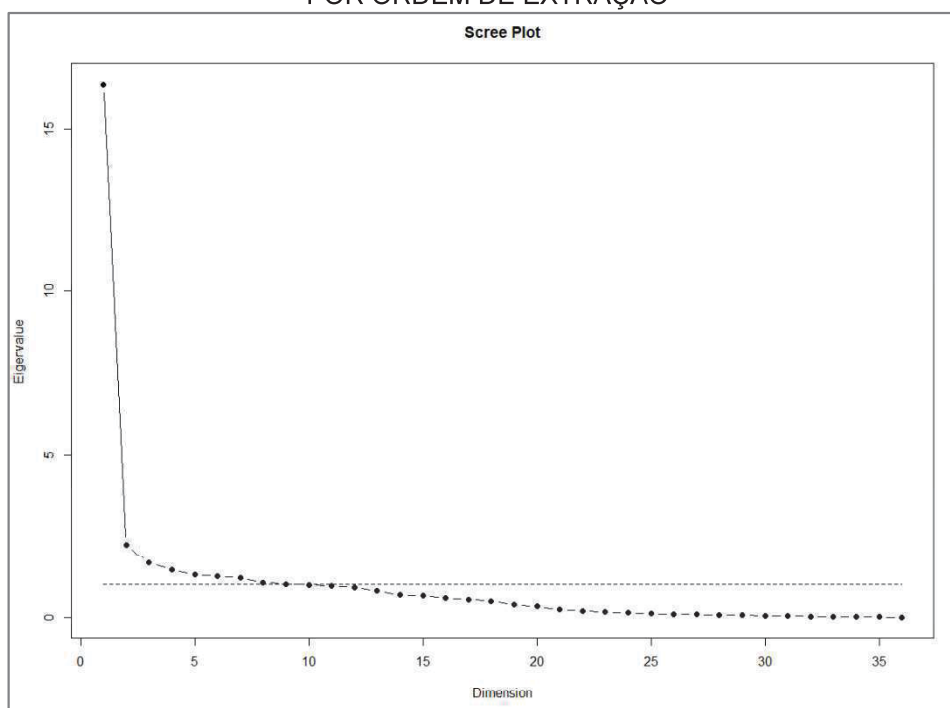
<b>Componente</b>	<b>Desvio-Padrão</b>	<b>Variância explicada (%)</b>	<b>Variância acumulada (%)</b>
1	32,656	0,2962	0,2962
2	122,245	0,04151	0,33775
3	121,124	0,04075	0,3785
4	107,743	0,03225	0,41074
5	102,193	0,02901	0,43975
6	101,186	0,02844	0,46819
7	0,99963	0,02776	0,49595
8	0,9969	0,0276	0,5236
9	0,98648	0,02703	0,55059
10	0,9725	0,02627	0,57686
11	0,96967	0,02612	<b>0,60298</b>
12	0,95962	0,02558	0,62856
13	0,95142	0,02514	0,6537
14	0,94999	0,02507	0,67877
15	0,94269	0,02469	0,70346
16	0,92733	0,02389	0,72734
17	0,92281	0,02365	0,751
18	0,90316	0,02266	0,77366
19	0,89822	0,02241	0,79607
20	0,88251	0,02163	0,8177
21	0,84474	0,01982	0,83752
22	0,81496	0,01845	0,85597
23	0,76494	0,01625	0,87223
24	0,71445	0,01418	0,8864

Componente	Desvio-Padrão	Variância explicada (%)	Variância acumulada (%)
25	0,68954	0,01321	0,89961
26	0,6618	0,01217	0,91178
27	0,65776	0,01202	0,9238
28	0,64713	0,01163	0,93543
29	0,63348	0,01115	0,94658
30	0,60514	0,01017	0,95675
31	0,57503	0,00919	0,96593
32	0,54712	0,00831	0,97425
33	0,52614	0,00769	0,98194
34	0,49628	0,00684	0,98878
35	0,47045	0,00615	0,99493
36	0,42738	0,00507	100000

FONTE: A autora (2020).

O GRÁFICO 1 apresenta o critério de Kaiser no gráfico de Scree. Considerou-se portanto, 11 componentes a serem extraídos, visto que representa uma variância acumulada no mínimo de 60%, além de satisfazer estes últimos critérios avaliados.

GRÁFICO 1 - GRÁFICO SCREE DOS AUTOVALORES VERSUS O NÚMERO DE COMPONENTES, POR ORDEM DE EXTRAÇÃO



FONTE: A autora (2020).

Inicialmente gerou-se a matriz de componentes sem rotação e avaliou-se os seus componentes. As cargas fatoriais não rotacionadas podem ser frequentemente difíceis de interpretar. Portanto, foram realizadas tentativas de diferentes rotações

como "Varimax", "Quartimax", "Promax" e "Oblimin" para que fosse utilizada aquela com resultados mais interpretáveis (Anexo 2). O método de rotação utilizado foi o método ortogonal Varimax por simplificar a interpretação dos componentes principais. A FIGURA 27 apresenta as cargas fatoriais, que indicam quanto um componente explica uma variável.

FIGURA 27 - REPRESENTAÇÃO DA MATRIZ COM CARGAS FATORIAIS APÓS A ROTAÇÃO ORTOGONAL COM MÉTODO VARIMAX E COMUNALIDADES

Variáveis	Cargas Fatoriais											Comunalidades
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	
C.1.1	0,31	0,06	<b>0,35</b>	0,04	0,04	0,28	0,03	0,11	0,08	0,03	0,02	<b>0,32</b>
C.1.2	0,09	0,03	<b>0,7</b>	0,08	0,15	-0,02	-0,05	-0,1	0,05	-0,02	0,07	<b>0,55</b>
C.1.3	0,21	0,22	<b>0,39</b>	-0,09	0,01	0,08	0,03	0,04	-0,05	0,01	-0,05	<b>0,27</b>
C.1.4	<b>0,66</b>	-0,05	0,35	0,11	-0,12	0,14	0,06	0,03	0,01	0,01	-0,01	<b>0,62</b>
C.1.5	0,12	0,08	0,17	-0,01	-0,02	<b>0,55</b>	0,06	0,21	-0,2	0,02	-0,07	<b>0,45</b>
C.1.6	0,14	0,01	-0,02	0,07	0,23	<b>0,67</b>	0	-0,12	0,09	0	0,07	<b>0,55</b>
C.1.7	0,01	0,01	-0,01	0,02	0	0,01	-0,01	-0,02	0,02	<b>1</b>	0	<b>0,99</b>
C.2.1	0,01	0	0,09	-0,07	0,03	-0,04	0,12	0,11	<b>0,89</b>	0,03	-0,02	<b>0,84</b>
C.2.2	0,05	0	0,04	0,1	0,01	0,04	<b>0,59</b>	0,12	0,09	0,01	0,1	<b>0,4</b>
C.2.3	0,02	0,01	0,03	-0,01	-0,03	0,01	0,08	0,05	-0,01	0	<b>0,98</b>	<b>0,96</b>
C.2.4	0,01	0,03	-0,04	-0,01	0,01	-0,01	<b>0,82</b>	-0,13	0	-0,01	-0,04	<b>0,69</b>
C.2.5	0,04	0,02	-0,06	0,09	0,01	0,03	0	<b>0,89</b>	0,09	-0,02	0,05	<b>0,82</b>
C.3.1	0,09	0,07	<b>0,44</b>	0,43	-0,12	0	0,04	-0,02	0,04	-0,01	-0,01	<b>0,41</b>
C.3.2	0,1	0,05	-0,19	0,28	-0,29	<b>0,36</b>	-0,15	-0,26	<b>0,36</b>	-0,08	0,07	<b>0,57</b>
C.3.3	0,08	0,05	0,12	<b>0,69</b>	0,04	0,02	0,05	0,08	-0,03	0,03	-0,01	<b>0,51</b>
C.3.4	0,12	0	-0,07	<b>0,68</b>	0,03	0,05	0,03	0,01	-0,03	0	0	<b>0,49</b>
C.4.1	<b>0,85</b>	-0,02	0,09	0,03	-0,01	0,04	0,02	0	0	0	0	<b>0,73</b>
C.4.2	<b>0,84</b>	0,01	0,11	0,06	-0,03	0,03	0,02	0,01	0	0	0,01	<b>0,72</b>
C.4.3	<b>0,71</b>	0,18	0	0,01	0,15	0,19	0,01	0,02	0	0,01	0	<b>0,6</b>
C.4.4	<b>0,8</b>	0,17	0,01	0,05	-0,04	0,16	0,02	0,03	0	0,01	0	<b>0,71</b>
C.4.5	<b>0,86</b>	-0,02	0,11	0,03	0,01	-0,04	0,01	-0,01	0	0	0,02	<b>0,75</b>
C.4.6	<b>0,84</b>	-0,05	0,21	0,03	-0,11	0,05	0,03	0,01	0	0,01	-0,01	<b>0,77</b>
C.4.7	<b>0,76</b>	-0,02	0,12	0,08	-0,18	0,14	0,03	0,02	0	0,01	-0,01	<b>0,66</b>
C.4.8	0,21	-0,06	0,11	0,01	<b>0,62</b>	0,24	0,03	0,02	0	-0,01	-0,04	<b>0,51</b>
C.5.1	<b>0,71</b>	0,17	0,1	0,09	0,32	-0,05	-0,02	-0,02	0,03	-0,01	0,02	<b>0,66</b>
C.5.2	<b>0,61</b>	0,21	0	0,08	0,36	0,03	-0,02	-0,04	0,04	0	0,03	<b>0,57</b>
C.5.3	<b>0,73</b>	0,14	0,05	0,05	0,03	0,11	0,02	0,02	0	0	-0,01	<b>0,57</b>
C.5.4	<b>0,64</b>	0,24	-0,04	0,05	0,14	0,1	0	0,02	0	0	0,01	<b>0,5</b>
C.5.5	0,41	<b>0,61</b>	0,03	0,06	0,26	-0,06	-0,02	-0,02	0,03	0	0,03	<b>0,62</b>
C.5.6	<b>0,61</b>	0,35	-0,04	0,05	0,3	0,01	-0,01	-0,01	0,03	0	0,02	<b>0,6</b>
C.5.7	0,11	<b>0,68</b>	0,1	0	-0,15	0,08	0,03	0,04	-0,03	0	-0,02	<b>0,51</b>
C.5.8	<b>0,71</b>	0,22	0,04	0,04	0,15	-0,03	-0,01	-0,01	0,01	0	0,02	<b>0,58</b>
C.5.9	<b>0,79</b>	0,09	0,09	0,03	0,1	0,01	0,01	0,01	0	0	0	<b>0,65</b>
C.5.10	<b>0,58</b>	0,2	-0,01	0,02	0,02	0,17	0,04	0,05	-0,02	0,01	-0,02	<b>0,41</b>
C.5.11	0,22	<b>0,71</b>	0,08	0,06	0,02	0,02	0,01	-0,01	0,02	0	0,01	<b>0,56</b>
C.5.12	<b>0,64</b>	0,11	0,03	0,08	0,34	-0,11	-0,03	-0,02	0,02	-0,01	0,02	<b>0,55</b>

FONTE: A autora (2020).

Com base na FIGURA 27, as variáveis C.1.1, C.1.3, C.1.5, C.2.2, C.3.1, C.3.4 e C.5.10, no entanto, obtiveram explicações menores que 50%. As variáveis C.1.5, C.2.2, C.3.4 e C.5.2 apresentam, todavia, associações positivas a determinados componentes superiores a 50%. Em relação às variáveis restantes, testou-se a adição de mais componentes, que não implicaram em uma contribuição com um ajuste mais significativo. Porém devido ao tamanho da amostra ser muito superior a 350 (como sugere a TABELA 1), as cargas fatoriais obtidas podem ser consideradas significativas. Assim, estas variáveis com proporção de variabilidade explicada acima de 30% foram relacionadas aos componentes principais.

Tendo em vista que os componentes podem ser representados pelo relacionamento linear das variáveis, os componentes determinantes para em relação ao consumo de itens dos clientes podem ser expressos pelas seguintes equações (70) a (80).

$$CP_1 = 0,86 * C. 4.5 + 0,85 * C. 4.1 + 0,84 * C. 4.2 + 0,84 * C. 4.6 + 0,8 * C. 4.4 \quad (70)$$

$$+ 0,79 * C. 5.9 + 0,76 * C. 4.7 + 0,73 * C. 5.3 + 0,71 * C. 4.3 + 0,71 * C. 5.1 + 0,71 * C. 5.8 + 0,66 * C. 1.4 + 0,64 * C. 5.4 + 0,64 * C. 5.12 + 0,61 * C. 5.6 + 0,61 * C. 5.2 + 0,58 * C. 5.10$$

$$CP_2 = 0,71 * C. 5.11 + 0,68 * C. 5.7 + 0,61 * C. 5.5 \quad (71)$$

$$CP_3 = 0,70 * C. 1.2 + 0,44 * C. 3.1 + 0,35 * C. 1.1 + 0,39 * C. 1.3 \quad (72)$$

$$CP_4 = 0,69 * C. 3.3 + 0,68 * C. 3.4 \quad (73)$$

$$CP_5 = 0,62 * C. 4.8 \quad (74)$$

$$CP_6 = 0,67 * C. 1.6 + 0,55 * C. 1.5 + 0,36 * C. 3.2 \quad (75)$$

$$CP_7 = 0,59 * C. 2.2 + 0,82 * C. 2.4 \quad (76)$$

$$CP_8 = 0,89 * C. 2.5 \quad (77)$$

$$CP_9 = 0,89 * C. 2.1 \quad (78)$$

$$CP_{10} = C. 1.7 \quad (79)$$

$$CP_{11} = 0,98 * C. 2.3 \quad (80)$$

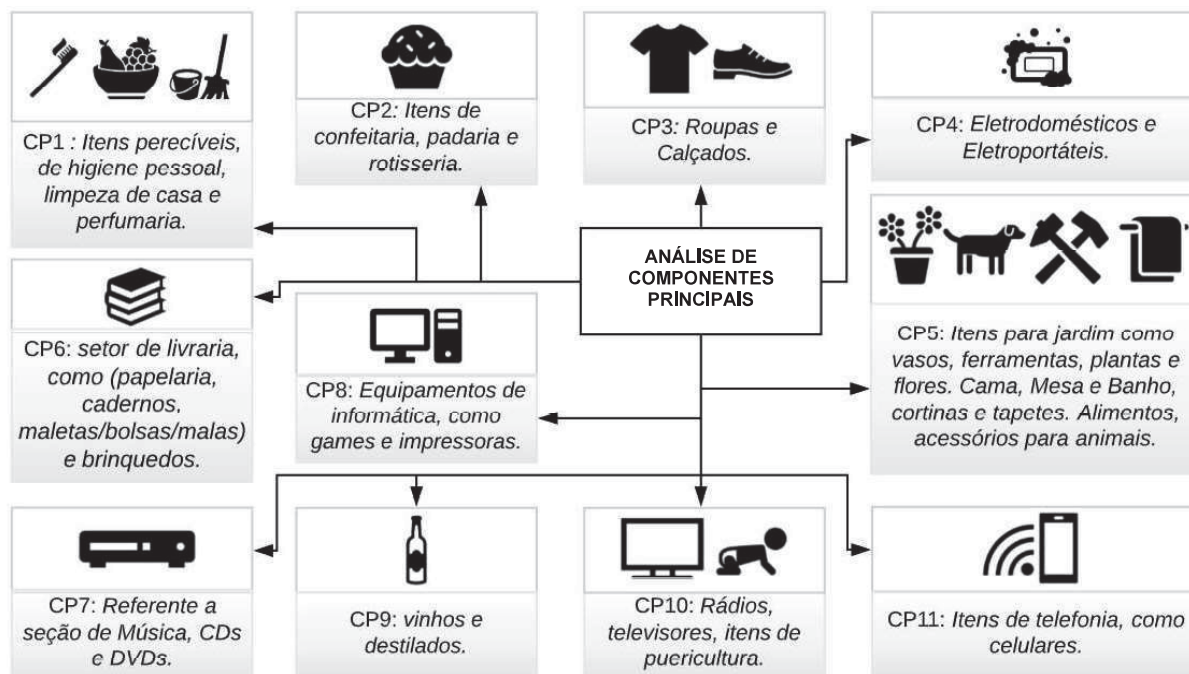
A interpretação e nomeação dos componentes por meio das cargas fatoriais é a última etapa da técnica de análise fatorial. A começar por:

- $CP_1$  refere-se a itens de necessidade básica, como açúcar, arroz, azeites, cafés, farináceos, feijão, temperos, leite, chás, biscoitos, massas, queijos, frutas e hortaliças, alimentos congelados, açougue, frutas e hortaliças, ovos limpeza, bebidas frias como água, chás, cervejas, energéticos, isotônicos, itens de limpeza de casa, de perfumaria e higiene pessoal;
- $CP_2$  está relacionado a um conjunto de itens do setor alimentício de perecíveis como itens de confeitaria, padaria e rotisseria. Dentre os itens relacionados estão os pães, biscoitos, bolos, dentre outros.
- $CP_3$  refere-se a roupas e calçados;
- $CP_4$  está relacionado a itens de maior valor como eletrodomésticos e eletroportáteis como por exemplo: bebedouros e purificadores, ar-condicionado, fogões, etc.;
- $CP_5$  está relacionado a itens de consumo de itens de consumo menos frequentes, como itens para jardim como vasos, ferramentas, plantas e flores. Também contém itens como cama mesa e banho, cortinas e tapetes. Além de ferragens, ferramentas (elétricas, hidráulica, iluminação) acessórios para automóveis. Por fim, neste grupo ainda se encontram itens relacionados a animais como alimentos, acessórios;
- $CP_6$  está relacionado ao setor de livraria, como (papelaria, cadernos, maletas/bolsas/malas) e brinquedos;
- $CP_7$  refere-se a seção de música, CDs e DVDs;
- $CP_8$  refere-se a equipamentos de informática, como games, *hardware* e impressoras;
- $CP_9$  refere-se a vinhos e destilados;
- $CP_{10}$  refere-se a aparelhos de DVDs, rádios, televisores, conjunto de som; além de itens de puericultura;
- $CP_{11}$  está relacionado a itens de telefonia como celulares.

Os componentes com os itens referentes estão representados na FIGURA 28.



FIGURA 28 - REPRESENTAÇÃO DOS ITENS RELACIONADOS AOS COMPONENTES



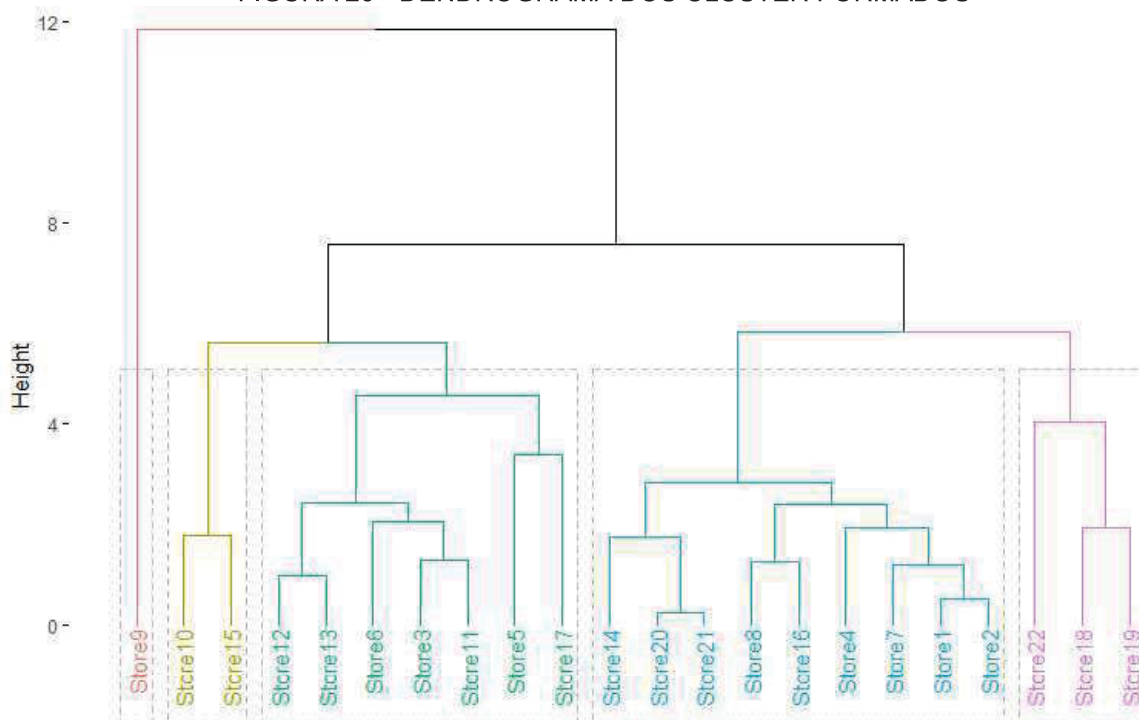
FONTE: A autora (2020).

Em seguida, com a obtenção dos escores das componentes principais, substituiu-se as variáveis originais relacionadas a itens de consumo pelas componentes principais e realizou-se a AC a fim de se classificar as lojas de acordo com os itens vendidos.

#### 4.1.1.2 Análise de Clusterização (AC)

A AC visou o agrupamento das 22 lojas pertencentes ao mesmo grupo. Para esta análise, parte-se do pressuposto de que características de padrões de consumo de itens das lojas possam ser usadas para separá-las em classes. O objetivo principal desta análise consiste em determinar em qual *cluster* se insere cada uma das lojas, permitindo avaliar algumas características semelhantes de sua atuação em função do que a loja pode oferecer e do tipo de perfil de consumidores que a acessam. Por meio da análise dos resultados, verificou-se que cinco *clusters* foram formados em função dos 11 componentes definidos na modelagem da ACP, como mostra a FIGURA 29.

FIGURA 29 - DENDROGRAMA DOS CLUSTER FORMADOS



FONTE: A autora (2020).

A TABELA 10 apresenta a descrição das características dos grupos em porcentagem para facilitar a análise. A coluna “Soma” apresenta a contribuição em termos de quantidade vendida de cada uma das classes.

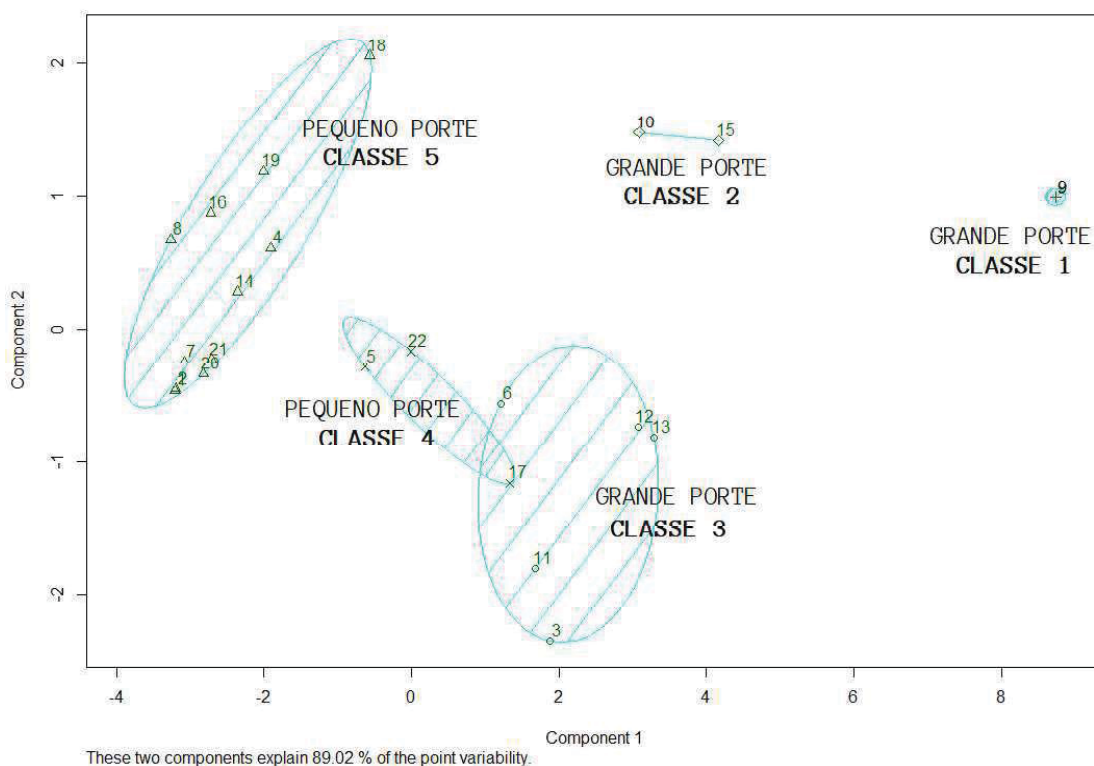
TABELA 10 – DESCRIÇÃO DOS GRUPOS

CLAS-SES	GRUPOS	CP <sub>1</sub>	CP	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>	CP <sub>6</sub>	CP <sub>7</sub>	CP <sub>8</sub>	CP <sub>9</sub>	CP <sub>10</sub>	CP <sub>11</sub>	POR-TE*	SOMA (%)
1	9	36%	9%	0%	0%	1%	0%	0%	0%	1%	0%	0%	G	36%
2	10, 15	22%	7%	0%	0%	1%	0%	0%	0%	0%	0%	0%	G	23%
3	3, 6, 11, 12, 13	17%	17%	26%	21%	17%	17%	10%	22%	15%	21%	23%	G	17%
4	5, 17, 22	15%	4%	0%	0%	0%	0%	0%	0%	0%	0%	0%	P	15%
5	1, 2, 4, 8, 16, 7, 14, 18, 19, 20, 21	9%	10%	3%	1%	9%	10%	9%	0%	8%	5%	0%	P	9%

Nota: \*G refere-se a lojas de Grande porte, enquanto P refere-se a lojas de Pequeno porte. FONTE: A autora (2020).

Realizou-se também a representação dos *clusters* em duas dimensões, conforme pode-se observar na FIGURA 30. É interessante observar que a loja 17 encontra-se mesclada à Classe 3 e 4. Isso ocorre porque embora seja considerada uma loja de pequeno porte, ela apresenta uma quantidade vendida superior à loja 6 por exemplo, que é considerada de grande porte. No entanto, ela está na classe 4 por compartilhar de restrições físicas de espaço assim como as outras lojas do seu grupo.

FIGURA 30 - REPRESENTAÇÃO DOS CLUSTER FORMADOS  
2D representation of Cluster



FONTE: A autora (2020).

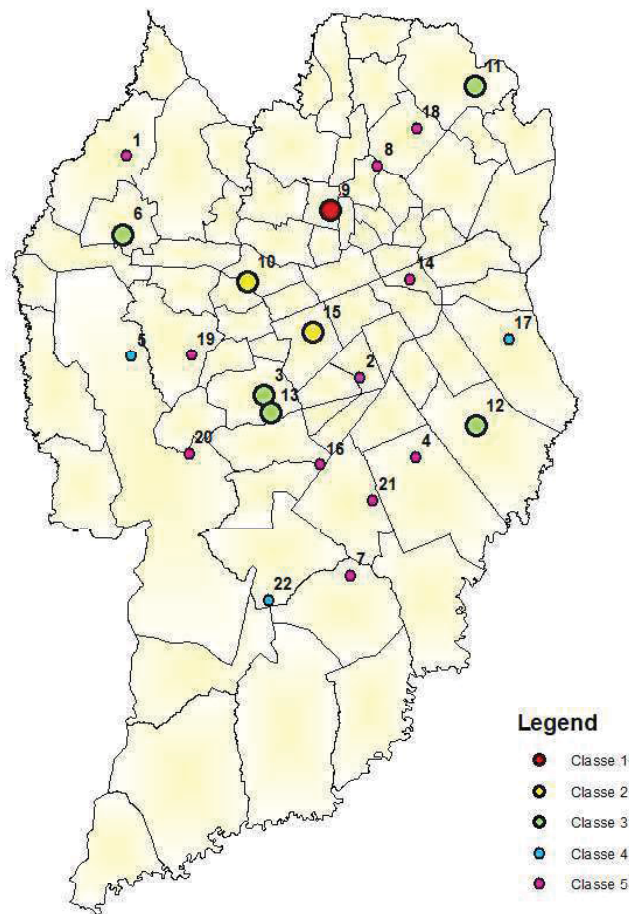
A seguir apresentam-se os aspectos mais determinantes de cada uma das classes:

- Classe 1: composta de apenas pela loja 9. A loja 9 desponta como uma das lojas mais completas em termos de quantidade e variedade de produtos da rede. Esta loja tem uma saída de produtos três vezes maior à média de vendas das lojas de pequeno porte. Conforme pode-se observar na TABELA 10, ela tem representatividade de 36% da quantidade de produtos vendidos da rede.
- Classe 2: composta por 2 lojas de grande porte (lojas 10 e 15), tem representatividade de 23% do total de itens comercializados. O que se observa é que estas lojas assim como a classe 1, concentram-se vendas de itens de necessidade básica, como alimentos, itens de higiene pessoal, de limpeza de casa, perfumaria, itens de confeitaria, padaria e rotisseria. Ou seja, itens de menor valor agregado, ao se comparar com itens comercializados na Classe 3 a seguir.

- Classe 3: composta por 5 lojas de grande porte, tem representatividade de 17% do total de itens comercializados. Pode-se observar que este grupo apresenta um índice maior de produtos comercializados que possuem uma maior margem de contribuição. Por exemplo, itens pertencentes aos componentes:  $CP_4$  (eletrodomésticos e eletroportáteis) com 21%;  $CP_8$  (equipamentos de informática, como games e impressoras) com 22%,  $CP_9$  (vinhos e destilados) com 15%;  $CP_{10}$  (rádios, televisores, itens de puericultura) com 21%; e  $CP_{11}$  (itens de telefonia como celulares) com 23%.
- Classe 4: composta por 3 lojas de pequeno porte, tem representatividade de 15% do total de itens comercializados. E de forma semelhante às classes 1 e 2 comercializam em sua maioria itens de menor valor agregado.
- Classe 5: composta pelo maior número de lojas (11 lojas) e que apresentam a menor representatividade (9%) do total de itens comercializados. Assim como a Classe 4, esta classe é constituída por lojas menores de pequeno porte. O que distingue esse grupo do anterior é o fato de haver a contribuição de mais componentes ( $CP_3$ ,  $CP_4$ ,  $CP_5$ ,  $CP_6$ ,  $CP_7$ ,  $CP_9$  e  $CP_{10}$ ).

A representação geográfica das lojas, bem como sua localização, está descrita na FIGURA 31.

FIGURA 31 - REPRESENTAÇÃO DAS 22 LOJAS LOCALIZADAS NO MUNICÍPIO DE CURITIBA



FONTE: A autora (2020).

#### 4.1.2 Técnicas de Predição

Nesta seção são apresentados os resultados das aplicações das técnicas RLM e RNA, cujo objetivo consiste na tarefa de predição de gastos de clientes em determinadas lojas a partir de suas características e preferências.

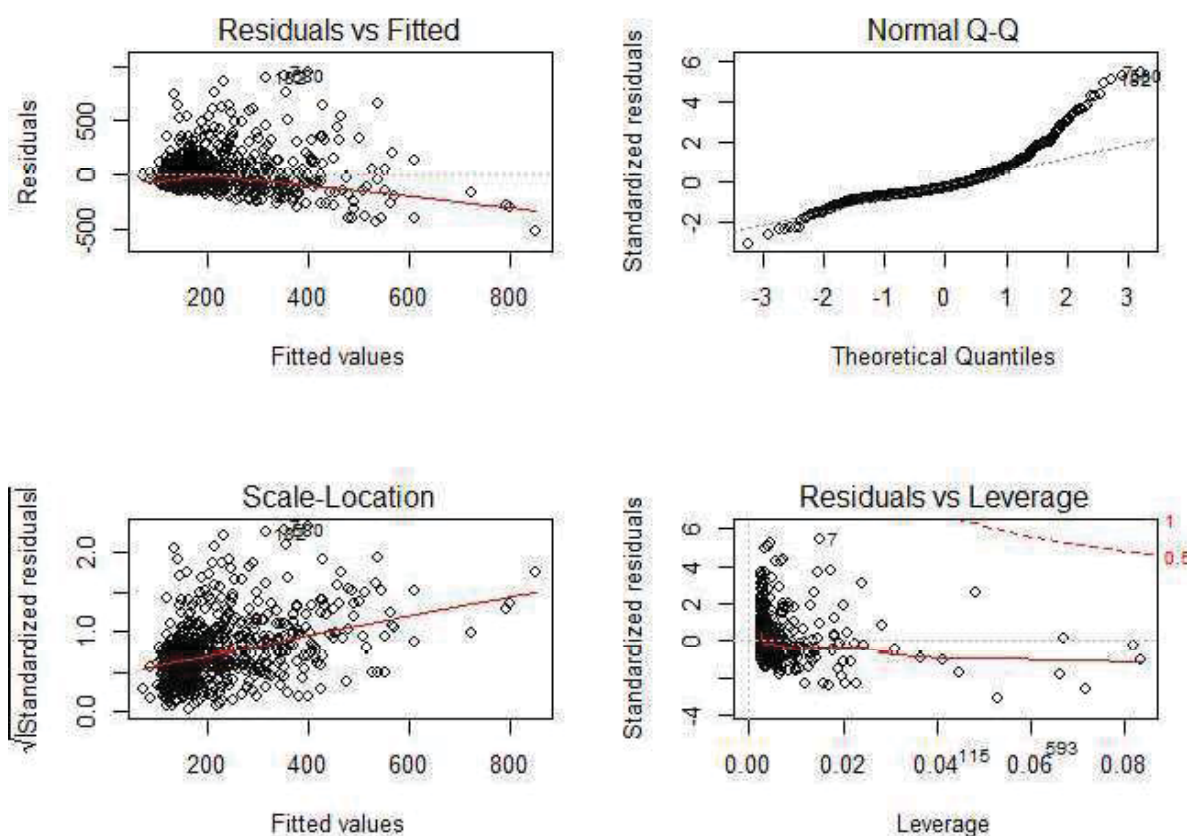
##### 4.1.2.1 Regressão Linear Múltipla (RLM)

Para a utilização do método *forward stepwise*, utilizou-se a estatística  $t$  ( $p$ -valor) com 95% de confiança. Para a classe de lojas 1, o primeiro modelo de regressão múltipla gerado foi ajustado e obteve  $p$ -valor de  $2,2e^{-16}$  para o teste F, que denota que as variáveis do modelo foram significativas conjuntamente. No entanto, o teste de Shapiro-Wilk para os resíduos rejeitou a hipótese de normalidade dos resíduos ( $W = 0,84005$ ,  $p$ -value  $< 2,2e^{-16}$ ). O teste de Anderson-Darling apresentou valores  $A =$

36,257,  $p\text{-value} < 2,2e^{-16}$ , reforçando a hipótese de que os dados não seguem distribuição normal. O teste de Durbin-Watson, por sua vez, foi de  $DW = 2,0097$ , e  $p\text{-valor} = 0,5554$ , constatando-se a não rejeição da hipótese nula, indicando a correlação serial ou dependência dos resíduos.

Hair et al. (2009) ressaltam que os testes de significância como o de Shapiro-Wilks são muito sensíveis em amostras grandes (que excedem 1000 observações). Assim, uma alternativa é a utilização de testes gráficos para avaliar o grau real de desvio da normalidade. Na FIGURA 32, percebe-se que os valores dos resíduos não estão perfeitamente em uma reta (quantil-quantil), indicando a falta de normalidade.

FIGURA 32 - REPRESENTAÇÃO GRÁFICA DOS RESÍDUOS VERSUS VALORES AJUSTADOS, QUANTIL QUANTIL DA NORMAL, RESÍDUOS VERSUS LEVERAGE

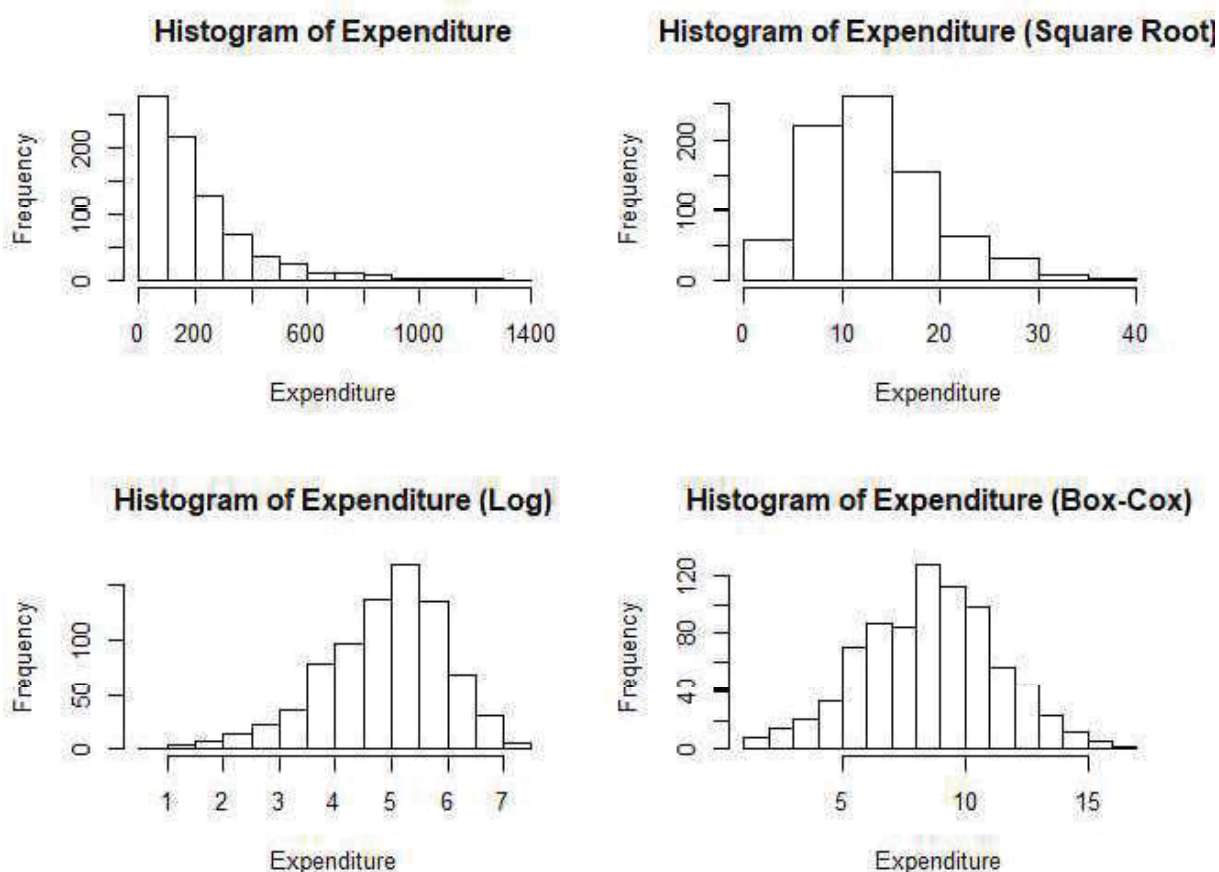


FONTE: A autora (2020).

Devido a isto, procedeu-se uma transformação de dados visando a modificação de alguma variável de modo a corrigir as violações das suposições

estatísticas e/ou melhorar as relações entre as variáveis. Fez-se tentativas de transformações tanto na variável dependente quanto nas variáveis independentes monitorando melhorias obtidas. Na FIGURA 33 apresentam-se o histograma da variável gasto original, seguido de transformações de aplicação de raiz-quadrada, logarítmica e por fim a Box-Cox. Essa ilustração é importante, pois a definição da melhor forma funcional é, por vezes, uma questão empírica a ser decidida a favor do melhor ajuste de variáveis.

FIGURA 33 - HISTOGRAMA DA VARIÁVEL DEPENDENTE EM RELAÇÃO ÀS FUNÇÕES DE TRANSFORMAÇÃO



FONTE: A autora (2020).

Para cada uma das transformações, foram avaliados os testes de normalidade, cujos valores estão descritos na TABELA 11. Pode-se observar que para o teste de Shapiro-Wilk realizado para a variável dependente, apenas a transformação Box-Cox não teve a hipótese de normalidade dos resíduos rejeitada ( $0,3208 > \alpha = 0,05$ ). Da mesma forma, para o teste de Anderson-Darling, o  $p$ -valor foi

maior do que o nível de significância ( $0,5016 > \alpha = 0,05$ ), o que permite concluir que os dados seguem a distribuição normal.

TABELA 11 – TESTES DE NORMALIDADE PARA A VARIÁVEL DEPENDENTE

Teste	Original, [p-valor]	Raiz Quadrada, [p-valor]	Logaritmo, [p-valor]	Box-Cox, [p-valor]
Shapiro-Wilk	W = 0,80229, [2,2e-16]	W = 0,96239, [1,788e-13]	W = 0,97421, [1,111e-10]	W = 0,99765, [0,3208]
Anderson-Darling	40,835, [2,2e-16]	A = 6,3131, [1,649e-15]	A = 4,8096, [6,502e-12]	A = 0,33955, [0,5016]

FONTE: A autora (2020).

Assim a transformação de Box-Cox foi adotada para a variável dependente, com o valor ótimo para transformação sendo  $\lambda = 0,2$ . Após a transformação ótima de Box-Cox, realizou-se uma nova análise com a variável transformada. Na TABELA 12, tem-se as estimativas dos parâmetros do modelo ajustado com a variável transformada. Rememora-se que o valor-p abaixo de 5% indica a rejeição da hipótese nula de que o coeficiente da variável em análise é igual a zero. Os valores dos coeficientes, por sua vez, estimam o impacto de cada variável no valor gasto pelo consumidor em determinada loja. Um coeficiente de valor maior indica uma variável que favorece ao consumidor gastos maiores.

TABELA 12 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO *STEPWISE* PARA CLASSE 1 (DADOS TREINAMENTO)

Parâmetros	Coefficientes	Erro padrão	Estatística <i>t</i>	Valor-p
(Intercepto)	7,988	0,1765	45,263	$<2e^{-16}$
SEX	-3,761	0,1645	-2,286	0,02249
DISTANCE	-0,00008006	0,00002587	-3,095	0,00204
F_1	0,05317	0,004027	13,203	$<2e^{-16}$
F_2	-0,04232	0,01447	-2,924	0,00356
F_4	1,347	0,6353	2,120	0,03435

FONTE: A autora (2020).

Conclusões que podem ser tiradas quanto ao modelo 1, é que a variável “SEX”, que refere-se ao gênero do cliente, apresentou significância relacionando-a ao valor gasto. Visto que ela tem ponderação negativa, conclui-se as mulheres tendem a gastar mais do que os homens nestas lojas. Além disso a variável “DISTANCE” sendo também negativa denota que pessoas que moram mais longe da loja tendem a fazer



compras com maiores gastos do que pessoas que moram mais próximas à loja em que a compra foi efetuada. Além disso, dentre as 11 classificações de produtos possíveis, os clientes com preferências pelos produtos do tipo 1, 2 e 4 representam aqueles com maiores gasto.

A equação de RLM que descreve a relação entre o valor gasto pelo cliente em relação às variáveis independentes escrita por (81).

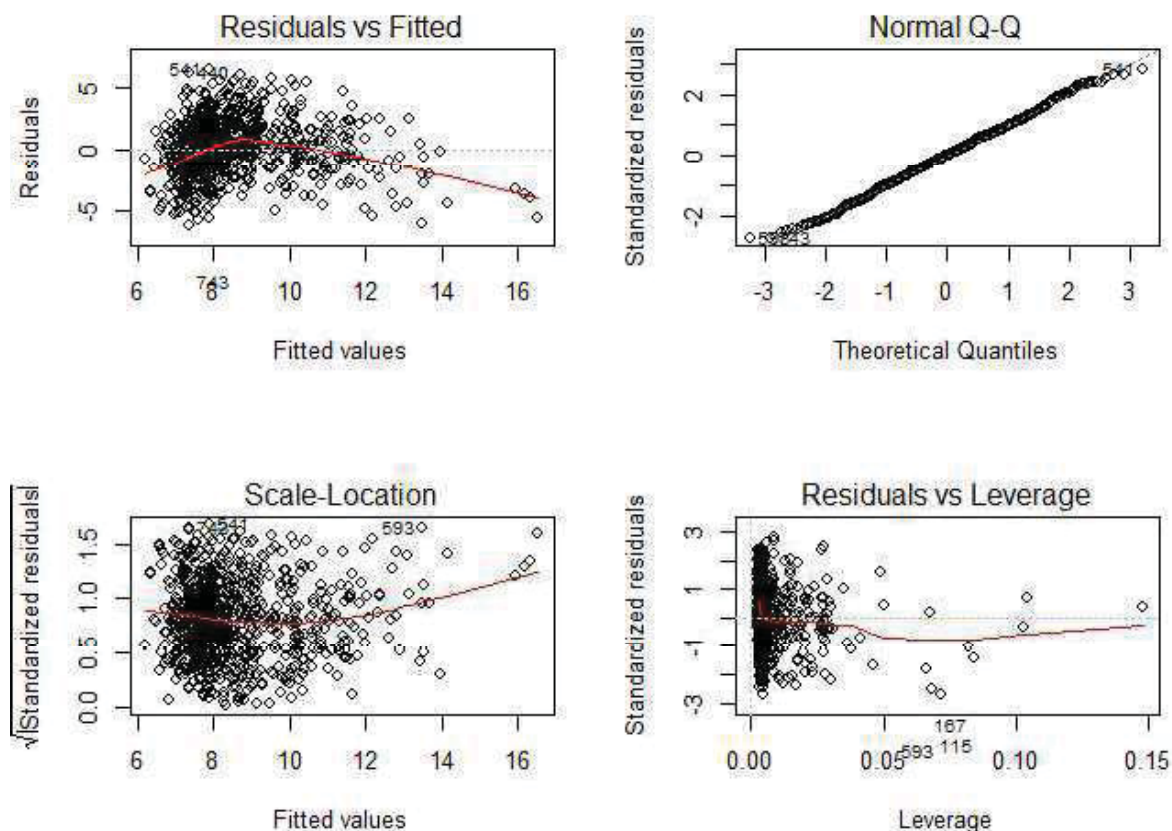
$$\frac{Expenditure^{0,2} - 1}{0,2} = 7,988 - 3,761 * SEX - 0,00008006 * DISTANCE + 0,05317 \quad (81)$$

$$* F_1 - 0,04232 * F_2 + 1,347 * F_4$$

O valor de  $R$  arredondado de 0,5433, o que mostra uma correlação moderada. Este valor representa a proporção dos dados que são explicados por meio do conjunto de variáveis explanatórias selecionadas. O coeficiente de determinação múltipla ( $R^2$ ) foi igual 0,2952 e o  $R^2$  ajustado foi 0,2907. O teste F (66,5) apresentou p-valor < 2,2e<sup>-16</sup> que indica que o modelo é significativo como um todo.

O teste de Shapiro-Wilks para os resíduos não rejeitou a hipótese de normalidade dos resíduos ( $W = 0,99773$ , p-valor = 0,3498). Da mesma forma, o teste de normalidade de Anderson-Darling, apresentou resultados de  $A = 0,24627$ , e p-valor = 0,756, que significa não rejeitar a hipótese nula de que os dados seguem uma distribuição normal. O teste de Durbin-Watson também não rejeitou a hipótese nula, ( $DW = 2,0189$ , p-valor = 0,6065), indicando que a transformação de Box-Cox corrigiu as suposições do modelo de regressão. Assim, na FIGURA 34 apresenta-se uma análise gráfica dos resíduos e por meio dela é possível verificar que houve melhora no ajuste da distribuição aos dados e que existe alguns pontos identificados que podem ser importantes e interferir na análise.

FIGURA 34 - REPRESENTAÇÃO GRÁFICA DOS RESÍDUOS APÓS TRANSFORMAÇÃO DE BOX-COX



FONTE: A autora (2020).

Os gráficos de diagnóstico apresentam os resíduos de quatro maneiras diferentes. O primeiro gráfico é um gráfico de dispersão dos valores de resíduos *versus* valores ajustados. Usado para verificar as suposições de relacionamento linear. Idealmente, o gráfico residual não mostrará nenhum padrão ajustado. Pode-se observar em evidência os pontos 743 e 541, que acabam por deslocar a linha vermelha da horizontal em zero. Esses pontos foram analisados e não há motivos evidentes para descartá-los da amostra. Vale ressaltar que há muito mais pontos com valores mais baixos e uma escassez de pontos são valores ajustados muito altos, essa alteração de densidade de pontos do gráfico, torna a interpretação mais difícil.

O segundo gráfico (QQ-normal) mostra como a distribuição de resíduos se encaixa na distribuição normal. Isso representa os resíduos padronizados em relação aos *quantis* normais teóricos. O que se pode observar é que todos os pontos caem

aproximadamente nessa linha de referência, o que reafirma a suposição de normalidade.

O terceiro apresenta (*Scale-Location*), um Gráfico de Escala-Localização, utilizado para verificar a homogeneidade da variância dos resíduos (homoscedasticidade). É essencialmente um gráfico de dispersão de resíduos normalizados com raiz quadrada absoluta e valores ajustados, com uma linha de regressão baixa. A linha horizontal com pontos igualmente dispersos é uma boa indicação de homoscedasticidade. Pode-se observar que a variabilidade (variâncias) dos pontos residuais aumenta com o valor da variável de resultado ajustado, sugerindo variações não constantes nos erros residuais (ou heterocedasticidade).

O quarto gráfico mostra quais pontos têm maior influência na regressão (pontos de alavancagem). Qualquer coisa fora do grupo e fora das linhas "Distância de Cook" pode ter um efeito influente no ajuste do modelo. Esses pontos também foram verificados e não apresentaram motivos evidentes para descartá-los da amostra.

Os valores de erros para os valores de teste são MAE=119,1538 e MSE =214,2868. Os mesmos passos aplicados para a Classe 1 são replicados para as outras classes. Iniciando pelo procedimento o *Stepwise Backward* com 95% de confiança. Os parâmetros de ajustes dos modelos para as Classes 2 a 5 estão descritos nas TABELA 13, TABELA 14, TABELA 15 e TABELA 16, respectivamente. Para cada análise, avaliou-se as possíveis transformações de variáveis para atender suposições da análise de regressão. A equação de RLM que descreve a relação entre o valor gasto pelo cliente em relação às variáveis independentes para a Classe 2 pode ser escrita por (39). A equação de RLM que descreve a relação entre o valor gasto pelo cliente em relação às variáveis independentes para a classe 5 de lojas, pode ser escrita pela equação (82).

$$\frac{Expenditure^{0,10} - 1}{0,10} = 5,387 - 0,00003414 * DISTANCE + 0,03282 * F1 - 0,02472 * F2 + 0,9188 * F4 + 1,312 * F10 + 1,274 * F11 \quad (82)$$

TABELA 13 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO STEPWISE PARA CLASSE 2

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
(Intercepto)	6,414	131,4	48,802	<2e <sup>-16</sup>
SEX	-0,4064	126	-3,226	0,00131

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
DISTANCE	-0,00003825	0,001879	-2,035	0,04214
F_1	0,03135	0,002399	13,066	<2e <sup>-16</sup>
F_9	0,121	0,04757	2,543	0,01118

FONTE: A autora (2019).

A equação de RLM que descreve a relação entre o valor gasto pelo cliente em relação às variáveis independentes para a classe de lojas 3, pode ser escrita por (83).

$$\frac{Expenditure^{0,2} - 1}{0,2} = 6,414 - 0,4064 * SEX - 0,00003825 * DISTANCE + 0,03135 * F_1 + 0,121 * F_9 \quad (83)$$

TABELA 14 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO STEPWISE PARA CLASSE 3

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
(Intercepto)	6,234	0,1314	47,460	<2e <sup>-16</sup>
SEX	-0,3168	0,1254	-2,527	0,011706
DISTANCE	-0,00004805	0,00001367	-3,516	0,000463
F_1	0,04412	0,002635	16,741	<2e <sup>-16</sup>
F_3	0,1961	0,09395	2,087	0,037225
F_4	1,442	0,2977	4,844	1,53e-06
F_11	0,9246	0,4525	2,043	0,041339

FONTE: A autora (2019).

A equação de RLM que descreve a relação entre o valor gasto pelo cliente em relação às variáveis independentes para a Classe 4 de lojas, pode ser escrita por (84).

$$\frac{Expenditure^{0,10} - 1}{0,10} = 5,869779 - 0,008491 * AGE - 0,046912 * INCOME + 0,032452 * F_1 - 0,027620 * F_2 + 0,853796 * F_4 + 0,901839 * F_{10} \quad (84)$$

TABELA 15 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO STEPWISE PARA CLASSE 4

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
(Intercepto)	5,869779	0,193868	30,277	<2e-16
AGE	-0,008491	0,003455	-2,458	0,014192
INCOME	-0,046912	0,020357	-2,305	0,021452

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
F_1	0,032452	0,002463	13,177	<2e-16
F_2	-0,027620	0,006902	-4,002	6,88e-05
F_4	0,853796	0,254608	3,353	0,000836
F_10	0,901839	0,348494	2,588	0,009835

FONTE: A autora (2019).

A equação de RLM que descreve a relação entre o valor gasto pelo cliente em relação às variáveis independentes para a classe 5 de lojas, pode ser escrita por (77).

$$\frac{Expenditure^{0,10} - 1}{0,10} = 5,387 - 0,00003414 * DISTANCE + 0,03282 * F_1 - 0,02472 * F_2 + 0,9188 * F_4 + 1,312 * F_{10} + 1,274 * F_{11} \quad (85)$$

TABELA 16 – AJUSTE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA UTILIZANDO STEPWISE PARA CLASSE 5

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
(Intercepto)	5,387	0,0927	58,100	<2e-16
DISTANCE	-0,00003414	0,0000115	-2,981	0,00296
F_1	0,03282	0,00271E	12,104	<2e-16
F_2	-0,02472	0,00951	-2,600	0,00949
F_4	0,9188	0,301	3,050	0,00236
F_10	1,312	0,452	2,906	0,00376
F_11	1,274	0,585	2,179	0,02963

FONTE: A autora (2019).

As estatísticas utilizadas para diagnóstico das suposições exigidas para a análise de regressão foram avaliadas para cada classe estão descritas na TABELA 17. Os resultados mostram que após a transformação de Box-Cox os modelos atendem aos pressupostos de normalidade com significância de 5% para os testes de Shapiro-Wilk e de Anderson-Darling. Da mesma forma, todos os modelos apresentaram a confirmação para todas as classes da independência dos resíduos pelo teste de Durbin-Watson.

TABELA 17 – RESULTADO DAS ESTATÍSTICAS ANALISADAS DOS RESÍDUOS PARA CADA CLASSE

Estatística	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5
Transformação de Box-Cox ( $\lambda$ )	0,2	0,15	0,10	0,10	0,10
Shapiro-Wilk (SW)	0,99773	0,99768	0,99748	0,99736	0,99771
p-valor	0,3498	0,3301	0,2605	0,2252	0,3419
Anderson-Darling (AD)	0,24627	0,40216	0,70382	0,30403	0,34986
p-valor	0,756	0,3576	0,06594	0,5705	0,4724

<b>Estatística</b>	<b>Classe 1</b>	<b>Classe 2</b>	<b>Classe 3</b>	<b>Classe 4</b>	<b>Classe 5</b>
Transformação de Box-Cox ( $\lambda$ )	0,2	0,15	0,10	0,10	0,10
Shapiro-Wilk (SW)	0,99773	0,99768	0,99748	0,99736	0,99771
p-valor	0,3498	0,3301	0,2605	0,2252	0,3419
Anderson-Darling (AD)	0,24627	0,40216	0,70382	0,30403	0,34986
Durbin-Watson (DW)	2,0189	1,9487	1,9674	2,0744	1,921
p-valor	0,6065	0,234	0,3223	0,855	0,1315

FONTE: A autora (2019).

Já a análise de variância, contida na TABELA 18, mostrou que os modelos são significativos, rejeitando-se a hipótese de não haver regressão. Por fim, para cada modelo proposto para cada classe, apresentam-se os coeficientes de correlação ( $R$ ) e de determinação ( $R^2$ ). Os coeficientes de correlações indicam a força de associação entre duas métricas. Pode-se observar que as correlações obtidas em todos os modelos foram moderadas.

TABELA 18 – ANÁLISE DE VARIÂNCIA DO AJUSTE DOS MODELOS E MEDIDAS DE ERROS PARA CADA CLASSE

<b>Descrição</b>	<b>Classe1</b>	<b>Classe2</b>	<b>Classe3</b>	<b>Classe4</b>	<b>Classe5</b>
Coefficiente de Correlação	0,5433	0,4780	0,5671	0,5073	0,5309
$R^2$ (coeficiente Correlação ao quadrado)	0,2952	0,2285	0,3216	0,2574	0,2819
$R^2$ ajustado	0,2907	0,2323	0,3165	0,2518	0,2765
Estatística F	66,5; <2e <sup>-16</sup>	60,15; <2e <sup>-16</sup>	62,65; <2e <sup>-16</sup>	45,81; <2e <sup>-16</sup>	51,89; <2e <sup>-16</sup>
Graus de Liberdade Resíduos	794	795	793	793	793
Soma dos Quadrados dos Resíduos	2,319	1,752	1762	1,343	1,394
MAE – Dados Treinamento	120,8408	90,95608	94,44668	84,8366	93,328
RMSE – Dados Treinamento	196,4526	154,0401	169,5457	149,4973	164,812
MAE – Dados Teste	119,1538	90,95608	133,0105	90,65113	107,689
RMSE – Dados Teste	214,2868	154,0401	324,3399	150,4197	212,473

FONTE: A autora (2020).

Embora o MAE seja importante para a interpretação dos erros, o RMSE deve ser mais útil quando erros grandes são particularmente indesejáveis, pois o RMSE atribui um peso relativamente alto a eles. Assim, em relação à essa métrica, pode-se observar que as RNAs apresentaram desempenho superior ao método de RLM para todas as classes.

#### 4.1.2.2 Redes Neurais Artificiais (RNA)

A arquitetura da rede, com topologia Perceptron de Múltiplas Camadas (*MultiLayer Perceptron* - MLP) *feedforward* apresentou uma camada de entrada de 16

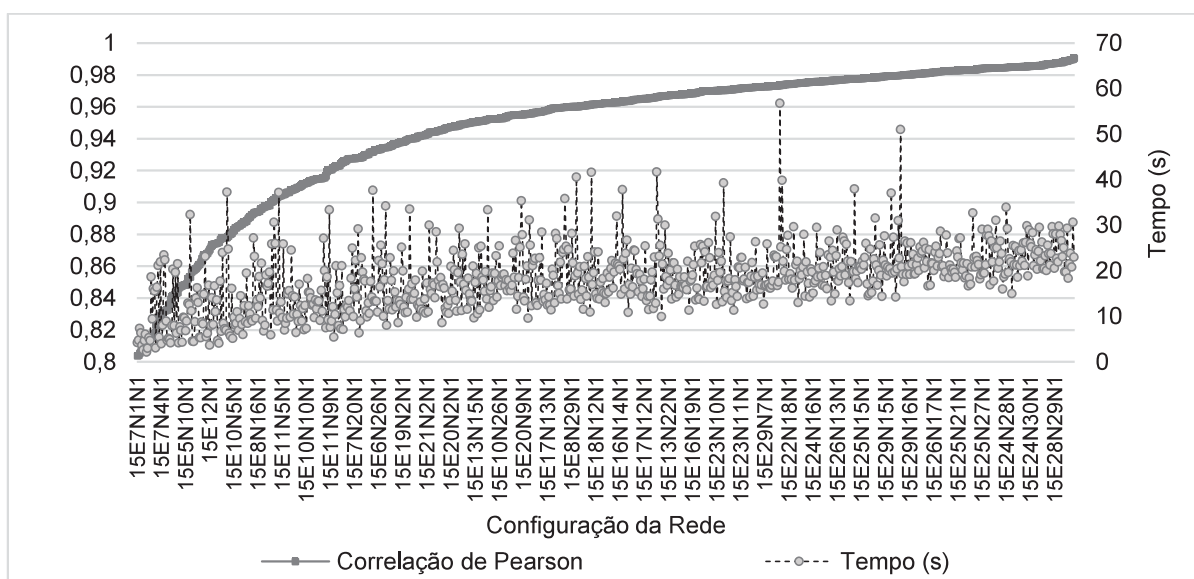
nós, constituído pelas 11 variáveis de itens de preferência, obtidas na ACP, além de variáveis como o sexo, idade, renda, endereço do cliente, distância até a loja. Para cada classe de lojas o número de variáveis de entrada pode variar de acordo com os produtos consumidos representados pelos componentes de cada uma.

A determinação do número de neurônios na camada intermediária deve ser analisada levando-se em conta que considerar poucos neurônios pode levar a uma capacidade de generalização maior, embora pode simplificar demais o modelo. Já o uso de muitos neurônios pode introduzir uma complexidade que o problema não possui, e fazer com que a rede fique muito restrita, e propensa a *overtraining*, ou seja, pode levar a uma superespecialização da rede e a uma perda da capacidade de generalização. Assim, foram realizados testes considerando até duas camadas escondidas, variando-se o número de neurônios de 1 a 30. A camada de saída é composta por um único neurônio que fornecerá o valor do gasto da compra. A taxa de treinamento foi considerada como 0,001, baixa o suficiente para evitar divergência.

A nomenclatura das redes é feita de forma que se identifiquem as camadas de entrada, a primeira e a segunda camadas ocultas, e por fim, a camada de saída. Como exemplo tem-se a rede 15E12N22N1, a qual apresenta 15 neurônios na camada de entrada, 12 neurônios na primeira camada oculta e 22 neurônios na segunda e um neurônio na camada de saída. Vale destacar que esta foi a rede que apresentou o melhor resultado de teste, ou seja, melhor capacidade de generalização, para a Classe 1 de lojas. O mesmo procedimento é realizado para as demais classes de lojas, e os resultados estão apresentados na TABELA 20.

A FIGURA 35 apresenta para uma análise do impacto do tempo de execução de redes de acordo com sua complexidade, ou seja, de acordo com um número maior de neurônios e/ou camadas escondidas. O que se pode observar é que o tempo aumenta de acordo com o aumento da complexidade da rede, as redes mais complexas não passam de 56 segundos. Porém, a correlação de Pearson embora tenha uma tendência de aumentar conforme a complexidade da rede, não apresenta essa relação de forma estritamente linear. Isto significa que boas correlações podem ser obtidas com redes mais simples também.

FIGURA 35 - TEMPOS DE EXECUÇÃO



FONTE: A autora (2020).

Os testes empíricos para a escolha do número de nós da camada oculta foram realizados considerando-se 1.000.000 como o número máximo de iterações. Na TABELA 19 são apresentadas as configurações de redes que obtiveram os melhores desempenho, dentre as várias testadas, por apresentarem os menores RMSE tanto no treinamento como no teste. Na TABELA 20 são apresentados os valores de RMSE e MAE, tanto do conjunto de treinamento, quanto do conjunto de testes.

TABELA 19 – RESULTADO DAS CONFIGURAÇÕES DE REDES COM MELHORES PERFORMANCE

REDE	TREINO				TESTE		Tempo (s)
	R*	R <sup>2**</sup>	RMSE	MAE	RMSE	MAE	
15E12N22N1	0,96224	0,92591	55,37083	37,73205	147,87727	103,32063	14,89
15E5N8N1	0,80417	0,64669	120,81593	81,44536	148,27907	101,56257	4,75
15E4N15N1	0,84920	0,72114	107,36269	71,60006	150,22863	104,11828	8,85
15E23N13N1	0,97012	0,94113	49,34457	32,79761	152,06031	112,07386	18,9
15E14N25N1	0,97819	0,95686	42,23861	28,82912	152,111	109,18146	24,67
15E9N9N1	0,89242	0,79641	91,72667	61,51668	152,36581	100,60649	8,88
15E6N30N1	0,90255	0,81460	87,53773	60,2891	152,37193	105,57912	25,98
15E6N26N1	0,93332	0,87109	73,08123	54,20605	152,48339	106,38594	22,25
15E4N18N1	0,85316	0,72788	106,02802	72,50861	152,54336	102,35525	12,54
15E18N26N1	0,97417	0,94901	45,91173	31,43276	152,66789	108,06258	19,24
15E16N19N1	0,96853	0,93805	50,63607	34,91651	152,94336	103,8473	19
15E5N24N1	0,90377	0,81680	87,17338	61,06048	152,99433	107,57201	37,18
15E7N9N1	0,88395	0,78137	95,11205	68,19886	153,21854	105,47496	8,41



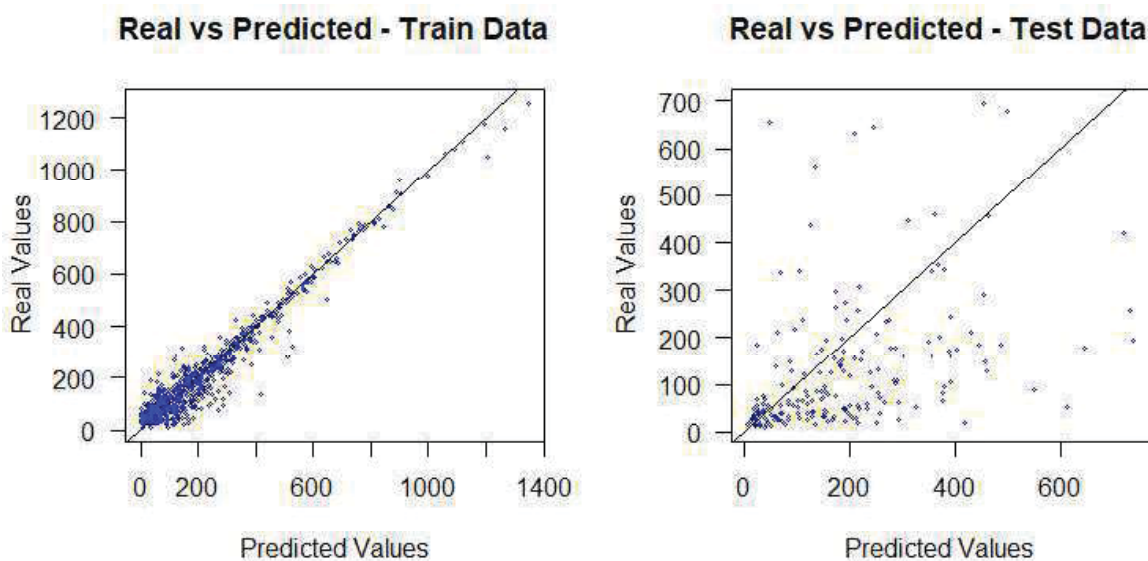
REDE	TREINO				TESTE		Tempo (s)
	R*	R <sup>2</sup> **	RMSE	MAE	RMSE	MAE	
15E29N6N1	0,97193	0,94465	47,87208	31,093	153,27735	106,21085	13,82
15E5N5N1	0,81526	0,66465	117,70049	80,87267	153,42817	104,24858	9,4

Nota: \*Correlação de Pearson. \*\*Coeficiente de Determinação.

FONTE: A autora (2020).

Pode-se observar que a configuração de rede 15E12N22N1 teve a melhor correlação de Pearson com 0,96224, o que representa um coeficiente de determinação de 0,92591, que representa o grau de ajuste com os dados de treino. Em relação aos erros de teste, esta rede apresentou os melhores resultados com RMSE de 147,87727 e MAE de 103,32063. A fim de visualizar a diferença de erros de treinamento e teste para esta rede, apresenta-se a FIGURA 36.

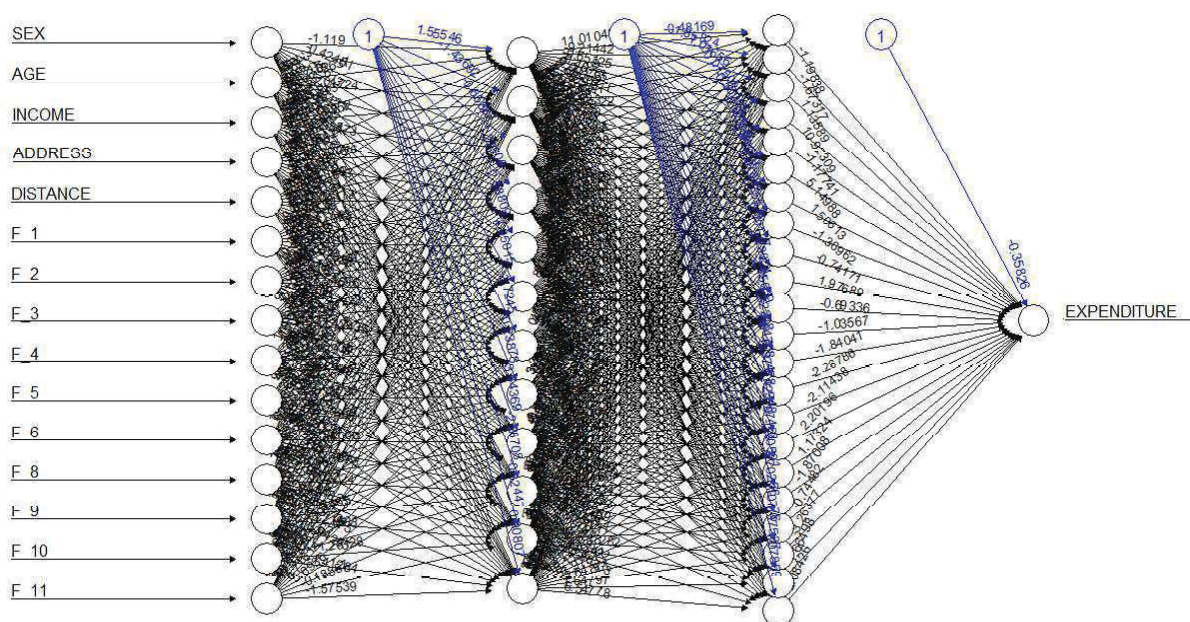
FIGURA 36 - VALORES OBSERVADOS E ESTIMADOS DA REDE 15E12N22N1



FONTE: A autora (2020).

A rede 15E12N22N1 com 12 neurônios na primeira camada oculta e 22 neurônios na segunda, apresentou o melhor resultado de teste, ou seja, melhor capacidade de generalização. A FIGURA 37 apresenta a representação da rede e no Anexo 3 apresenta as demais redes para as classes 2 a 5.

FIGURA 37 - REPRESENTAÇÃO DA REDE 15E12N22N1 – CLASSE 1



FONTE: A autora (2020).

O mesmo procedimento é realizado para as demais classes de lojas, e os resultados estão apresentados na TABELA 20.

TABELA 20 – RESULTADOS DE TREINO E TESTE PARA AS RNAS

Classes	REDE	TREINO				TESTE		Tempo (s)
		R*	R <sup>2</sup> **	RMSE	MAE	RMSE	MAE	
1	15E12N22N1	0,96224	0,92591	55,37083	37,73205	147,87727	103,32063	14,89
2	14E27N3N1	0,86477	0,7478	79,23064	53,30499	148,69015	<b>100,74604</b>	1,6
3	15E8N8N1	0,8796	0,7737	80,03575	57,67117	146,6951	92,88676	2,38
4	15E5N4N1	0,80094	0,64150	87,84139	61,59092	148,3372	<b>91,89039</b>	3,01
5	15E3N24N1	0,81463	0,6636	97,47816	67,17992	187,2185	<b>109,1</b>	5,8

FONTE: A autora (2020).

Comparando-se os resultados obtidos com as RNAs da TABELA 20 com os resultados da RLM da TABELA 18, constata-se que as RNAs apresentaram melhores resultados para todas as classes no conjunto de teste, visto que os erros apresentados em relação ao RMSE são menores. Portanto, os gastos considerados na etapa de roteirização são previstos com base nesta técnica.

#### 4.2 RESULTADOS COMPUTACIONAIS PARA O VRP

A seguir são apresentados os resultados para o VRP a partir da abordagem com o modelo matemático de PLIM o HGA-SP proposto.

#### 4.2.1 Descrição das Instâncias

Todos os valores dos parâmetros usados nas instâncias são relatados a seguir:

- Número de clientes ( $|V^C|$ ): 20 e 50
- Número de lojas ( $|V^D|$ ): 3 e 5, considerando-se as lojas com Classes distintas (conforme sugerido na AC).
- Tempo de serviço no cliente  $i$  ( $\tau_i$ ): por tempo de serviço entenda o tempo de realização da entrega do produto ao cliente após o trajeto até ele ter sido percorrido. Assume-se que esse tempo depende dos tipos de itens adquiridos pelo cliente. São considerados: 30 minutos para itens do tipo F4, F8 e F10, que são eletrodomésticos, por exemplo, e 15 minutos para outros.
- Demanda do cliente  $i$  na loja  $d$  ( $q_{id}$ ): é definida a partir dos métodos de mineração de dados conforme visto na seção anterior, assumindo-se que o tamanho da carga é proporcional a demanda ou valor gasto pelo cliente.

A distância percorrida do nó  $i$  para  $j$  ( $c_{ij}$ ): obtida por meio distância euclidiana entre as coordenadas geográficas obtidas a partir do endereço do cliente.

Parâmetros relacionados à periodicidade:

- Número de padrões de dias com uma frequência de entrega  $f_i = 1$  do cliente  $i$  ( $|L_i|$ ): 3
- Número de dias do horizonte de planejamento ( $|T|$ ): 1 e 5

Parâmetros relacionados às características dos veículos:

- Número de tipos de veículos de frota própria ( $|K_{s=0}|$ ): 3
- Número de tipos veículos de frota terceirizada ( $|K_{s=1}|$ ): 3
- Duração máxima da rota do veículo  $k$  ( $H_{isk}$ ): 8 (em horas)
- Custo Variável por distância percorrida do veículo  $k$  ( $V_k$ ): 1 (em R\$/m)

- Velocidade média dos veículos ( $\vartheta$ ): 60 (em km/h)
- As características dos veículos são demonstradas na TABELA 21.

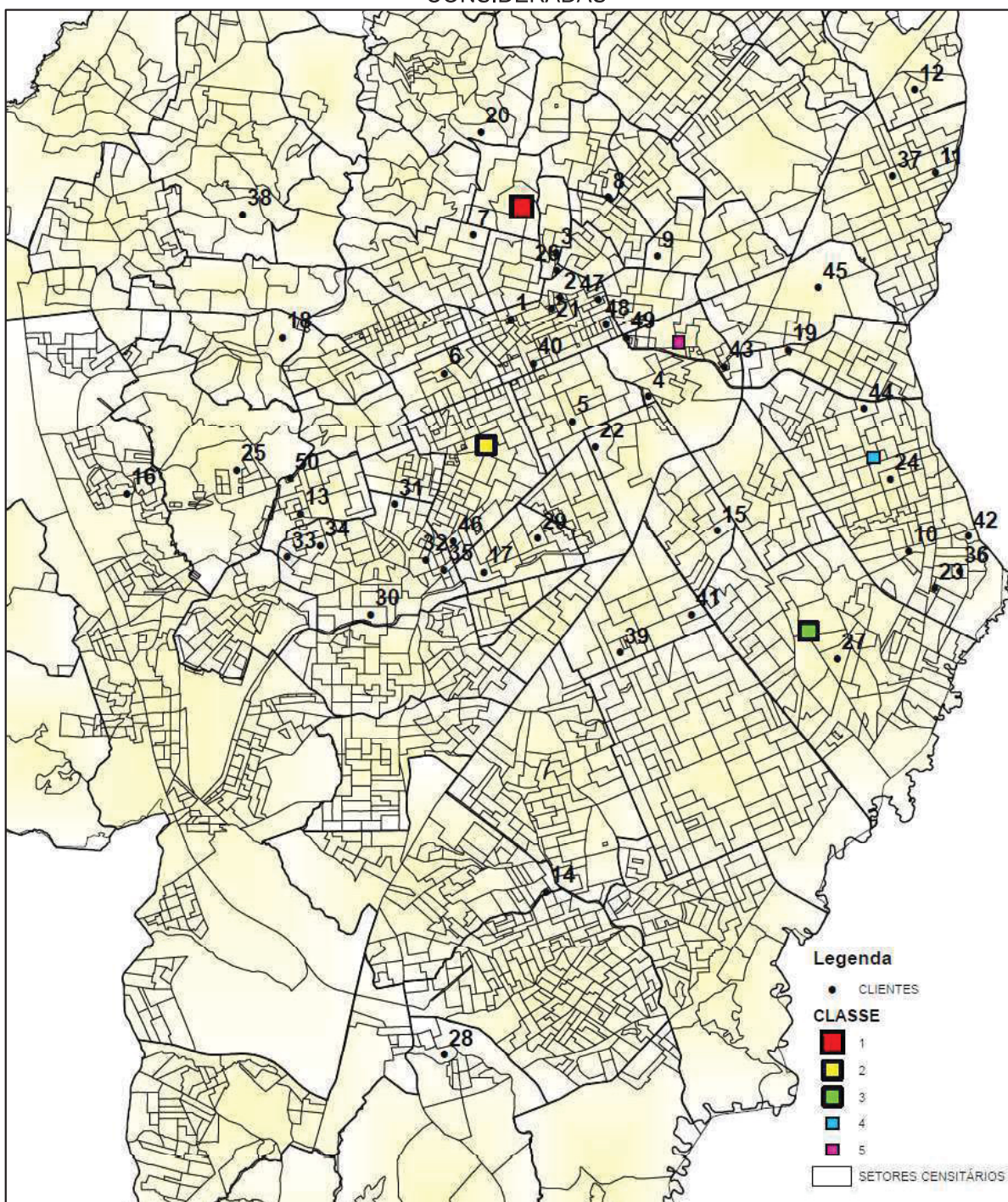
TABELA 21 – CARACTERÍSTICAS RELACIONADAS AOS TIPOS DE VEÍCULOS

Tipo de frota	Heterogêneo						Homogêneo	
	0			1			0	1
Tipos de veículos por tipo de frota ( $K_s$ )	1	2	3	1	2	3	1	1
Número de veículos do tipo $k$ ( $ M_{s,k} $ ):	2	1	1	1	2	3	2	2
Custo Fixo do veículo $k$ ( $F_k$ ) (em R\$):	0	0	0	4500	4500	4500	0	4500
Capacidade do veículo $k$ da frota $s$ ( $Q_{sk}$ ):	500	1000	2500	500	1000	2500	2500	2500

FONTE: A autora (2020).

A FIGURA 38 apresenta a configuração real da localização dos 50 clientes e das 5 lojas avaliadas.

FIGURA 38 - ILUSTRAÇÃO DO POSICIONAMENTO GEOGRÁFICO DOS CLIENTES E DAS LOJAS CONSIDERADAS



FONTE: A autora (2020).

Em relação a periodicidade dos itens, neste exemplo, considera-se que uma frequência única de entrega, o que significa que o cliente irá receber seu produto uma única vez ao longo do horizonte de planejamento. Essa característica é compatível

com a ideia de entrega em domicílio realizada pela Amazon® ou outros sites de compras *on-line*. A possibilidade de entregas aos clientes com uma frequência predeterminada, característica do problema PVRP clássico, foi mantida, no entanto, para representar casos reais que podem surgir em aplicações práticas de outras empresas. Nas quais o cliente deseja receber o produto com frequência maior, poderia se citar como exemplos específicos a entrega de marmitas ou alimentos orgânicos, que são entregues duas vezes na semana, ou diariamente.

Para a geração das instancias, verificou-se a influência da periodicidade, número de veículos e do custo fixo no tempo computacional. O modelo é testado em instâncias com características diferentes, a fim de recriar situações realistas possíveis para o cenário analisado. A TABELA 22 apresenta uma visão geral dos dados.

TABELA 22 - CARACTERÍSTICAS DAS INSTANCIAS GERADAS

Instancia	Número de clientes	Número de Depósitos	Periodicidade	Veículos
n20d3p1v1	20	3	Único dia	Homogêneo
n20d3p1v2	20	3	Único dia	Heterogêneo
n20d3p2v1	20	3	T=5*	Homogêneo
n20d3p2v2	20	3	T=5*	Heterogêneo
n20d5p1v1	20	5	Único dia	Homogêneo
n20d5p1v2	20	5	Único dia	Heterogêneo
n20d5p2v1	20	5	T=5*	Homogêneo
n20d5p2v2	20	5	T=5*	Heterogêneo
n50d3p1v1	50	3	Único dia	Homogêneo
n50d3p1v2	50	3	Único dia	Heterogêneo
n50d3p2v1	50	3	T=5*	Homogêneo
n50d3p2v2	50	3	T=5*	Heterogêneo
n50d5p1v1	50	5	Único dia	Homogêneo
n50d5p1v2	50	5	Único dia	Heterogêneo
n50d5p2v1	50	5	T=5*	Homogêneo
n50d5p2v2	50	5	T=5*	Heterogêneo

Nota: Clientes em 3 padrões possíveis:  $\{1,0,0,0,0\},\{0,1,0,0,0\},\{1,1,0,0,0\}$  com frequência de atendimento de um dia. FONTE: A autora (2020).

#### 4.2.2 Exemplo Numérico

Considerou-se para o exemplo numérico:

- Número de clientes ( $|V^C|$ ): 6
- Número de lojas ( $|V^D|$ ): 3
- A distância percorrida do nó  $i$  para  $j$  ( $c_{ij}$ ): obtida por meio distância euclidiana entre as coordenadas geográficas obtidas a partir do endereço do cliente.

Parâmetros relacionados à periodicidade:

- Número de dias do horizonte de planejamento ( $|T|$ ): 5
- Número de padrões de dias com uma frequência de entrega  $f_i = 1$  do cliente  $i$  ( $|L_i|$ ), cujos padrões estão descritos na TABELA 23: 1

Parâmetros relacionados às características dos veículos:

- Número de tipos de veículos de frota própria ( $|K_{s=0}|$ ): 3
- Número de tipos veículos de frota terceirizada ( $|K_{s=1}|$ ): 3
- Duração máxima da rota do veículo  $k$  ( $H_{sk}$ ): 8 (em horas)
- Velocidade média dos veículos ( $\vartheta$ ): 60 (em km/h)

O TABELA 23 apresenta os dados dos clientes bem como as coordenadas geográficas das lojas utilizado no exemplo.

TABELA 23 – CARACTERÍSTICAS DO EXEMPLO NUMÉRICO

Cliente	Coordenada x	Coordenada y	Tempo de serviço ( $\tau_i$ ) (em h)	Demanda do cliente $i$ na loja $d$ ( $q_{id}$ )			Frequência de entrega ( $f_i$ )	Número de padrões de dias	Lista de Padrões ( $L_i$ )
				Loja 0	Loja 1	Loja 2			
1	673396	7186223	0.25	474	259	2137	1	2	{16, 8}
2	675007	7186588	0.25	275	122	49	1	2	{16, 8}
3	675152	7186139	0.25	176	70	161	1	1	{16}
4	675930	7184814	0.25	43	74	103	1	1	{16}
5	674527	7184365	0.25	129	115	275	1	1	{8}
6	672694	7187787	0.25	87	68	32	1	1	{8}
7	673595	7188319							
8	676497	7185810							
9	672932	7183890							

FONTE: A autora (2020).

Em respeito à lista de padrões de dias apresentada, seguiu-se o formato da instancia clássica do PVRP proposta por Cordeau, Gendreau e Laporte (1997a). Aqui cada combinação de visita é codificada com o equivalente decimal da sequência de *bits* binários correspondente. Por exemplo, considerando-se uma frequência de dois dias de entregas, em um horizonte de planejamento de 5 dias, o código 10, que é equivalente à sequência de bits 01010, significa que um cliente é visitado nos dias 2

e 4, no caso da frequência de atendimento ser 2; ou 2 ou 4, no caso da frequência de atendimento ser 1. Observe que os dias são numerados da esquerda para a direita. O código 8 e o código 16, utilizados no exemplo, são equivalentes a 01000, que significa que um cliente é visitado no dia 2 e 10000, que significa que um cliente é visitado no dia 1.

TABELA 24 – PARÂMETROS CONSIDERADOS PARA O EXEMPLO

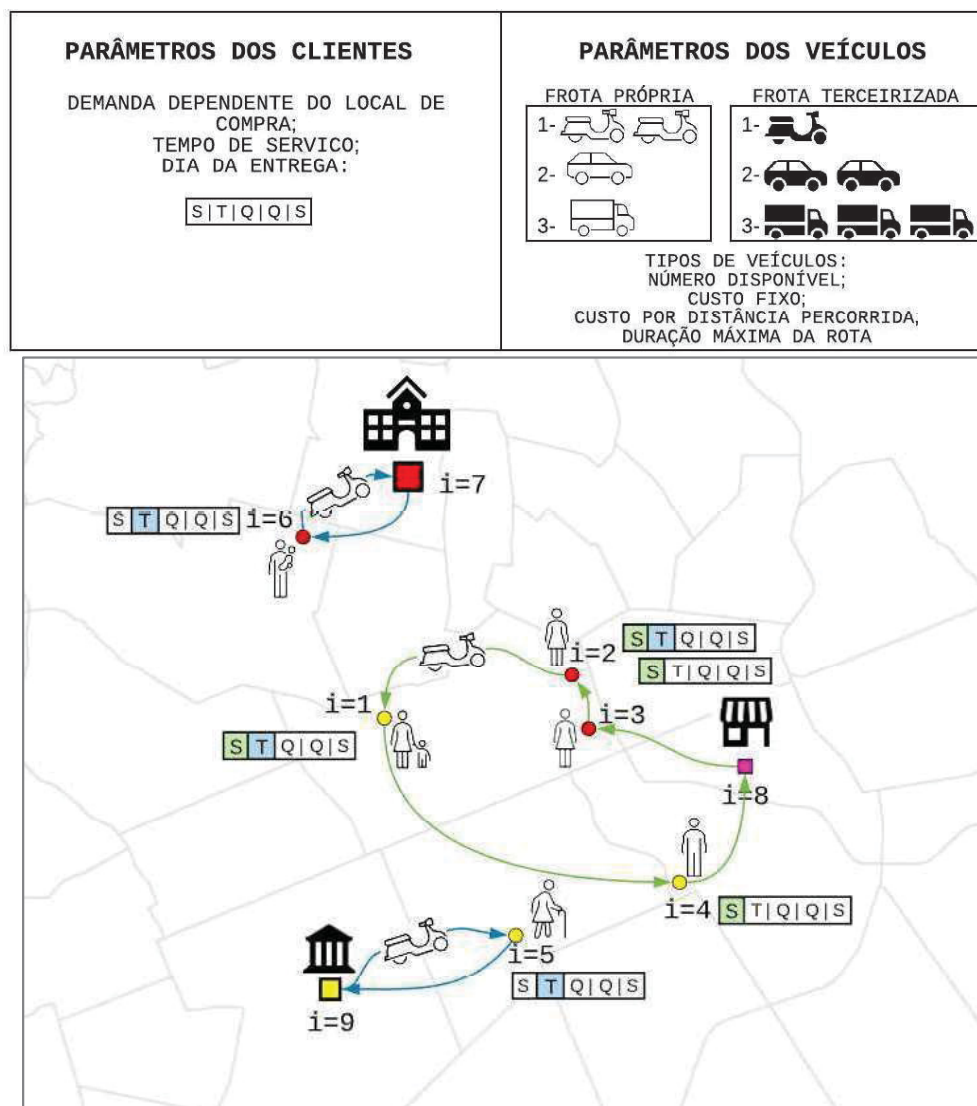
Tipo de frota	0			1		
	1	2	3	1	2	3
Tipos de veículos por tipo de frota ( $K_s$ )	1	2	3	1	2	3
Número de veículos do tipo $k$ ( $ M_{s,k} $ ):	2	1	1	1	2	3
Custo Fixo do veículo $k$ ( $F_k$ ) (em R\$):	0	0	0	4500	4500	4500
Custo Variável por distância percorrida ( $V_k$ ) (em R\$/km)	1,0	1,1	1,2	1,0	1,1	1,2
Capacidade do veículo $k$ da frota $s$ ( $Q_{sk}$ ):	500	1000	2500	500	1000	2500

FONTE: A autora (2020).

Apresenta-se na FIGURA 39, uma ilustração dos resultados obtidos com o modelo matemático e com o padrão de resposta por rota para cada caso. Cada círculo da figura ilustra a posição geográfica de um cliente, enquanto os quadrados representam a localização das diferentes lojas de um grupo de empresas. Supôs-se um horizonte de planejamento de 5 dias. Destacou-se na figura as preferências de atendimento dos clientes, em azul, os clientes atendidos na terça-feira e em verde os clientes que desejam ser atendidos na quarta-feira, sendo que ele pode desejar uma opção ou outra independentemente, como é o caso dos clientes 1 e 2.



FIGURA 39 - ILUSTRAÇÃO DO RESULTADO DO EXEMPLO NUMÉRICO CONSIDERANDO A (I) MINIMIZAÇÃO DE DISTÂNCIA E A (II) MAXIMIZAÇÃO DE BENEFÍCIOS



Solução Ótima =12974,91800

Limitante Inferior =12974,91800

Gap =0.00000

Tempo (s) =3,10715

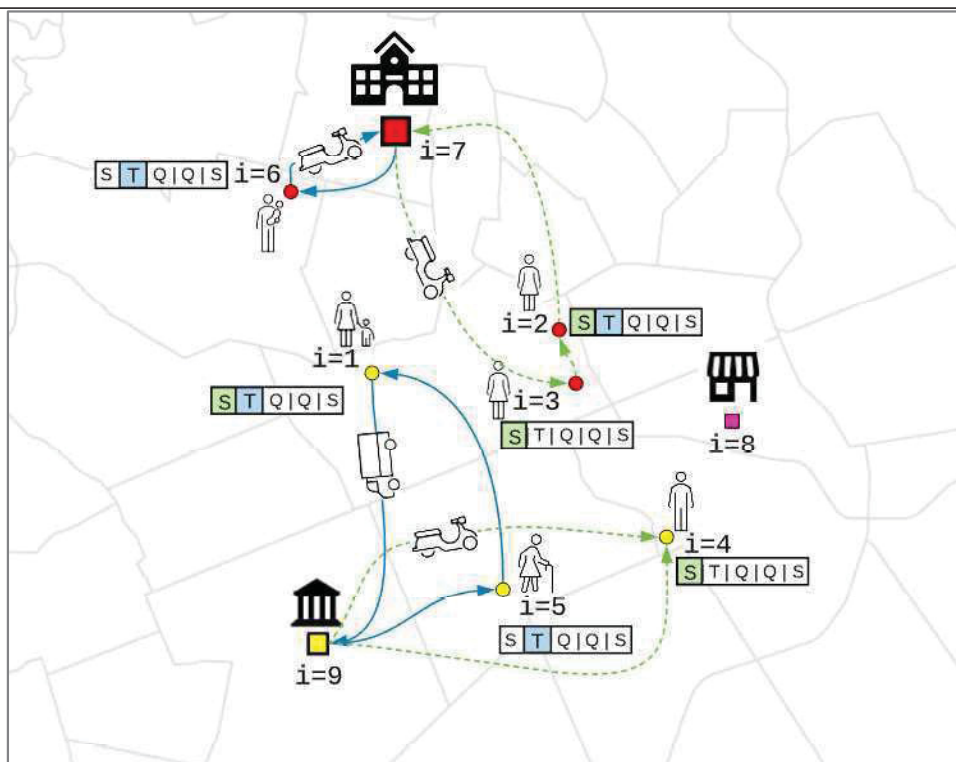
Legenda: dep=Depósito, Q = Carga Máxima Permitida; q=Demanda, h=Duração (h); H= Duração Máxima Permitida; s=Internal[0]/External[1] Frota; k =Tipo de veiculo; v = Vehicle;

Rota = 8 4 8, dep=8, q=275, Q=500, h=0.30547, H=8, s =0, k=0, v=1, t=1, dist=3328.45400

Rota = 7 2 1 0 3 7, dep=7, q=462, Q=500, h=1.12590, H=8, s =0, k=0, v=0, t=0, dist=7553.78600

Rota = 6 5 6, dep=6, q=100, Q=500, h=0.28488, H=8, s =0, k=0, v=0, t=1, dist=2092.67800

(i)



Solução Ótima =2995,05584

Limitante Inferior =2995,05584

Gap =0,00000

Time (s) =1,46201

Legenda: dep=Depósito, Q = Carga Máxima Permitida; q=Demanda, h=Duração (h); H= Duração Máxima Permitida; s=Internal[0]/External[1] Frota; k =Tipo de veículo; v = Vehicle;

Rota = 7 2 3 7, dep=0, q=445, Q=500, h=0.589744, H=8, s=0, k=0, t=0, dist=5384.61

Rota = 9 4 9, dep=2, q=100, Q=500, h=0.354572, H=8, s=0, k=0, t=0, dist=6274.32

Rota = 7 6 7, dep=0, q=100, Q=500, h=0.284878, H=8, s=0, k=0, t=1, dist=2092.68

Rota = 9 1 5 9, dep=2, q=2375, Q=2500, h=0.603635, H=8, s=0, k=4, t=1, dist=6218.08

(ii)

FONTE: A autora (2020).

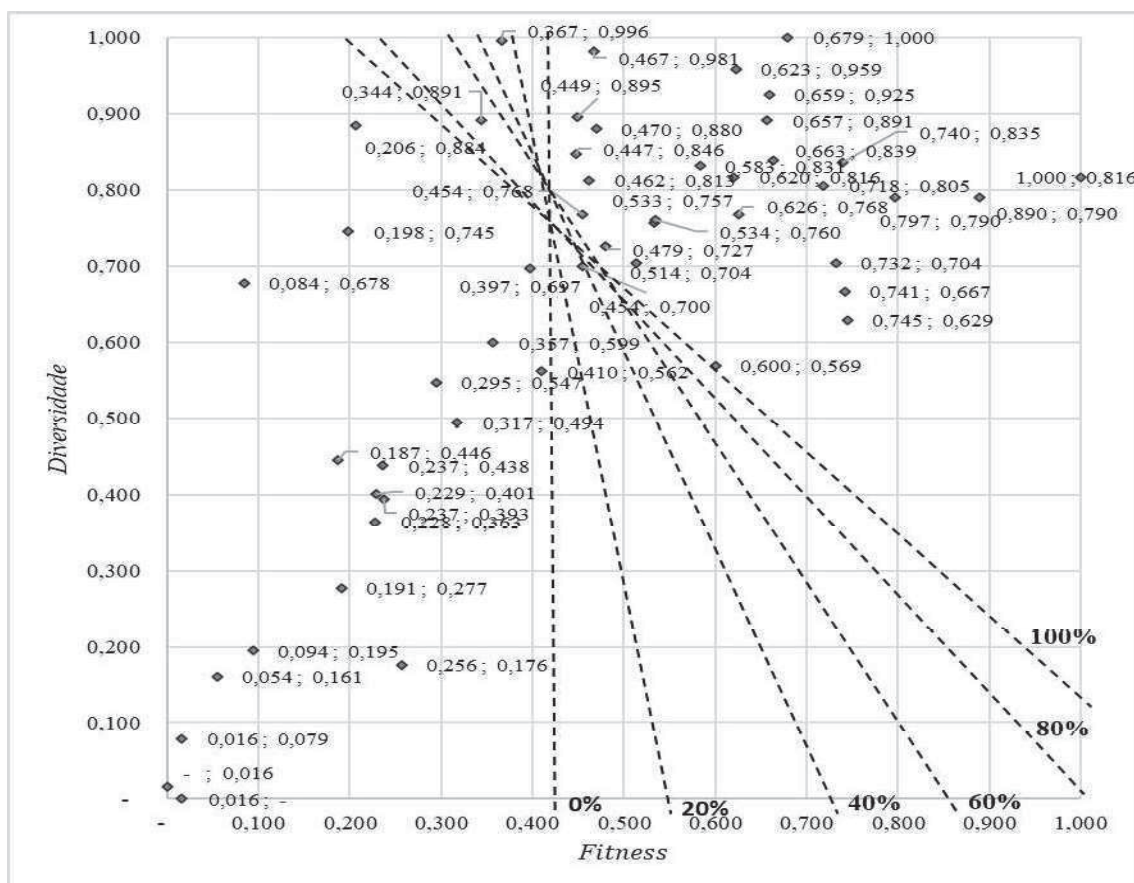
Pode-se observar que para a segunda situação, em que os benefícios com a entrega são maximizados, as rotas são definidas considerando não somente a distância percorrida, mas também a preferência de compra de cada cliente. Os clientes 2 e 3 por exemplo, antes eram atendidos pela loja situada em  $i=8$  devido a sua proximidade, e para a versão de maximização passam a ser atendidos pela loja situada no ponto  $i=7$ , pela sua propensão de gastos serem maiores nesta loja.

#### 4.2.3 Avaliação dos Parâmetros

Os parâmetros que foram avaliados são: o  $fd$ , conforme já mencionado, utilizado para o cálculo de avaliação de soluções; o  $h$ , coeficiente de granulometria.

Na FIGURA 15 apresenta-se soluções relacionando-se seus valores de *fitness* e diversidade, e empiricamente selecionou-se o parâmetro  $fd=0,5$ , por considerar mais soluções com maior diversidade porém não descartando as soluções com maiores *fitness*.

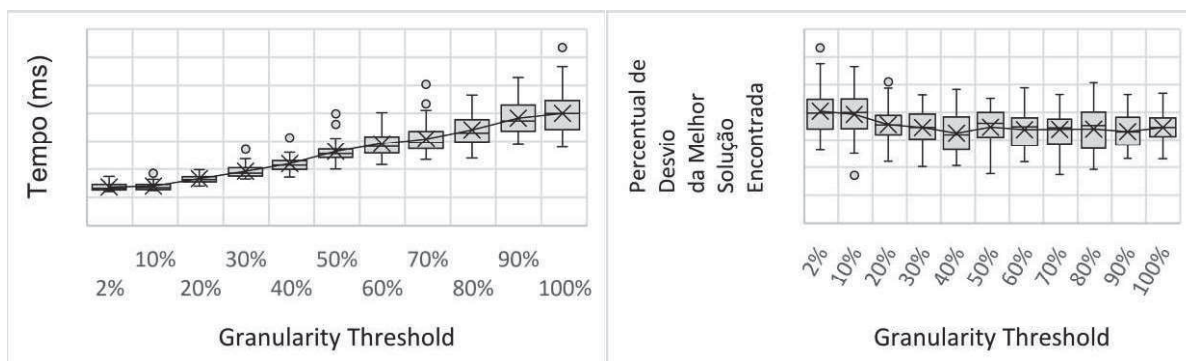
FIGURA 40 - EXEMPLO DA INFLUÊNCIA DO FATOR DE DIVERSIDADE EM RELAÇÃO A ESCOLHA DE INDIVÍDUOS PARA A POPULAÇÃO.



FONTE: A autora (2020).

A fim de investigar a influência do coeficiente de granulometria ou *granularity threshold* ( $h$ ) no tempo computacional e na qualidade da solução, realizou-se testes com  $h$  igual a 2% a 100% distribuídos de acordo com o que se apresenta na FIGURA 41. Os testes foram realizados aplicando-se o procedimento de Melhoria de Rota em soluções geradas de forma aleatória. Observa-se que o tempo de solução cresce linearmente com  $h$ . Por outro lado, não se pode afirmar em relação a qualidade da solução, que acaba estagnando independente do aumento de  $h$ . Assim, utilizou-se um valor de  $h=20\%$ , pois constatou-se uma vantagem em termos de tempo computacional em relação a coeficientes superiores que não garantem soluções melhores.

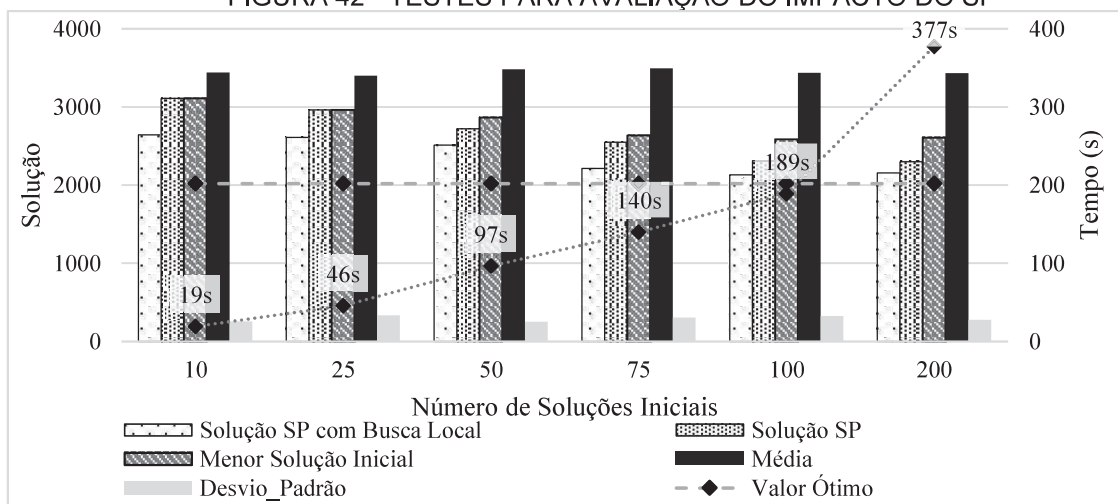
FIGURA 41 - IMPACTO DO COEFICIENTE DE GRANULOMETRIA EM TERMOS DE QUALIDADE DE SOLUÇÃO E TEMPO COMPUTACIONAL



FONTE: A autora (2020).

Com a finalidade de avaliar os resultados com o SP foram realizados testes com e sem este mecanismo. Para os testes, considerou-se a inicialização de um conjunto de soluções de tamanho  $n$ . Seguido do SP a fim de avaliar a interferência do número de soluções. Assim, a melhor solução obtida para um conjunto de soluções de tamanho  $n$  (com  $n = 10, 25, 50, 75, 100$  e  $200$ ), foi armazenada (“Menor Solução Inicial” na Figura 4), bem como a média e o desvio-padrão das soluções. Em seguida, aplicou-se o SP para a verificação da nova solução. Os valores da solução SP e da solução após as técnicas de Melhoria de Solução foram descritos na FIGURA 42. O que se pode observar é que a utilização do SP traz consideráveis melhorias para as soluções do problema, independentemente do número de soluções geradas.

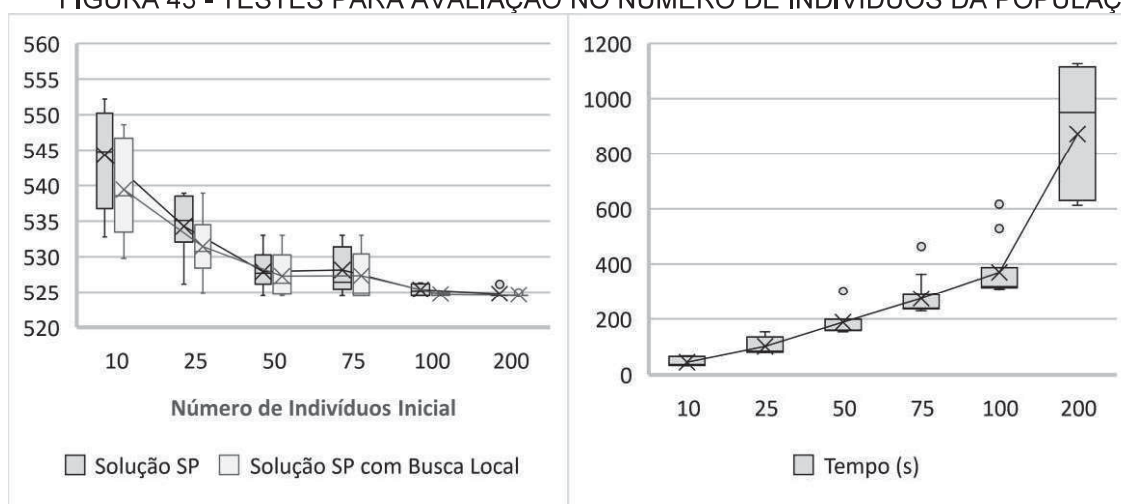
FIGURA 42 - TESTES PARA AVALIAÇÃO DO IMPACTO DO SP



FONTE: A autora (2020).

Os mesmos testes foram replicados 10 vezes a fim de se observar a dispersão em torno das soluções obtidas e em relação ao tempo computacional de acordo com o número de soluções iniciais (FIGURA 43).

FIGURA 43 - TESTES PARA AVALIAÇÃO NO NÚMERO DE INDIVÍDUOS DA POPULAÇÃO.



FONTE: A autora (2020).

Observou-se que, a partir de um determinado tamanho do conjunto de soluções  $n$ , a melhoria na solução passa a ser relativamente não compensatória quando comparado ao tempo de execução necessário.

#### 4.2.4 Resultados Testes Computacionais

O HGA-SP proposto é avaliado para um total de 10 registros de soluções, em um prazo de 3600s. Os resultados matemáticos deste algoritmo são então, comparados com os resultados obtidos com a execução do modelo de PLIM em 48600s. Os resultados detalhados obtidos são relatados nas TABELA 25 para o problema HMDPCOMVRP e TABELA 26 para a sua versão de maximização, e organizados da seguinte maneira. A primeira coluna contém o nome da instância, da qual o número de clientes, número de depósitos, padrão de dias de atendimento do cliente e padrão de veículos são relatados na TABELA 24.

Na coluna 2 apresenta-se o *ZBound* que representa o limitante do valor da solução ótima do problema inteiro. Ao lidar com o problema de minimização *ZBound* representa limitante inferior para o problema inteiro, ou ainda um *lower bound*, o que significa que o valor da solução do problema inteiro não será inferior a esse

determinado valor já conhecido. Para o problema de maximização, analogamente, *ZBound* é o limitante superior, ou *upper bound*, quando a solução ótima inteira não é maior que esse valor. Na coluna 3, apresenta-se a melhor solução (BKS – *Best Known Solution*) obtida pelo modelo dentro do prazo. A coluna 4 mostra a diferença percentual entre BKS e *ZBound*, para calcular o *Gap* utilizou-se a expressão (86).

$$Gap(\%) = \left| \frac{Z - Zbound}{Z} \right| * 100 \quad (86)$$

Para o HGA-SP, os resultados foram apresentados da seguinte forma, na coluna 5 e 6 são apresentadas as melhores soluções encontradas, e suas diferenças percentuais, respectivamente. Em seguida, apresenta-se os valores médios obtidos e o seu *gap* em relação a *Zbound*. Em negrito destacam-se os *gaps* que superaram o modelo matemático.

TABELA 25 – RESULTADOS OBTIDOS COM MODELO MATEMÁTICO PLIM E COM O HGA-SP PARA O HMDPCOMVRP

Instância	Modelo Matemático Solver		Algoritmo HGASP				
	ZBound	BKS	Gap (%)	Melhor Solução Encontrada	Gap (%)	Média	GAP (%) médio
n20d3p1v1	63598,5818	63598,6	0,000	63598,6	<b>0,000</b>	63674,84	0,12
n20d3p1v2	71425,7644	74212,149	3,75	74212,1	3,90	74284,54	4,00
n20d3p2v1	76127,331	76127,331	0,000	76127,3	<b>0,000</b>	76127,3	<b>0,000</b>
n20d3p2v2	86311,4051	87026,7653	0,822	87026,8	0,829	87026,8	0,83
n20d5p1v1	61277,4711	63598,5912	3,65	63598,6	3,79	64486,95	5,24
n20d5p1v2	66751,9871	82446,2348	19,04	72754,4	<b>8,99</b>	72754,4	<b>8,99</b>
n20d5p2v1	75298,1492	75303,098	0,006	75303,1	<b>0,006</b>	75311,99	<b>0,02</b>
n20d5p2v2	78759,3501	85257,1545	7,62142	84315,7	<b>7,05</b>	84415,45	<b>7,18</b>
n50d3p1v1	-84512,9979	-	-	112590	<b>233,22</b>	117811,5	<b>239,40</b>
n50d3p1v2	95424,68	174661,621	45,36597	124249	<b>30,21</b>	131354	<b>37,65</b>
n50d3p2v1	89621,2552	138296,53	35,19631	119319	<b>33,14</b>	120916,4	<b>34,92</b>
n50d3p2v2	102477,745	179211,561	42,82	137606	<b>34,28</b>	142410,8	<b>38,97</b>
n50d5p1v1	-82241,5514	-	-	108801	<b>232,29</b>	122126	<b>248,50</b>
n50d5p1v2	-85550,2346	-	-	124995	<b>246,11</b>	136162,3	<b>259,16</b>
n50d5p2v1	87172,0984	164276,87	46,93587	120884	<b>38,67</b>	125216,6	<b>43,64</b>
n50d5p2v2	99278,9818	192429,147	48,40751	135523	<b>36,51</b>	140647,4	<b>41,67</b>

FONTE: A autora (2020). Nota: em negrito, apresentam-se os *gaps* em que o HGA-SP apresentou resultados melhores ou iguais ao *solver*.

TABELA 26 – RESULTADOS OBTIDOS COM MODELO MATEMÁTICO PLIM E COM O HGAS-SP PARA O HMDPCOMVRP COM VERSÃO ESTENDIDA

Instância	Modelo Matemático (MIP Solver)			Algoritmo HGASP			
	ZBound	BKS	Gap (%)	Melhor Solução Encontrada	Gap (%)	Média	GAP (%) médio
n20d3p1v1	4936,4284	4902,93561	0,68	4914,74	<b>0,44</b>	4892,691	0,89
n20d3p1v2	7191,64884	9319,79743	177,17	3900,8	<b>45,76</b>	3869,547	<b>46,19</b>
n20d3p2v1	7629,71252	7165,72422	6,48	7165,72	<b>6,08</b>	7165,72	<b>6,08</b>
n20d3p2v2	7620,93524	6523,83881	16,82	7169,35	<b>5,93</b>	7139,402	<b>6,32</b>
n20d5p1v1	4938,95337	4708,06111	0,05	4868,62	1,42	4868,45	1,43
n20d5p1v2	7211,48758	-9023,32437	179,94	3874,52	<b>46,27</b>	3856,99	<b>46,52</b>
n20d5p2v1	8453,28749	-7581,5743	2,12	7246,72	14,27	7246,72	14,27
n20d5p2v2	8456,71209	-7171,76268	217,99	7169,68	<b>15,22</b>	7161,382	<b>15,32</b>
n50d3p1v1	-4915,30442	-	-	-7652,55	<b>55,69</b>	-7674,875	<b>56,14</b>
n50d3p1v2	-7284,36903	-	-	-8718,17	<b>19,68</b>	-8744,889	<b>20,05</b>
n50d3p2v1	9910,82474	-	-	9751,06	<b>1,61</b>	9692,831	<b>2,20</b>
n50d3p2v2	14,628	-	-	7714,42	<b>47,26</b>	7677,632	<b>47,51</b>
n50d5p1v1	4916,78819	-	-	-7687,33	<b>256,35</b>	-7709,221	<b>256,79</b>
n50d5p1v2	-7296,95236	-	-	-8749,1	<b>19,90</b>	-8791,865	<b>20,49</b>
n50d5p2v1	-9922,02584	-	-	9686,91	<b>197,63</b>	9632,928	<b>197,09</b>
n50d5p2v2	-14634,9977	-	-	7794,06	<b>153,26</b>	7636,552	<b>152,18</b>

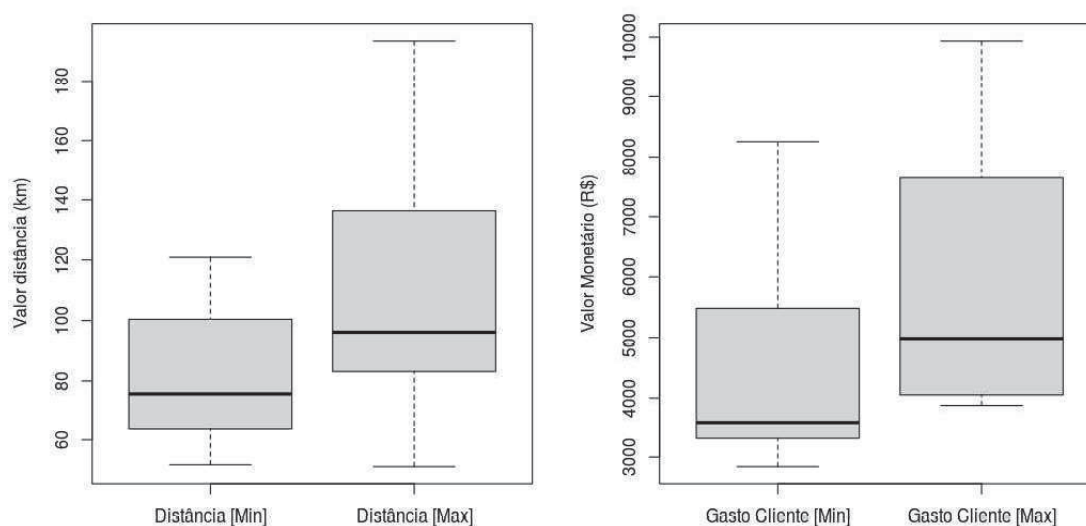
FONTE: A autora (2020). Nota: em negrito, apresentam-se os *gaps* em que o HGA-SP apresentou resultados melhores ou iguais ao *solver*.

De um modo geral, observou-se que na TABELA 25 e TABELA 26 para instâncias menores com 20 clientes, o modelo matemático de programação inteira mista resolvido com o Solver apresenta algumas soluções com *gaps* melhores que o algoritmo HGA-SP. Porém, instancias maiores, com 50 clientes, são mais difíceis de serem resolvidas pelo modelo e as mais desafiadoras para o algoritmo.

Pôde-se observar também, que algumas soluções são negativas, o que pode representar um prejuízo para a empresa, como também um resultado da penalização em relação ao uso de veículos além da sua disponibilidade. Essa situação normalmente acontece ao se fazer uso de frota terceirizada, visto que dessa forma, os benefícios adquiridos com a entrega podem não ser suficientes para cobrir os custos fixos de terceirização de veículos, ou ao se utilizar mais veículos do que os disponíveis em ambas as frotas.

Os resultados relatados mostraram o desempenho de alta qualidade da HGA-SP. Afinal, para o HMDPCOMVRP obtêm-se uma melhoria percentual média de 75% das soluções do modelo. E de 81,25% considerando-se as melhores soluções obtidas. Já para a versão estendida, esses valores são ainda melhores, superando o modelo em 87,5% considerando-se as melhores soluções e 81,25% considerando-se as soluções médias.

FIGURA 44 - BOXPLOT COM DISTRIBUIÇÃO DE DISTÂNCIA E VALOR MONETÁRIO EM RELAÇÃO AS SOLUÇÕES POR MODELO



FONTE: A autora (2020).

A FIGURA 44 mostra os valores de distância percorrida por rota de acordo com a versão do problema: considerando a minimização de distância ou a maximização de benefícios em virtude de depósitos heterogêneos.

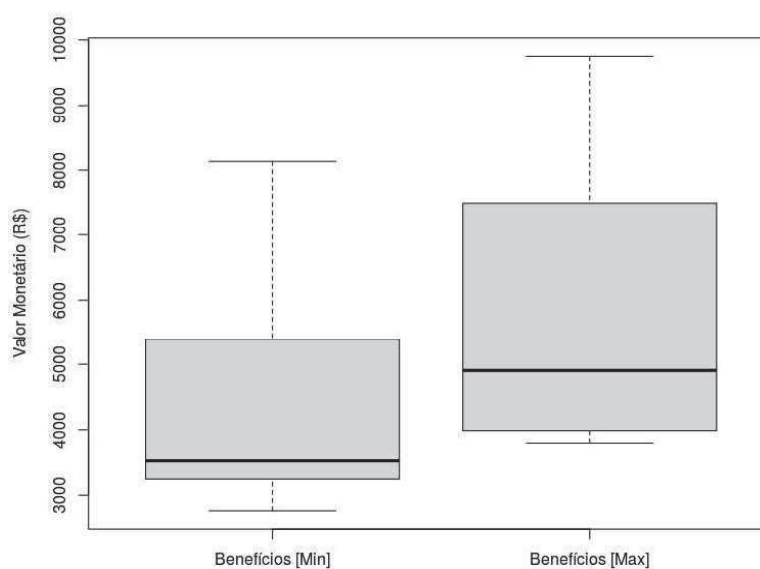
Esta análise é realizada a fim de analisar o comportamento associados às soluções que consideram apenas os custos de distribuição (estratégia de Minimização (Min)) e aquelas que consideram também as preferências do cliente (gastos previstos) (estratégia de Maximização (Max)). Neste último caso, pode-se perceber que o primeiro quartil da distância inicia na mesma posição do primeiro quartil da estratégia de minimização, indicando que mesmo maximizando benefícios pode-se ter soluções com distâncias menores. No entanto, ao se observar o comportamento dos gastos dos clientes, nota-se que o primeiro quartil inicia acima da mediana de gastos de cliente para a estratégia de minimização. Esse gasto maior é revertido em benefícios



ao grupo varejista, visto que estes estão relacionados às aquisições dos clientes, além dos custos logísticos envolvidos.

Os benefícios obtidos em cada solução também foram descritos na FIGURA 45. Os benefícios estão relacionados aos custos de distribuição e aos rendimentos esperados.

FIGURA 45 - BOXPLOT COM DISTRIBUIÇÃO DE BENEFÍCIOS OBTIDOS DAS SOLUÇÕES POR MODELO



FONTE: A autora (2020).

O que se pôde observar é que para a versão estendida do problema, os benefícios são estritamente positivos. Esse fato é um forte indício de que há um *trade-off* entre custos logísticos e receitas obtidas. Isto é, por vezes, são necessários custos de distribuição mais altos para obter um aumento nos rendimentos esperados.

## 5 CONCLUSÕES

Neste trabalho apresentou-se uma metodologia que associa estratégias de segmentação de mercado, por meio do processo KDD com o problema de roteamento de veículos. Para a primeira etapa, utilizou-se técnicas de agrupamento e técnicas preditivas. Já para o planejamento de rotas elaborou-se um modelo matemático de programação linear inteira mista bem como uma versão expandida deste modelo considerando-se depósitos heterogêneos. Devido à sua complexidade, apresentou-se uma meta-heurística híbrida com particionamento de conjuntos (HGA-SP). A abordagem proposta foi ilustrada por meio da sua aplicação em caso real de um grupo varejista brasileiro de lojas do segmento de super e hipermercados.

A fim de evidenciar as características dos problemas que têm sido mais populares em estudos sobre o VRP, realizou-se uma revisão sistemática da literatura contemplando 229 artigos de periódicos internacionais. Os resultados indicaram que a maioria dos problemas utilizam meta-heurísticas como método de solução, com métodos exatos e heurísticas clássicas sendo aplicadas com menos frequência. Outro ponto, refere-se às características dos dados, a maioria dos autores usa dados artificiais para testar seus métodos de solução, enquanto 1,8% dos artigos utilizaram dados reais ou uma combinação de dados artificiais e reais. O COMVRP foi evidenciado como uma oportunidade de estudo devido à literatura ainda escassa.

Ainda relacionado à perspectiva teórica, o presente estudo também tratou sobre as estratégias promissoras relacionadas a variantes do VRP com múltiplos atributos. Evidenciaram-se escolhas relacionadas ao espaço de busca, representação da solução e vizinhanças utilizadas por algoritmos estado-da-arte. Neste contexto, os algoritmos em destaque em suas estratégias são: UHGS (VIDAL et al., 2014), HEA (KOÇ et al., 2015) e HILS-RVRP (PENNA et al., 2017). Sendo que o primeiro deles constituiu um grande marco para este campo de pesquisa pela qualidade das soluções obtidas, e pela quantidade de variantes contempladas. Os métodos bem-sucedidos em MAVRP não são determinados por um único fator, mas são o resultado de um bom equilíbrio entre vários elementos da metodologia: o uso de diferentes espaços de busca, vizinhanças variáveis, *trade-off* entre a diversificação e intensificação, e hibridização por exemplo. Devido a tendência de se trabalhar com um número de variantes crescente e métodos de solução capazes de abordar esses

atributos simultaneamente, este estudo é pertinente. Além de ser um importante passo para o direcionamento do método unificado proposto nesta tese, de forma a lidar eficientemente com as aplicações práticas de roteamento.

Em respeito ao aspecto metodológico, a indicação de uma abordagem sistemática capaz de analisar o perfil de um cliente é importante tanto para direcionar os estímulos mais adequados a cada perfil, alavancando as vendas, quanto para extração de outras informações importantes no direcionamento de estratégias de *marketing*. Para que isso seja realizado, o conhecimento armazenado em bases de dados tem se mostrado ser um elemento fundamental para o direcionamento das pesquisas. Assim, a primeira parte desta tese apresentou por meio do processo de KDD, uma metodologia capaz de extrair informações de um conjunto de lojas distintas com uma variedade significativa de produtos comercializados.

As conclusões obtidas com cada uma das técnicas aplicadas serão expostas na sequência. Por meio da análise exploratória de dados, buscou-se identificar dentre as variáveis armazenadas em sua base, aquelas que, tratadas estatisticamente e por meio de técnicas de agrupamento e predição permitam prever atitudes de consumo, com vistas às características de similaridades que segregam esse mercado em grupos. A pesquisa realizada partiu do pressuposto de que é possível prever o gasto gerado pelos clientes em determinada loja, considerando-se suas características e seu perfil de consumo. Considerou-se como variável resposta (dependente) o gasto realizado pelo consumidor e como variáveis explicativas as características dos clientes, como idade, sexo, endereço, renda, preferências e distância do cliente até a loja. Alguns pontos fortes do uso do processo apresentado, permitindo que ele seja aplicada em outras situações, são o fato de que os dados podem ser facilmente obtidos como, por exemplo, em ambientes de comércio eletrônico, onde os clientes precisam se registrar antes de comprar; e, também, a utilização de *software* livre como o R para a aplicação das técnicas.

Como resultados, primeiramente, organizou-se uma classificação de 11 tipos de itens consumidos, por meio da AF. Com estas informações de consumo e relacionadas ao porte das lojas, identificou-se por meio da AC, que as 22 lojas se enquadram em 5 grupos (classes) de lojas distintas. Para cada classe de lojas, foram avaliados os modelos preditivos de gastos por meio dos métodos de RLM e RNAs. De modo geral, observou-se que os modelos preditivos obtidos com as RNAs

apresentaram desempenho superior. Mais especificamente, para as RNAs, realizou-se o teste de 900 configurações distintas com duas camadas, selecionando-se a aquela que forneceu o menor erro no conjunto de testes.

Observou-se que por mais instigantes que os resultados fossem, eles encontram-se limitados pela impossibilidade de se compreender, em toda a sua dimensão, as razões que levam alguém a gastar. Além disso, é importante observar que outras variáveis adicionais como, número de filhos, escolaridade, além de variáveis psicográficas como interesses, opiniões, necessidades e valores, também são importantes na determinação do gasto dos clientes. Atualmente, essas variáveis não são levantadas pela empresa analisada, mas assim que o sejam, poderiam também ser levadas em consideração no modelo para melhorar a sua previsão. Tais limitações, ao mesmo tempo, recomendam cautela com os resultados do presente trabalho, e encorajam o aprofundamento da investigação e o desenvolvimento de futuras pesquisas, com a avaliação de outras técnicas de processamento.

Quanto aos resultados obtidos com o planejamento de rotas, inicialmente, desenvolveu-se um modelo matemático descrito por HMDPCOMVRP. Este modelo matemático é capaz de lidar com restrições complexas características do mercado varejista. O HMDPCOMVRP incorpora diversas características que se adequam ao segmento de redes varejistas, como por exemplo, a possibilidade de se ter uma frota própria e terceirizada, que tenha característica de frota heterogênea. Além destas considerou-se no modelo clássico de minimização, a possibilidade de o cliente escolher o dia em que deseja ser atendido e também múltiplas lojas pertencentes à rede. Uma versão expandida deste problema é proposta, incorporando a ideia de que as lojas varejistas podem ser heterogêneas em termos de suas ofertas comerciais. Portanto, esta versão incorpora resultados da segmentação de mercado realizada pelo processo KDD, devido a necessidade de avaliar as preferências dos clientes.

Para a solução do problema, resolveu-se o modelo matemático e devido à complexidade do problema, propôs-se um algoritmo genético híbrido com Particionamento de Conjuntos denominado HGA-SP para a sua solução. Os resultados relatados mostram o desempenho de alta qualidade do HGA-SP. Afinal, para o HMDPCOMVRP obtêm-se uma melhoria percentual média acima de 75% em relação às soluções do modelo exato. Já para a versão estendida, esses valores são

ainda melhores, superando o modelo em 87,5% considerando-se as melhores soluções e 81,25% considerando-se as soluções médias.

Outra conclusão importante é obtida analisando-se o comportamento associados às soluções que consideram apenas os custos de distribuição (estratégia de minimização) e aquelas que consideram também as preferências do cliente (gastos previstos) (estratégia de maximização). O que se pode observar é que para esta versão estendida do problema, os benefícios são estritamente positivos. Esse fato indica que há um *trade-off* entre custos logísticos e receitas obtidas. Isto é, por vezes, são necessários custos de distribuição mais altos para obter um aumento nos rendimentos esperados.

Por fim, quanto ao aspecto prático ou gerencial, este trabalho demonstrou que é possível utilizar a tecnologia para explorar bancos de dados para obter informações específicas e *insights* sobre perfil, características e preferências do comportamento de consumo dos clientes. Os resultados encontrados indicaram que lojas varejistas podem se beneficiar da metodologia apresentada de forma a obter informações úteis e uma maior vantagem competitiva ao projetar soluções e produtos mais adequados às necessidades de seus clientes.

A fim de aumentar a sua competitividade no mercado, as empresas precisam ser capazes de transformar dados em informações estratégicas assim como precisam entender e ajustar continuamente seus relacionamentos com os clientes. Assim, a integração destas estratégias com o planejamento de rotas eficiente, pode permitir não apenas uma economia de custos operacionais, como também um aumento de receita. Além disso, pode-se ter um impacto positivo na taxa de entrega dos produtos, afinal, considera-se os dias em que o cliente pode receber o produto, evitando-se perder a viagem ou até mesmo a rejeição de produtos por uma entrega mal sucedida.

O modelo exige que o cliente seja atendido em um determinado dia conforme solicitado. Tal restrição pode acarretar casos extremos a contratação de frota terceirizada para atender a demanda de determinado dia. Como sugestões de trabalhos futuros, pode-se incorporar penalizações para o não cumprimento de prazos de atendimento.

Outra sugestão de trabalhos futuros está relacionada à adição de restrições de carregamentos ao modelo. Afinal, os custos relacionados à distribuição de produtos, principalmente do *e-commerce*, têm relação direta não apenas com a distância percorrida, mas também com o peso/volume transportado, assim, quanto mais pesado/volumoso for o produto e mais longe for a entrega, mais caro será o custo de transporte.

## REFERÊNCIAS

AAKER, D. A.; KUMAR, V.; DAY, G. S. Pesquisa de marketing. São Paulo: Atlas. eating in Brazil. **Rev. Saude Publica**, v. 44, n. 2, p. 221–229, 2001.

ABRAS. **SETOR SUPERMERCADISTA FATURA R\$ 355,7 BILHÕES EM 2018. Elaborada por Redação Portal ABRAS.**

AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. **Mining association rules between sets of items in large databases.** Proceedings of the 1993 ACM SIGMOD international conference on Management of data. **Anais...**1993

AHMADIZAR, F.; ZEYNIVAND, M.; ARKAT, J. Two-level vehicle routing with cross-docking in a three-echelon supply chain: A genetic algorithm approach. **Applied Mathematical Modelling**, v. 39, n. 22, p. 7065–7081, 2015.

AMALIA, A. D. H.; SUPRAJITNO, H.; PRATIWI, A. B. Solving Close-Open Mixed Vehicle Routing Problem Using Bat Algorithm. **Contemporary Mathematics and Applications**, v. 2, n. 1, p. 46–56, 2020.

ANSARI, A.; RIASI, A. Customer clustering using a combination of fuzzy c-means and genetic algorithms. **International Journal of Business and Management**, v. 11, n. 7, p. 59, 2016.

ARAÚJO, C. E. D. G. Algoritmos Genéticos Híbridos sem Delimitadores de Rotas para Problemas de Roteirização de Veículos. 2008.

ARCHETTI, C.; FERNÁNDEZ, E.; HUERTA-MUÑOZ, D. L. The flexible periodic vehicle routing problem. **Computers & Operations Research**, v. 85, p. 58–70, 2017.

ARCHETTI, C.; SPERANZA, M. G. Vehicle routing problems with split deliveries. **International transactions in operational research**, v. 19, n. 1–2, p. 3–22, 2012.

AVCI, M.; TOPALOGLU, S. A hybrid metaheuristic algorithm for heterogeneous vehicle routing problem with simultaneous pickup and delivery. **Expert Systems with Applications**, v. 53, p. 160–171, jul. 2016.

AZADEH, A.; FARROKHI-ASL, H. The close–open mixed multi depot vehicle routing problem considering internal and external fleet of vehicles. **Transportation Letters**, v. 11, n. 2, p. 78–92, 2019.

BACH, T. M. et al. Online customer behavior: perceptions regarding the types of risks incurred through online purchases. **Palgrave Communications**, v. 6, n. 1, p. 1–12, 2020.

BAI, R. et al. A set-covering model for a bidirectional multi-shift full truckload vehicle routing problem. **Transportation Research Part B: Methodological**, v. 79, p. 134–148, 2015.

BALDACCI, R. et al. An exact algorithm for the period routing problem. **Operations research**, v. 59, n. 1, p. 228–241, 2011.

BALDACCI, R.; BATTARRA, M.; VIGO, D. Routing a heterogeneous fleet of vehicles. In: **The vehicle routing problem: latest advances and new challenges**. [s.l.] Springer, 2008. p. 3–27.

BEANE, T. P.; ENNIS, D. M. Market segmentation: a review. **European journal of marketing.**, v. 21, n. 5, p. 20–42, 1987.

BELHAIZA, S. A Game Theoretic Approach for the Real-Life Multiple-Criterion Vehicle Routing Problem with Multiple Time Windows. **IEEE Systems Journal**, 2016.

BELTRAMI, E. J.; BODIN, L. D. Networks and vehicle routing for municipal waste collection. **Networks**, v. 4, n. 4, p. 65–94, 1974.

BORTFELDT, A. et al. Hybrid algorithms for the vehicle routing problem with clustered backhauls and 3D loading constraints. **European Journal of Operational Research**, v. 243, n. 1, p. 82–96, 2015.

BOUDIA, M.; PRINS, C.; REGHIOUI, M. **An effective memetic algorithm with population management for the split delivery vehicle routing problem**. International Workshop on Hybrid Metaheuristics. **Anais...Springer**, 2007

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 26, n. 2, p. 211–243, 1964.

BRAEKERS, K.; RAMAEKERS, K.; VAN NIEUWENHUYSE, I. The vehicle routing problem: State of the art classification and review. **Computers & Industrial Engineering**, v. 99, p. 300–313, 2016.

BRANDÃO, J. A tabu search algorithm for the heterogeneous fixed fleet vehicle routing problem. **Computers & Operations Research**, v. 38, n. 1, p. 140–151, 2011.

BRANDÃO, J. Iterated local search algorithm with ejection chains for the open vehicle routing problem with time windows. **Computers & Industrial Engineering**, v. 120, p. 146–159, 2018.

BRÄYSY, O.; GENDREAU, M. Vehicle routing problem with time windows, Part I: Route construction and local search algorithms. **Transportation Science**, v. 39, n. 1, p. 104–118, 2005.

BRITO, J. et al. An ACO hybrid metaheuristic for close-open vehicle routing problems with time windows and fuzzy constraints. **Applied Soft Computing**, v. 32, p. 154–163, jul. 2015.

BRITO, J.; EXPÓSITO, A.; MORENO, J. A. Variable neighbourhood search for close-open vehicle routing problem with time windows. **IMA Journal of Management Mathematics**, v. 27, n. 1, p. 25–38, 2016.

BRUGLIERI, M. et al. A new Mathematical Programming Model for the Green Vehicle Routing Problem. **Electronic Notes in Discrete Mathematics**, v. 55, p. 89–92, 2016.

CACERES-CRUZ, J. et al. Rich Vehicle Routing Problem: Survey. **ACM COMPUTING SURVEYS**, v. 47, n. 2, jan. 2015.

CAIGNY, A. DE; COUSSEMENT, K.; BOCK, K. W. DE. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. v. 269, p. 760–772, 2018.

CALVET, L. et al. Combining statistical learning with metaheuristics for the Multi-Depot Vehicle Routing Problem with market segmentation. **Computers and Industrial Engineering**, v. 94, p. 93–104, 2016.

CAMPBELL, A. M.; WILSON, J. H. Forty years of periodic vehicle routing problem. **Networks**, 2014.

CAMPBELL, J. F.; O'KELLY, M. E. Twenty-Five Years of Hub Location Research - Tags: TRANSPORTATION Science (Periodical) NETWORK hubs. **Transportation Science**, v. 46, n. 2, p. 153–169, 2012.

CARRABS, F.; CERULLI, R.; CORDEAU, J. F. An additive branch-and-bound algorithm for the pickup and delivery traveling salesman problem with LIFO or FIFO loading. **INFOR: Information Systems and Operational Research**, v. 45, n. 4, p. 223–238, 2007.



CARVALHO, A. C. P. DE L. F.; BRAGA, A. DE P.; LUDEMIR, T. B. **Fundamentos de redes neurais artificiais**. [s.l: s.n.].

CARVALHO, S. et al. Segmentação do mercado consumidor cativo alta e média tensão de uma distribuidora de energia elétrica : aplicação do Mapa Auto- Organizável de Kohonen para descoberta de padrões de inadimplência no setor. 2016.

CATTARUZZA, D. et al. Vehicle routing problems for city logistics. **Euro Journal on Transportation and Logistics**, v. 6, n. 1, SI, p. 51–79, mar. 2017.

CATTARUZZA, D.; ABSI, N.; FEILLET, D. Vehicle routing problems with multiple trips. **4OR**, v. 14, n. 3, p. 223–259, 2016.

CATTELL, R. B. The scree test for the number of factors. **Multivariate behavioral research**, v. 1, n. 2, p. 245–276, 1966.

CENSO. Sinopse do Censo 2010 - Pessoas Residentes - 0 a 5 anos de idade. p. 2015, 2015.

CHARNET, R. et al. Análise de modelos de regressão linear com aplicações. **Campinas: Unicamp**, 1999.

CHERKESLY, M. et al. Branch-price-and-cut algorithms for the pickup and delivery problem with time windows and multiple stacks. **European Journal of Operational Research**, v. 250, n. 3, p. 782–793, 2016.

CHRISTOFIDES, N. The vehicle routing problem. **Combinatorial optimization**, 1979.

CHU, J. C.; YAN, S.; HUANG, H.-J. A Multi-Trip Split-Delivery Vehicle Routing Problem with Time Windows for Inventory Replenishment Under Stochastic Travel Times. **Networks & Spatial Economics**, v. 17, n. 1, p. 41–68, mar. 2017.

ÇIMEN, M.; SOYSAL, M. Time-dependent green vehicle routing problem with stochastic vehicle speeds: An approximate dynamic programming algorithm. **Transportation Research Part D: Transport and Environment**, v. 54, p. 82–98, 2017.

CLARKE, G.; WRIGHT, J. W. Scheduling of vehicles from a central depot to a number of delivery points. **Operations Research**, v. 12, n. 4, p. 568–581, 1964.

CORDEAU, J.-F. et al. A guide to vehicle routing heuristics. **Journal of the Operational Research society**, v. 53, n. 5, p. 512–522, 2002.

CORDEAU, J.-F. et al. Vehicle routing. **Handbooks in operations research and management science**, v. 14, p. 367–428, 2007.

CORDEAU, J.-F. J.; LAPORTE, G.; MERCIER, A. A unified tabu search heuristic for vehicle routing problems with time windows. **Journal of the Operational Research Society**, v. 52, p. 928–936, 2001.

CORDEAU, J. F.; GENDREAU, M.; LAPORTE, G. A tabu search heuristic for periodic and multi-depot vehicle routing problems. **Networks**, v. 30, p. 105–119, 1997.

CUNHA, J. V; COELHO, A. C. Análise multivariada: para os cursos de administração, ciências contábeis e economia. FIPECAFI–Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras. **Regressão linear múltipla. São Paulo: Atlas**, p. 131–231, 2009.

DANTZIG, G. B.; RAMSER, J. H. The truck dispatching problem. **Management science**, v. 6, n. 1, p. 80–91, 1959.

- DAVIS, L. **Applying adaptive algorithms to epistatic domains**. *IJCAI. Anais...1985*
- DAYARIAN, I. et al. A branch-and-price approach for a multi-period vehicle routing problem. **Computers & Operations Research**, v. 55, p. 167–184, mar. 2015.
- DE ARMAS, J.; MELIÁN-BATISTA, B. Variable neighborhood search for a dynamic rich vehicle routing problem with time windows. **Computers & Industrial Engineering**, v. 85, p. 120–131, 2015.
- DEMIR, E.; BEKTAŞ, T.; LAPORTE, G. An adaptive large neighborhood search heuristic for the pollution-routing problem. **European Journal of Operational Research**, v. 223, n. 2, p. 346–359, 2012.
- DETTI, P.; PAPALINI, F.; DE LARA, G. Z. M. A multi-depot dial-a-ride problem with heterogeneous vehicles and compatibility constraints in healthcare. **OMEGA-International Journal of Management Science**, v. 70, n. 1, p. 1–14, 2017.
- DÍAZ, A. et al. Journal of Retailing and Consumer Services A segmentation study of cinema consumers based on values and lifestyle. **Journal of Retailing and Consumer Services**, v. 41, n. August 2017, p. 79–89, 2018.
- DOMINGUEZ, O. et al. A Biased-Randomised Large Neighbourhood Search for the two-dimensional Vehicle Routing Problem with Backhauls. **European Journal of Operational Research**, v. 255, n. 2, p. 442–462, 2016.
- DONG, Z.; TURNQUIST, M. A. Combining service frequency and vehicle routing for managing supplier shipments. **Transportation Research Part E: Logistics and Transportation Review**, v. 79, p. 231–243, jul. 2015.
- DROR, M.; TRUDEAU, P. Savings by split delivery routing. **Transportation Science**, v. 23, n. 2, p. 141–145, 1989.
- DUHAMEL, C.; LACOMME, P.; TOUSSAINT, H. A GRASP x ELS for the vehicle routing problem with three-dimensional loading constraints. 2011.
- EHMKE, J. F.; CAMPBELL, A. M.; THOMAS, B. W. Vehicle routing to minimize time-dependent emissions in urban areas. **European Journal of Operational Research**, v. 251, n. 2, p. 478–494, jun. 2016.
- EKSIOGLU, B.; VURAL, A. V; REISMAN, A. The vehicle routing problem: A taxonomic review. **Computers & Industrial Engineering**, v. 57, n. 4, p. 1472–1483, 2009.
- ENDERER, F.; CONTARDO, C.; CONTRERAS, I. Integrating dock-door assignment and vehicle routing with cross-docking. **Computers & Operations Research**, v. 88, p. 30–43, 2017.
- ERDOGAN, G. **VRP Spreadsheet Solver.**, 2013.
- ERDOĞAN, G. An open source Spreadsheet Solver for Vehicle Routing Problems. **Computers & Operations Research**, v. 84, p. 62–72, 2017.
- ESCOBAR, J. W. et al. A hybrid granular tabu search algorithm for the multi-depot vehicle routing problem. **Journal of heuristics**, v. 20, n. 5, p. 483–509, 2014.
- FÁVERO, L. P. L. et al. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: [s.n.].
- FAYYAD, U. M. et al. **Advances in knowledge discovery and data mining**. Califórnia: American Association for Artificial Intelligence, 1996.
- FEILLET, D. et al. An exact algorithm for the elementary shortest path problem with resource

constraints: Application to some vehicle routing problems. **Networks**, v. 44, n. 3, p. 216–229, 2004.

FISHER, M. L.; JAIKUMAR, R. A generalized assignment heuristic for vehicle routing. **Networks**, v. 11, n. 2, p. 109–124, 1981.

FLESZAR, K.; OSMAN, I. H.; HINDI, K. S. A variable neighbourhood search algorithm for the open vehicle routing problem. **European Journal of Operational Research**, v. 195, n. 3, p. 803–809, 2009.

FOEDERMAYR, E. K.; DIAMANTOPOULOS, A. Market Segmentation in Practice : Review of Empirical Studies , Methodological Assessment , and Agenda for Future Research. v. 4488, 2008.

FRANCIS, P. M.; SMILOWITZ, K. R.; TZUR, M. The period vehicle routing problem and its extensions. In: **The vehicle routing problem: latest advances and new challenges**. [s.l.] Springer, 2008. p. 73–102.

FRANCOIS, V. et al. Large neighborhood search for multi-trip vehicle routing. **European Journal of Operational Research**, v. 255, n. 2, p. 422–441, 2016.

FU, L.-L.; ALOULOU, M. A.; TRIKI, C. Integrated production scheduling and vehicle routing problem with job splitting and delivery time windows. **International Journal of Production Research**, v. 55, n. 20, p. 5942–5957, 2017.

FUKASAWA, R.; HE, Q.; SONG, Y. A branch-cut-and-price algorithm for the energy minimization vehicle routing problem. **Transportation Science**, v. 50, n. 1, p. 23–34, 2016.

GENDREAU, M. et al. Metaheuristics for the vehicle routing problem and its extensions: A categorized bibliography. In: **The vehicle routing problem: latest advances and new challenges**. [s.l.] Springer, 2008. p. 143–169.

GENDREAU, M.; JABALI, O.; REI, W. Stochastic vehicle routing problems. Vehicle routing: problems, methods, and application. **Society for Industrial and Applied Mathematics**, p. 213–239, 2014.

GENDREAU, M.; JABALI, O.; REI, W. Future Research Directions in Stochastic Vehicle Routing. **Transportation Science**, v. 50, n. 4, p. 1163–1173, nov. 2016.

GENDREAU, M.; TARANTILIS, C. D. **Solving large-scale vehicle routing problems with time windows: The state-of-the-art**. [s.l.] Cirrelt Montreal, 2010.

GLOVER, F. Future paths for integer programming and links to artificial intelligence. **Computers operations research**, v. 13, n. 5, p. 533–549, 1986.

GOEL, A.; IRNICH, S. An exact method for vehicle routing and truck driver scheduling problems. **Transportation Science**, v. 51, n. 2, p. 737–754, 2017.

GOLDEN, B. L. et al. The impact of metaheuristics on solving the vehicle routing problem: algorithms, problem sets, and computational results. In: **Fleet management and logistics**. [s.l.] Springer, 1998. p. 33–56.

GOLDEN, B. L.; RAGHAVAN, S.; WASIL, E. A. **The Vehicle Routing Problem: Latest Advances and New Challenges**. [s.l.: s.n.]. v. 43

GOODSON, J. C. A priori policy evaluation and cyclic-order-based simulated annealing for the multi-compartment vehicle routing problem with stochastic demands. **European Journal of Operational Research**, v. 241, n. 2, p. 361–369, mar. 2015.

GRANGIER, P. et al. A matheuristic based on large neighborhood search for the vehicle routing

problem with cross-docking. **Computers & Operations Research**, v. 84, p. 116–126, 2017.

GUIMARÃES, T. A. et al. The two-echelon multi-depot inventory-routing problem. **Computers & Operations Research**, v. 101, p. 220–233, 2019.

GUJARATI, D. N. *Dasar-dasar Ekonometrika jilid 1*. Jakarta: Penerbit, 2006.

HAIR, J. F. et al. **Análise multivariada de dados**. A. S. Sant ed. Porto Alegre: [s.n.].

HAN, H.; PONCE-CUETO, E. Waste collection vehicle routing problem: Literature review [Problema de programación de rutas en la recogida de residuos: Revisión de la literatura]. **Promet-Traffic & Transportation**, v. 27, n. 4, p. 345–358, 2015.

HAYKIN, S. **Redes neurais: princípios e prática**. [s.l: s.n.].

HEIN, F.; ALMEDER, C. Quantitative insights into the integrated supply vehicle routing and production planning problem. **International Journal of Production Economics**, v. 177, p. 66–76, jul. 2016.

HENKE, T. et al. The multi-compartment vehicle routing problem with flexible compartment sizes. **European Journal of Operational Research**, v. 246, n. 3, p. 730–743, nov. 2015.

HIZIROGLU, A. Soft computing applications in customer segmentation: State-of-art review and critique. **Expert Systems with Applications**, v. 40, n. 16, p. 6491–6507, 2013.

HOF, J.; SCHNEIDER, M.; GOEKE, D. Solving the battery swap station location-routing problem with capacitated electric vehicles using an AVNS algorithm for vehicle-routing problems with intermediate stops. **Transportation Research Part B: Methodological**, v. 97, p. 102–112, mar. 2017.

HOKAMA, P.; MIYAZAWA, F. K.; XAVIER, E. C. A branch-and-cut approach for the vehicle routing problem with loading constraints. **Expert Systems with Applications**, v. 47, p. 1–13, 2016.

HOLLAND, J. H.; HOLLAND, J. H. **Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence**. [s.l.] University of Michigan press, 1975.

HUANG, S.-H.; LIN, P.-C. Vehicle routing-scheduling for municipal waste collection system under the “Keep Trash off the Ground” policy. **OMEGA-International Journal of Management Science**, v. 55, p. 24–37, 2015.

HUERTA-MUÑOZ, D. L.; RÍOS-MERCADO, R. Z.; RUIZ, R. An iterated greedy heuristic for a market segmentation problem with multiple attributes. **European Journal of Operational Research**, v. 261, n. 1, p. 75–87, 2017.

IORI, M.; MARTELLO, S. Routing problems with loading constraints. **Top**, v. 18, n. 1, p. 4–27, 2010.

IORI, M.; RIERA-LEDESMA, J. Exact algorithms for the double vehicle routing problem with multiple stacks. **Computers & Operations Research**, v. 63, p. 83–101, nov. 2015.

IYODA, E. M. *Inteligência computacional no projeto automático de redes neurais híbridas e redes neurofuzzy heterogêneas*. 2000.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM computing surveys (CSUR)**, v. 31, n. 3, p. 264–323, 1999.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. [s.l.] Prentice hall Upper Saddle River, NJ, 2002. v. 5

JOLLIFFE, I. T. Principal components in regression analysis. In: **Principal component analysis**. [s.l.] Springer, 1986. p. 129–155.

JOZEFOWIEZ, N.; SEMET, F.; TALBI, E.-G. Multi-objective vehicle routing problems. **European journal of operational research**, v. 189, n. 2, p. 293–309, 2008.

KAISER, H. F. An index of factorial simplicity. **Psychometrika**, v. 39, n. 1, p. 31–36, 1974.

KAIWARTYA, O. et al. Multiobjective dynamic vehicle routing problem and time seed based solution using particle swarm optimization. **Journal of Sensors**, v. 2015, 2015.

KALAYCI, C. B.; KAYA, C. An ant colony system empowered variable neighborhood search algorithm for the vehicle routing problem with simultaneous pickup and delivery. **Expert Systems with Applications**, v. 66, p. 163–175, 2016.

KARAKATIĆ, S. et al. A survey of genetic algorithms for solving multi depot vehicle routing problem. **Applied Soft Computing**, v. 27, p. 519–532, 2015.

KENDALL, M. **Multivariate analysis**. [s.l.] Charles Griffin, 1975.

KHASHEI, M.; BIJARI, M. An artificial neural network (p, d, q) model for timeseries forecasting. **Expert Systems with applications**, v. 37, n. 1, p. 479–489, 2010.

KHODABANDEH, E. et al. Modelling and solution of a large-scale vehicle routing problem at GE appliances & lighting. **International Journal of Production Research**, v. 55, n. 4, p. 1100–1116, 2017.

KIANI, G. H. Determining profitable products in the retail market with consideration of cash limitation and exhibition periods. **Journal of Retailing and Consumer Services**, v. 55, p. 102079, 2020.

KIRBY, D. Is your fleet the right size? **Journal of the Operational Research Society**, v. 10, n. 4, p. 252, 1959.

KLEN, A. M. Algoritmo para agrupamento de descontinuidades em famílias baseado no Método Fuzzy K-Means. 2015.

KNIGHT, K.; HOFER, J. Vehicle scheduling with timed and connected calls: A case study. **Operational Research Quarterly**, v. 19, n. 3, p. 229–310, 1968.

KOÇ, Ç. et al. A hybrid evolutionary algorithm for heterogeneous fleet vehicle routing problems with time windows. **Computers and Operations Research**, v. 64, p. 11–27, 2015.

KOÇ, Ç. et al. Thirty years of heterogeneous vehicle routing. **European Journal of Operational Research**, v. 249, n. 1, p. 1–21, 2016.

KOÇ, Ç.; KARAOGLAN, I. The green vehicle routing problem: A heuristic based exact solution approach. **Applied Soft Computing**, v. 39, p. 154–164, 2016.

KOÇ, Ç.; LAPORTE, G. Vehicle routing with backhauls: Review and research perspectives. **Computers & Operations Research**, v. 91, p. 79–91, mar. 2018.

KOTLER, P. **Administração de marketing**. 10. ed. São Paulo: São Paulo: Prentice Hall, 2000.

KRAMER, R. H. F. R.; SUBRAMANIAN, A.; PENNA, P. H. V. Asymmetric vehicle routing problem with heterogeneous limited feet: A case study in a beverage industry. **Gestao e Producao**, v. 23, n. 1, p. 165–176, 2016.

KÜÇÜKOĞLU, I. et al. A memory structure adapted simulated annealing algorithm for a green vehicle routing problem. **Environmental Science and Pollution Research**, v. 22, n. 5, p. 3279–3297, 2015.

KUMAR, R.; GOPAL, G.; KUMAR, R. Novel crossover operator for genetic algorithm for permutation problems. **International Journal of Soft Computing and Engineering (IJSCE)**, v. 3, n. 2, p. 252–258, 2013.

KURPEL, D. V. et al. **The Exact Solutions of Several Classes of Container Loading Problems**. [s.l.] CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise ..., 2018.

KURPEL, D. V. et al. The exact solutions of several types of container loading problems. **European Journal of Operational Research**, v. 284, n. 1, p. 87–107, 2020.

LAHYANI, R. et al. A multi-compartment vehicle routing problem arising in the collection of olive oil in Tunisia. **OMEGA-International Journal of Management Science**, v. 51, p. 1–10, 2015.

LAHYANI, R.; KHEMAKHEM, M.; SEMET, F. Rich vehicle routing problems: From a taxonomy to a definition. **European Journal of Operational Research**, v. 241, n. 1, p. 1–14, 2015.

LAI, D. S. W. et al. A tabu search heuristic for the heterogeneous vehicle routing problem on a multigraph. **Transportation Research Part E: Logistics and Transportation Review**, v. 86, p. 32–52, 2016.

LALLA-RUIZ, E. et al. An improved formulation for the multi-depot open vehicle routing problem. **OR Spectrum**, v. 38, n. 1, p. 175–187, jan. 2016.

LAPORTE, G. Fifty years of vehicle routing. **Transportation Science**, v. 34, n. 4, p. 408–416, 2009.

LAPORTE, G. Scheduling issues in vehicle routing. **Annals of Operations Research**, v. 236, n. 2, p. 463–474, jan. 2016.

LAU, H. C. W. et al. A credibility-based fuzzy location model with Hurwicz criteria for the design of distribution systems in B2C e-commerce. **Computers & Industrial Engineering**, v. 59, n. 4, p. 873–886, 2010.

LEGGIERI, V.; HAOUARI, M. A practical solution approach for the green vehicle routing problem. **Transportation Research Part E: Logistics and Transportation Review**, v. 104, p. 97–112, 2017.

LI, F.; GOLDEN, B.; WASIL, E. The open vehicle routing problem: Algorithms, large-scale test problems, and computational results. **Computers & operations research**, v. 34, n. 10, p. 2918–2930, 2007.

LI, H. et al. The two-echelon time-constrained vehicle routing problem in linehaul-delivery systems considering carbon dioxide emissions. **Transportation Research Part D: Transport and Environment**, v. 49, p. 231–245, 2016.

LI, H. et al. The vehicle flow formulation and savings-based algorithm for the rollon-rolloff vehicle routing problem. **European Journal of Operational Research**, v. 257, n. 3, p. 859–869, mar. 2017.

LI, K.; TIAN, H. A two-level self-adaptive variable neighborhood search algorithm for the prize-collecting vehicle routing problem. **Applied Soft Computing**, v. 43, p. 469–479, jun. 2016.

LIN, C. et al. Survey of green vehicle routing problem: past and future trends. **Expert Systems with Applications**, v. 41, n. 4, p. 1118–1138, 2014.

LIU, R.; JIANG, Z. The close–open mixed vehicle routing problem. **European Journal of Operational Research**, v. 220, n. 2, p. 349–360, 2012.

LIU, S.; HUANG, W.; MA, H. An effective genetic algorithm for the fleet size and mix vehicle routing problems. **Transportation Research Part E: Logistics and Transportation Review**, v. 45, n. 3, p. 434–445, 2009.

LUTZ, C., & NEWLANDS, G.; LUTZ, C.; NEWLANDS, G. Consumer segmentation within the sharing economy: The case of Airbnb. **Journal of Business Research**, v. 88, p. 187–196, 2018.

MADEIRA, A. B. Green marketing mix: A case study of Brazilian retail enterprises. **Environmental Quality Management**, v. 28, n. 3, p. 111–116, 2019.

MAENNEL, D.; BORTFELDT, A. A hybrid algorithm for the vehicle routing problem with pickup and delivery and three-dimensional loading constraints. **European Journal of Operational Research**, v. 254, n. 3, p. 840–858, nov. 2016.

MAHMOUDI, M.; ZHOU, X. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state-space-time network representations. **Transportation Research Part B: Methodological**, v. 89, p. 19–42, jul. 2016.

MAISCHBERGER, M.; CORDEAU, J.-F. **Solving variants of the vehicle routing problem with a simple parallel iterated tabu search**. International Conference on Network Optimization. **Anais...** Springer, 2011

MAKNOON, Y.; LAPORTE, G. Vehicle routing with cross-dock selection. **Computers & Operations Research**, v. 77, p. 254–266, jan. 2017.

MALHOTRA, N. K. **Pesquisa de Marketing: Uma Orientação Aplicada**. [s.l.] Bookman Editora, 2001.

MANCINI, S. A real-life Multi Depot Multi Period Vehicle Routing Problem with a Heterogeneous Fleet: Formulation and Adaptive Large Neighborhood Search based Matheuristic. **Transportation Research Part C-Emerging Technologies**, v. 70, n. SI, p. 100–112, 2016.

MANLY, B. F. J.; ALBERTO, J. A. N. **Multivariate statistical methods: a primer**. [s.l.] CRC press, 2016.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate Analysis** Academic Press Inc. **London) LTD**, v. 15, p. 518, 1979.

MASMOUDI, M. A. et al. A hybrid genetic algorithm for the heterogeneous dial-a-ride problem. **Computers & Operations Research**, v. 81, p. 1–13, 2017.

MINGOTI, S. A. Análise de Componentes Principais. **MINGOTI, SA. Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada**. Belo Horizonte: UFMG, p. 59–95, 2007.

MOLENBRUCH, Y. et al. Multi-directional local search for a bi-objective dial-a-ride problem in patient transportation. **Computers & Operations Research**, v. 77, p. 58–71, 2017.

MONTOYA-TORRES, J. R. et al. A literature review on the vehicle routing problem with multiple depots. **Computers and Industrial Engineering**, v. 79, p. 115–129, jan. 2015.

MONTOYA, A. et al. A multi-space sampling heuristic for the green vehicle routing problem. **Transportation Research Part C: Emerging Technologies**, v. 70, n. SI, p. 113–128, 2016.

MOONS, S. et al. Integrating production scheduling and vehicle routing decisions at the

operational decision level: A review and discussion. **Computers and Industrial Engineering**, v. 104, p. 224–245, 2017.

MORRISON, D. F. The structure of multivariate observations: I. Principal components. **Multivariate statistical methods**, v. 2, p. 266–301, 1976.

MOU, S.; ROBB, D. J.; DEHORATIUS, N. Retail store operations: Literature review and research directions. **European Journal of Operational Research**, v. 265, n. 2, p. 399–422, 2018.

NAGATA, Y.; BRÄYSY, O.; DULLAERT, W. A penalty-based edge assembly memetic algorithm for the vehicle routing problem with time windows. **Computers & operations research**, v. 37, n. 4, p. 724–737, 2010.

NAGY, G.; SALHI, S. Heuristic algorithms for single and multiple depot vehicle routing problems with pickups and deliveries. **European Journal of Operational Research**, v. 162, n. 1, p. 126–141, 2005.

NGUEVEU, S. U.; PRINS, C.; CALVO, R. W. An effective memetic algorithm for the cumulative capacitated vehicle routing problem. **Computers & Operations Research**, v. 37, n. 11, p. 1877–1885, 2010.

NOROUZI, N.; SADEGH-AMALNICK, M.; TAVAKKOLI-MOGHADDAM, R. A time-dependent vehicle routing problem solved by improved Simulated Annealing. **Proceedings of the Romanian Academy Series A-Mathematics Physics Technical Sciences Information Science**, v. 16, n. 3, p. 458–465, 2015.

NOROUZI, N.; SADEGH-AMALNICK, M.; TAVAKKOLI-MOGHADDAM, R. Modified particle swarm optimization in a time-dependent vehicle routing problem: minimizing fuel consumption. **OPTIMIZATION LETTERS**, v. 11, n. 1, p. 121–134, jan. 2017.

OLIVEIRA-BROCHADO, A.; MARTINS, F. V. Aspectos Metodológicos da Segmentação de Mercado: base de segmentação e métodos de classificação. **RBN: Revista Brasileira de Gestão de Negócios**, v. 10, n. 27, p. 132–149, 2008.

OSMAN, I. H. Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem. **Annals of operations research**, v. 41, n. 4, p. 421–451, 1993.

PARENTE, J. **Varejo no Brasil: gestão e estratégia**. São Paulo: [s.n.].

PARTYKA J.; HALL, R. Vehicle routing software survey: VR delivers the goods. **OR/MS Today**, v. 41, n. 1, p. 40–46, 2014.

PENNA, P. H. V. et al. **New compound neighborhoods structures for the heterogeneous fixed fleet vehicle routing problem**. Proceedings of the XLV Symposium Bras. Pesqui. Operac.(Natal, Brazil, September 16–19, 2013). **Anais...2013**

PENNA, P. H. V. et al. A hybrid heuristic for a broad class of vehicle routing problems with heterogeneous fleet. **Annals of Operations Research**, p. 1–70, 2017.

PENNA, P. H. V.; SUBRAMANIAN, A.; OCHI, L. S. An iterated local search heuristic for the heterogeneous fleet vehicle routing problem. **Journal of Heuristics**, v. 19, n. 2, p. 201–232, 2013.

PILLAC, V. et al. A review of dynamic vehicle routing problems. **European Journal of Operational Research**, v. 225, n. 1, p. 1–11, 2013.

PISINGER, D.; ROPKE, S. A general heuristic for vehicle routing problems. **Computers & Operations Research**, v. 34, n. 8, p. 2403–2435, 2007.



PLA, L. E. **Análisis multivariado: método de componentes principales**. [s.l.] OEA, Washington, DC (EUA). Secretaría General. Programa Regional de ..., 1986.

POHLMANN, M. C. Análise de conglomerados. **Análise Multivariada**. São Paulo: Atlas, 2007.

POLAT, O. et al. A perturbation based variable neighborhood search heuristic for solving the Vehicle Routing Problem with Simultaneous Pickup and Delivery with Time Limit. **European Journal of Operational Research**, v. 242, n. 2, p. 369–382, 2015.

POLLARIS, H. et al. Vehicle routing problems with loading constraints: state-of-the-art and future directions. **OR Spectrum**, v. 37, n. 2, p. 297–330, mar. 2015.

PRINS, C. A simple and effective evolutionary algorithm for the vehicle routing problem. **Computers & Operations Research**, v. 31, n. 12, p. 1985–2002, 2004.

PRINS, C. Two memetic algorithms for heterogeneous fleet vehicle routing problems. **Engineering Applications of Artificial Intelligence**, v. 22, n. 6, p. 916–928, 2009.

PSARAFTIS, H. N. Dynamic vehicle routing problems. **Vehicle Routing: Methods and Studies**, v. 61, n. 1, p. 223–248, 1988.

PSARAFTIS, H. N.; WEN, M.; KONTOVAS, C. A. Dynamic vehicle routing problems: Three decades and counting. **Networks**, v. 67, n. 1, p. 3–31, jan. 2016.

PULLEN, H.; WEBB M. A computer application to a transport scheduling problem. **he Computer Journal**, v. 10, n. 1, p. 10–13, 1967.

QIAN, J.; EGLESE, R. Fuel emissions optimization in vehicle routing problems with time-varying speeds. **European Journal of Operational Research**, v. 248, n. 3, p. 840–848, 2016.

REINARTZ, W.; WIEGAND, N.; IMSCHLOSS, M. The impact of digital transformation on the retailing value chain. **International Journal of Research in Marketing**, v. 36, n. 3, p. 350–366, 2019.

REPOUSSIS, P. P. et al. A hybrid evolution strategy for the open vehicle routing problem. **Computers & Operations Research**, v. 37, n. 3, p. 443–455, 2010.

RITZINGER, U.; PUCHINGER, J.; HARTL, R. F. A survey on dynamic and stochastic vehicle routing problems. **International Journal of Production Research**, v. 54, n. 1, p. 215–231, jan. 2016.

RIVERA, J. C. et al. Mathematical formulations and exact algorithm for the multitrip cumulative capacitated single-vehicle routing problem. **European Journal of Operational Research**, v. 249, n. 1, p. 93–104, 2016.

RIVERA, J. C.; AFSAR, H. M.; PRINS, C. A multistart iterated local search for the multitrip cumulative capacitated vehicle routing problem. **COMPUTATIONAL OPTIMIZATION AND APPLICATIONS**, v. 61, n. 1, p. 159–187, 2015.

SAAB, W. G. L.; GIMENEZ, L. C. P. **Aspectos atuais do varejo de alimentos no mundo e no Brasil**. Disponível em: <[http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes\\_pt/Galerias/Arquivos/conhecimento/bnset/set1106.pdf](http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/bnset/set1106.pdf)>. Acesso em: 1 mar. 2020.

SÁNCHEZ-HERNÁNDEZ, G. et al. Ranking and selection of unsupervised learning marketing segmentation. **Knowledge-based systems**, v. 44, p. 20–33, 1 maio 2013.

SBVC. **300 maiores empresas de varejo brasileiro**. Disponível em: <[http://sbvc.com.br/o-papel-dovarejo-na-economia-brasileira-2019/o-papel-do-varejo-na-economia-brasileira-sbvc\\_2019/](http://sbvc.com.br/o-papel-dovarejo-na-economia-brasileira-2019/o-papel-do-varejo-na-economia-brasileira-sbvc_2019/)>.

Acesso em: 16 ago. 2020.

SCHAWB, A. J. **Eletronic classroom**.

SCHENEKEMBERG, C. M. et al. The two-echelon production-routing problem. **European Journal of Operational Research**, 2020a.

SCHENEKEMBERG, C. M. et al. The two-echelon inventory-routing problem with fleet management. **Computers & Operations Research**, v. 121, p. 104944, 2020b.

SCHIFFMAN, L. G.; KANUK, L. L. **Consumer behavior**. 7. ed. [s.l.] Prentice Hall, Upper Saddle River, 1991.

SCHRAGE, L. Formulation and structure of more complex/realistic routing and scheduling problems. **Networks**, v. 11, n. 2, p. 229–232, 1981.

SCHYNS, M. An ant colony system for responsive dynamic vehicle routing. **European Journal of Operational Research**, v. 245, n. 3, p. 704–718, 2015.

SILVESTRIN, P. V.; RITT, M. An iterated tabu search for the multi-compartment vehicle routing problem. **Computers & Operations Research**, v. 81, p. 192–202, 2017.

SMITH, W. R. Product differentiation and market segmentation as alternative marketing strategies. **Journal of marketing**, v. 21, n. 1, p. 3–8, 1956.

SOLOMON, M. M. Algorithms for the vehicle routing and scheduling problems with time window constraints. **Operations Research**, v. 35, n. 2, p. 254–265, 1987.

SON, L. H.; LOUATI, A. Modeling municipal solid waste collection: A generalized vehicle routing model with multiple transfer stations, gather sites and inhomogeneous vehicles in time windows. **Waste Management**, v. 52, p. 34–49, jun. 2016.

SOTO, M. et al. Multiple neighborhood search, tabu search and ejection chains for the multi-depot open vehicle routing problem. **Computers and Industrial Engineering**, v. 107, p. 211–222, 2017.

SOYSAL, M. et al. A Simulation Based Restricted Dynamic Programming approach for the Green Time Dependent Vehicle Routing Problem. **Computers & Operations Research**, v. 88, p. 297–305, 2017.

SRIKANT, R.; AGRAWAL, R. Mining generalized association rules. **Future generation computer systems**, v. 13, n. 2–3, p. 161–180, 1997.

STOCK, J. H.; WATSON, M. W. Combination forecasts of output growth in a seven-country data set. **Journal of forecasting**, v. 23, n. 6, p. 405–430, 2004.

SUBRAMANIAN, A. et al. A parallel heuristic for the vehicle routing problem with simultaneous pickup and delivery. **Computers & Operations Research**, v. 37, n. 11, p. 1899–1911, 2010.

SUBRAMANIAN, A. et al. A hybrid algorithm for the heterogeneous fleet vehicle routing problem. **European Journal of Operational Research**, v. 221, n. 2, p. 285–295, 2012.

TANG, J.; YU, Y.; LI, J. An exact algorithm for the multi-trip vehicle routing and scheduling problem of pickup and delivery of customers to the airport. **Transportation Research Part E: Logistics and Transportation Review**, v. 73, p. 114–132, jan. 2015.

TAVAKKOLI-MOGHADDAM, R. et al. **A Multi-Depot Close and Open Vehicle Routing Problem with Heterogeneous Vehicles**. 2019 International Conference on Industrial Engineering and

Systems Management (IESM). **Anais...IEEE**, 2019

TEAM, R. C. R: A language and environment for statistical computing. 2013.

TIWARI, A.; CHANG, P.-C. P.-C. A block recombination approach to solve green vehicle routing problem. **International Journal of Production Economics**, v. 164, p. 379–387, jun, 2015.

TIWARI, R. et al. Market segmentation using supervised and unsupervised learning techniques for E-commerce applications. **Journal of Intelligent & Fuzzy Systems**, v. 35, n. 5, p. 5353–5363, 2018.

TKACZYNSKI, A.; RUNDLE-THIELE, S. R. Event segmentation: A review and research agenda. **Tourism Management**, v. 32, n. 2, p. 426–434, 2011.

TOTH, P.; VIGO, D. **The vehicle routing problem**. Philadelphia, USA: Society for Industrial and Applied Mathematics, 2002a.

TOTH, P.; VIGO, D. Models, relaxations and exact approaches for the capacitated vehicle routing problem. **Discrete Applied Mathematics**, 2002b.

TOTH, P.; VIGO, D. Vehicle routing: problems, methods, and applications. **Society for Industrial and Applied Mathematics**, 2014.

TURAN, S. et al. A VNS approach to multi-location inventory redistribution with vehicle routing. **Computers & Operations Research**, v. 78, p. 526–536, 2017.

TURKENSTEEN, M. The accuracy of carbon emission and fuel consumption computations in green vehicle routing. **European Journal of Operational Research**, v. 262, n. 2, p. 647–659, 2017.

TYMCIOS, F. S. et al. Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation. **Solar energy**, v. 78, n. 6, p. 752–762, 2005.

VAROTTO, L. F. Varejo no Brasil–Resgate Histórico e Tendências. **Revista Brasileira de Marketing**, v. 17, n. 3, p. 429–443, 2018.

VIDAL, T. et al. A Hybrid Genetic Algorithm for Multidepot and Periodic Vehicle Routing Problems. **Operations Research**, v. 60, n. 3, p. 611–624, 2012.

VIDAL, T. et al. A hybrid genetic algorithm with adaptive diversity management for a large class of vehicle routing problems with time-windows. **Computers and Operations Research**, v. 40, n. 1, p. 475–489, 2013a.

VIDAL, T. et al. Heuristics for multi-attribute vehicle routing problems: A survey and synthesis. **European Journal of Operational Research**, v. 231, n. 1, p. 1–21, 2013b.

VIDAL, T. et al. A unified solution framework for multi-attribute vehicle routing problems. **European Journal of Operational Research**, v. 234, n. 3, p. 658–673, 2014.

WANG, K.; LAN, S.; ZHAO, Y. A genetic-algorithm-based approach to the two-echelon capacitated vehicle routing problem with stochastic demands in logistics service. **JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY**, v. 68, n. 11, p. 1409–1421, nov. 2017.

WANG, Z.; LI, Y.; HU, X. A heuristic approach and a tabu search for the heterogeneous multi-type fleet vehicle routing problem with time windows and an incompatible loading constraint. **Computers and Industrial Engineering**, v. 89, n. SI, p. 162–176, nov. 2015.

WASSAN, N. et al. The Multiple Trip Vehicle Routing Problem with Backhauls: Formulation and a Two-Level Variable Neighbourhood Search. **Computers & Operations Research**, v. 78, p. 454–467, 2017.

WEDEL, M.; KAMAKURA, W. A. **Market segmentation: Conceptual and methodological foundations**. [s.l: s.n.].

XIAO, Y.; KONAK, A. The heterogeneous green vehicle routing and scheduling problem with time-varying traffic congestion. **Transportation Research Part E: Logistics and Transportation Review**, v. 88, p. 146–166, 2016.

XIAO, Y.; KONAK, A. A genetic algorithm with exact dynamic programming for the green vehicle routing & scheduling problem. **JOURNAL OF CLEANER PRODUCTION**, v. 167, p. 1450–1463, nov. 2017.

YALCIN, G. D.; ERGINEL, N. Fuzzy multi-objective programming algorithm for vehicle routing problems with backhauls. **Expert Systems with Applications**, v. 42, n. 13, p. 5632–5644, 2015.

YAO, B. et al. An improved particle swarm optimization for carton heterogeneous vehicle routing problem with a collection depot. **Annals of Operations Research**, v. 242, n. 2, p. 303–320, jul. 2016.

YU, V. F.; JEWPNAYA, P.; REDI, A. A. N. P. Open vehicle routing problem with cross-docking. **Computers & Industrial Engineering**, v. 94, p. 6–17, 2016.

ZACHARIADIS, E. E.; KIRANOUDIS, C. T. An open vehicle routing problem metaheuristic for examining wide solution neighborhoods. **Computers & Operations Research**, v. 37, n. 4, p. 712–723, 2010.

ZACHARIADIS, E. E.; TARANTILIS, C. D.; KIRANOUDIS, C. T. The vehicle routing problem with simultaneous pick-ups and deliveries and two-dimensional loading constraints. **European Journal of Operational Research**, v. 251, n. 2, p. 369–386, 2016.

ZHANG, Z.; WEI, L.; LIM, A. An evolutionary local search for the capacitated vehicle routing problem minimizing fuel consumption under three-dimensional loading constraints. **Transportation Research Part B: Methodological**, v. 82, p. 20–35, 2015.

**ANEXO 1 – MEDIDAS DE POSIÇÃO REFERENTE À ETAPA DE ESTATÍSTICA DESCRITIVA DOS DADOS**

	C.1.1	C.1.2	C.1.3	C.1.4	C.1.5	C.1.6	C.1.7	C.2.1	C.2.2	C.2.3	C.2.4	C.2.5	C.3.1	C.3.2	C.3.3	C.3.4	C.4.1	C.4.2	
Min.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1st Qu.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Median	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	2,00	2,00
Mean	0,32	0,08	0,83	1,54	0,16	0,11	0,00	0,01	0,02	0,00	0,01	0,01	0,05	0,02	0,08	0,05	4,01	4,73	4,73
3rd Qu.	0,00	0,00	0,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,00	6,00	6,00
Max.	87,00	31,00	486,00	89,00	145,00	50,00	9,00	6,00	7,00	5,00	5,00	6,00	20,00	17,00	25,00	9,00	185,00	231,00	231,00

	C.4.2	C.4.3	C.4.4	C.4.5	C.4.6	C.4.7	C.5.1	C.5.2	C.5.3	C.5.4	C.5.5	C.5.6	C.5.7	C.5.8	C.5.9	C.5.10	C.5.11	C.5.12	
Min.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1st Qu.	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Median	2,00	4,00	2,00	2,00	1,00	0,00	3,00	0,00	1,00	0,00	0,00	0,00	0,00	1,00	1,00	0,00	0,00	0,00	0,00
Mean	6,26	10,02	5,50	6,05	4,17	0,44	9,55	2,43	3,61	1,96	1,79	2,16	0,32	2,33	1,83	1,15	0,38	0,77	0,77
3rd Qu.	7,00	12,00	7,00	8,00	5,00	0,00	10,00	3,00	4,00	2,00	2,00	2,00	0,00	3,00	2,00	1,00	0,00	1,00	1,00
Max.	855,00	444,00	409,00	266,00	125,00	97,00	597,00	255,00	199,00	114,00	309,00	173,00	181,00	123,00	118,00	120,00	97,00	70,00	70,00

## ANEXO 2 – CARGAS FATORIAIS

### MATRIZ COM COMPONENTES NÃO ROTACIONADOS

Variáveis	Cargas Fatoriais											Comuna- lidades
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	
C.1.1	<b>0,4</b>	0,14	0,23	0	-0,18	0,18	0	-0,04	-0,12	-0,04	0,08	<b>0,32</b>
C.1.2	0,22	0,12	0,27	-0,13	<b>-0,55</b>	-0,04	-0,04	0,12	0,04	-0,09	0,27	<b>0,55</b>
C.1.3	0,3	-0,07	0,15	-0,04	<b>-0,32</b>	0,01	-0,01	-0,04	<b>-0,19</b>	0,01	0,05	<b>0,27</b>
C.1.4	<b>0,68</b>	0,33	0,05	-0,05	-0,14	-0,05	0	-0,04	-0,13	0,01	0,04	<b>0,62</b>
C.1.5	0,22	0,09	0,24	-0,09	-0,07	0,4	-0,07	-0,16	<b>-0,26</b>	0,21	-0,11	<b>0,45</b>
C.1.6	0,24	0,06	0,22	-0,06	0,21	<b>0,55</b>	-0,05	0,2	-0,07	0,13	0,16	<b>0,55</b>
C.1.7	0,02	0,03	0,05	-0,01	-0,04	0,1	<b>0,97</b>	-0,05	0,09	0,17	-0,01	<b>0,99</b>
C.2.1	0,04	0,08	0,18	<b>0,46</b>	0,08	0,01	0,18	0,19	-0,15	<b>-0,67</b>	0,21	<b>0,84</b>
C.2.2	0,09	0,16	0,27	<b>0,45</b>	0	-0,07	-0,02	0,15	-0,04	0,15	-0,19	<b>0,4</b>
C.2.3	0,03	0,07	0,17	0,44	0,06	-0,06	-0,09	<b>-0,32</b>	0,35	0,37	0,6	<b>0,96</b>
C.2.4	0,03	0,08	0,16	<b>0,48</b>	0	-0,18	-0,03	0,42	-0,19	0,32	-0,28	<b>0,69</b>
C.2.5	0,07	0,08	0,2	0,32	0	0,21	-0,06	<b>-0,63</b>	0,12	-0,29	-0,36	<b>0,82</b>
C.3.1	0,19	0,24	0,43	-0,21	<b>-0,18</b>	<b>-0,22</b>	-0,01	0,03	0,04	-0,06	0,02	<b>0,41</b>
C.3.2	0,11	0,17	0,21	-0,18	<b>0,56</b>	0,01	0	0,1	-0,16	-0,13	0,29	<b>0,57</b>
C.3.3	0,19	0,21	<b>0,47</b>	-0,21	0,09	-0,13	0	0,03	0,32	-0,02	-0,2	<b>0,51</b>
C.3.4	0,19	0,2	0,36	-0,21	0,26	-0,12	-0,02	0,07	0,34	0,01	-0,2	<b>0,49</b>
C.4.1	<b>0,81</b>	0,17	-0,18	0,02	0,02	-0,05	0,01	-0,02	-0,02	0,01	0	<b>0,73</b>
C.4.2	<b>0,82</b>	0,17	-0,14	0	0,02	-0,09	0,01	-0,04	-0,03	0	0	<b>0,72</b>
C.4.3	<b>0,75</b>	-0,06	-0,04	0,02	0,08	0,13	0	0	-0,03	0,04	-0,01	<b>0,6</b>
C.4.4	<b>0,82</b>	0,03	-0,05	0	0,13	-0,01	0,01	-0,07	-0,11	0,05	-0,02	<b>0,71</b>
C.4.5	<b>0,82</b>	0,16	-0,19	0,02	-0,03	-0,1	0,01	-0,02	0,02	-0,01	0,01	<b>0,75</b>
C.4.6	<b>0,81</b>	0,27	-0,14	-0,01	-0,06	-0,1	0,02	-0,05	-0,1	0,01	0,03	<b>0,77</b>
C.4.7	<b>0,74</b>	0,27	-0,07	-0,04	0,06	-0,07	0,01	-0,09	-0,14	0,03	0,01	<b>0,66</b>
C.4.8	0,31	-0,08	0,03	0,05	-0,19	0,48	-0,07	0,26	0,25	-0,03	-0,05	<b>0,51</b>
C.5.1	<b>0,77</b>	-0,12	-0,06	0,02	-0,07	0	-0,01	0,09	0,19	-0,06	0	<b>0,66</b>
C.5.2	<b>0,68</b>	-0,19	-0,03	0,03	0,01	0,09	-0,01	0,12	0,19	-0,03	0	<b>0,57</b>
C.5.3	<b>0,75</b>	0,03	-0,05	0	0,06	0	0	-0,03	-0,06	0,02	-0,02	<b>0,57</b>
C.5.4	<b>0,69</b>	-0,13	-0,03	0,02	0,1	0,04	0	-0,01	0,01	0,01	-0,03	<b>0,5</b>
C.5.5	<b>0,56</b>	-0,51	0,18	0	-0,01	-0,09	0,01	0,02	0,06	-0,04	0,01	<b>0,62</b>
C.5.6	<b>0,7</b>	-0,3	0	0,04	0,05	0,03	0	0,06	0,12	-0,02	-0,02	<b>0,6</b>
C.5.7	0,25	-0,43	0,35	-0,05	0,02	-0,15	0,02	-0,16	<b>-0,29</b>	0,07	0,02	<b>0,51</b>
C.5.8	<b>0,74</b>	-0,12	-0,08	0,03	0,01	-0,06	0,01	0	0,06	-0,02	0	<b>0,58</b>
C.5.9	<b>0,8</b>	0,03	-0,12	0,02	-0,02	-0,04	0	0	0,02	-0,01	-0,01	<b>0,65</b>
C.5.10	<b>0,62</b>	-0,05	0,01	0,01	0,09	0,06	0	-0,05	-0,1	0,06	-0,05	<b>0,41</b>
C.5.11	0,39	-0,5	0,33	-0,04	0,03	-0,15	0,02	-0,07	-0,15	0,01	0,04	<b>0,56</b>
C.5.12	<b>0,67</b>	-0,12	-0,12	0,03	-0,05	0	-0,01	0,1	0,24	-0,08	-0,02	<b>0,55</b>

## MATRIZ COM COMPONENTES ROTACIONADOS COM MÉTODO QUARTIMAX

Variáveis	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	Comunidades
C.1.1	0,36	0,04	0,02	0,32	0,26	-0,01	0,02	0,12	0,08	0,02	0,03	0,32
C.1.2	0,17	0,04	0,06	0,69	0	0,12	-0,06	-0,12	0,09	0,07	-0,02	0,55
C.1.3	0,26	0,21	-0,12	0,36	0,06	-0,02	0,03	0,04	-0,03	-0,05	0,01	0,27
C.1.4	0,69	-0,11	0,06	0,28	0,09	-0,18	0,05	0,03	0,01	0	0,01	0,62
C.1.5	0,17	0,07	-0,05	0,17	0,51	-0,07	0,07	0,24	-0,21	-0,07	0,02	0,45
C.1.6	0,19	0	0,07	-0,03	0,68	0,16	0	-0,09	0,05	0,06	0,01	0,55
C.1.7	0,01	0	0,01	-0,01	0	0	-0,01	-0,01	0,02	0	1	0,99
C.2.1	0,02	0	-0,04	0,05	0	0,01	0,07	0,1	0,91	-0,01	0,03	0,84
C.2.2	0,07	-0,01	0,1	0,04	0,03	0	0,57	0,15	0,12	0,13	0,01	0,4
C.2.3	0,02	0,01	-0,01	0,02	0,01	-0,01	0,02	0,04	-0,01	0,98	0	0,96
C.2.4	0,02	0,02	0	-0,04	0	0,01	0,82	-0,1	0,04	0	-0,02	0,69
C.2.5	0,05	0,01	0,06	-0,05	-0,01	0,01	-0,04	0,89	0,1	0,06	-0,02	0,82
C.3.1	0,15	0,07	0,41	0,43	-0,01	-0,15	0,03	-0,01	0,04	-0,01	0	0,41
C.3.2	0,1	0,04	0,29	-0,23	0,36	-0,34	-0,16	-0,23	0,3	0,06	-0,07	0,57
C.3.3	0,13	0,04	0,68	0,13	0,02	0,02	0,04	0,11	-0,04	-0,01	0,04	0,51
C.3.4	0,15	-0,01	0,68	-0,06	0,05	0,01	0,03	0,04	-0,06	-0,01	0,01	0,49
C.4.1	0,84	-0,1	-0,01	0,01	-0,01	-0,06	0,01	-0,01	0	0	0	0,73
C.4.2	0,84	-0,06	0,02	0,03	-0,03	-0,08	0	0	0	0,01	0	0,72
C.4.3	0,74	0,12	-0,03	-0,07	0,16	0,09	0,01	0,03	-0,01	0	0,01	0,6
C.4.4	0,82	0,1	0,01	-0,07	0,11	-0,1	0,01	0,04	-0,02	-0,01	0,01	0,71
C.4.5	0,86	-0,09	-0,01	0,03	-0,09	-0,03	0	-0,02	0	0,02	0	0,75
C.4.6	0,84	-0,12	-0,01	0,13	-0,02	-0,17	0,02	0	-0,01	0	0,01	0,77
C.4.7	0,77	-0,08	0,04	0,05	0,07	-0,24	0,02	0,03	-0,01	-0,01	0,01	0,66
C.4.8	0,26	-0,08	0	0,12	0,28	0,58	0,02	0,03	0,02	-0,05	-0,01	0,51
C.5.1	0,75	0,11	0,06	0,04	-0,06	0,27	-0,03	-0,03	0,03	0,01	-0,01	0,66
C.5.2	0,6	0,15	-0,01	-0,06	0,14	-0,02	0,04	0,06	-0,04	-0,02	0,01	0,41
C.5.3	0,29	0,69	0,04	0,03	0,01	0	0,01	-0,01	0,02	0,01	0	0,56
C.5.4	0,66	0,05	0,05	-0,02	-0,12	0,31	-0,04	-0,03	0,03	0,01	-0,01	0,55
C.5.5	0,65	0,15	0,05	-0,05	0,03	0,33	-0,03	-0,04	0,04	0,02	0	0,57
C.5.6	0,75	0,07	0,01	-0,02	0,07	-0,02	0,01	0,02	-0,01	-0,01	0	0,57
C.5.7	0,67	0,18	0,02	-0,1	0,07	0,1	-0,01	0,02	-0,01	0	0	0,5
C.5.8	0,48	0,57	0,04	-0,02	-0,06	0,23	-0,03	-0,03	0,03	0,02	0	0,62
C.5.9	0,66	0,3	0,03	-0,1	0	0,27	-0,02	-0,01	0,02	0,01	0	0,6
C.5.10	0,17	0,67	-0,02	0,06	0,06	-0,16	0,04	0,05	-0,03	-0,01	0,01	0,51
C.5.11	0,73	0,15	0	-0,03	-0,06	0,11	-0,02	-0,01	0,01	0,02	0	0,58
C.5.12	0,8	0,02	-0,01	0,02	-0,03	0,05	0	0	0	0	0	0,65

### MATRIZ COM COMPONENTES ROTACIONADOS COM MÉTODO PROMAX

Variáveis	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	Comunidades
C.1.1	0,18	0,03	0,15	0,27	0,31	-0,02	0,09	-0,02	0,09	0,02	0,02	0,32
C.1.2	-0,06	0,03	0,25	-0,02	0,83	0,03	-0,17	-0,1	0,09	0,07	-0,03	0,55
C.1.3	0,09	0,27	0,03	0,04	0,37	-0,15	-0,01	0	-0,01	-0,04	0	0,27
C.1.4	0,75	-0,12	-0,1	0,05	0,26	0,03	-0,01	0,02	0,01	-0,01	0	0,62
C.1.5	-0,05	0,08	0,15	0,57	0,05	-0,06	0,19	0,02	-0,2	-0,06	0,01	0,45
C.1.6	-0,15	-0,09	0,47	0,82	-0,03	0,01	-0,13	0,01	0,03	0,06	0,01	0,55
C.1.7	0,01	0	-0,02	0,02	-0,04	0,03	-0,04	-0,02	0,03	0,01	1	0,99
C.2.1	-0,04	0	0,03	-0,06	0,14	-0,15	0,17	0,06	0,97	-0,06	0,03	0,84
C.2.2	0,01	-0,01	0,04	0,03	-0,04	0,11	0,11	0,58	0,11	0,1	0	0,4
C.2.3	0,01	0,01	-0,09	0,04	0,1	-0,05	0,05	-0,01	-0,07	1	0,01	0,96
C.2.4	0,01	0,04	0,02	0	-0,12	0,01	-0,19	0,88	0,03	-0,03	-0,02	0,69
C.2.5	-0,01	-0,01	0,02	-0,06	-0,22	0,12	1	-0,14	0,15	0,04	-0,03	0,82
C.3.1	0,01	0,12	-0,13	-0,08	0,47	0,4	-0,04	0	0,02	-0,02	-0,01	0,41
C.3.2	0,08	0,08	-0,33	0,36	-0,19	0,2	-0,21	-0,14	0,26	0,04	-0,06	0,57
C.3.3	-0,08	0,02	0,06	-0,02	0,12	0,73	0,12	0,02	-0,08	-0,03	0,03	0,51
C.3.4	0	-0,05	0,03	0,02	-0,07	0,72	0,06	0,03	-0,11	-0,03	0,01	0,49
C.4.1	0,96	-0,14	-0,04	-0,04	-0,01	-0,03	-0,01	0,01	-0,01	0	0	0,73
C.4.2	0,94	-0,09	-0,08	-0,07	0,01	0	0	0	-0,01	0	0	0,72
C.4.3	0,61	0,09	0,15	0,18	-0,1	-0,04	0,02	0,01	-0,02	-0,01	0,01	0,6
C.4.4	0,8	0,11	-0,1	0,08	-0,12	-0,02	0,03	0,01	-0,03	-0,01	0,01	0,71
C.4.5	0,97	-0,14	-0,03	-0,13	0,03	-0,02	-0,02	0	0	0,01	0	0,75
C.4.6	1	-0,14	-0,15	-0,06	0,1	-0,04	-0,02	0,01	-0,01	0	0	0,77
C.4.7	0,9	-0,09	-0,22	0,03	-0,01	0,01	0,01	0,01	-0,02	-0,01	0,01	0,66
C.4.8	-0,1	-0,25	0,9	0,39	0,17	0,02	0,01	0,02	0,02	-0,05	-0,01	0,51
C.5.1	0,58	0,05	0,31	-0,07	0,09	0,06	-0,03	-0,03	0,03	0,01	-0,01	0,66
C.5.2	0,51	0,16	0	0,13	-0,11	-0,03	0,05	0,04	-0,04	-0,02	0,01	0,41
C.5.3	-0,1	0,86	-0,16	-0,03	0,05	0,01	-0,02	0,01	0,01	0,01	0	0,56
C.5.4	0,54	-0,03	0,33	-0,12	0,03	0,07	-0,01	-0,03	0,02	0	-0,01	0,55
C.5.5	0,42	0,09	0,38	0,06	0	0,06	-0,03	-0,01	0,02	0,02	0	0,57
C.5.6	0,71	0,06	0	0,05	-0,05	-0,01	0,02	0,01	-0,01	-0,02	0	0,57
C.5.7	0,53	0,17	0,1	0,07	-0,12	0,01	0,03	0	-0,02	0	0	0,5
C.5.8	0,08	0,66	0,13	-0,08	0,03	0,04	-0,02	-0,01	0,03	0,01	0	0,62
C.5.9	0,39	0,29	0,25	0	-0,07	0,03	0	-0,01	0,01	0,01	0	0,6
C.5.10	-0,15	0,88	-0,33	0,01	0,03	-0,05	0,02	0,03	-0,03	-0,01	0	0,51
C.5.11	0,65	0,14	0,08	-0,08	-0,02	0	-0,01	-0,01	0,01	0,01	0	0,58
C.5.12	0,8	-0,01	0,06	-0,06	0,01	-0,02	0	0	0	0	0	0,65

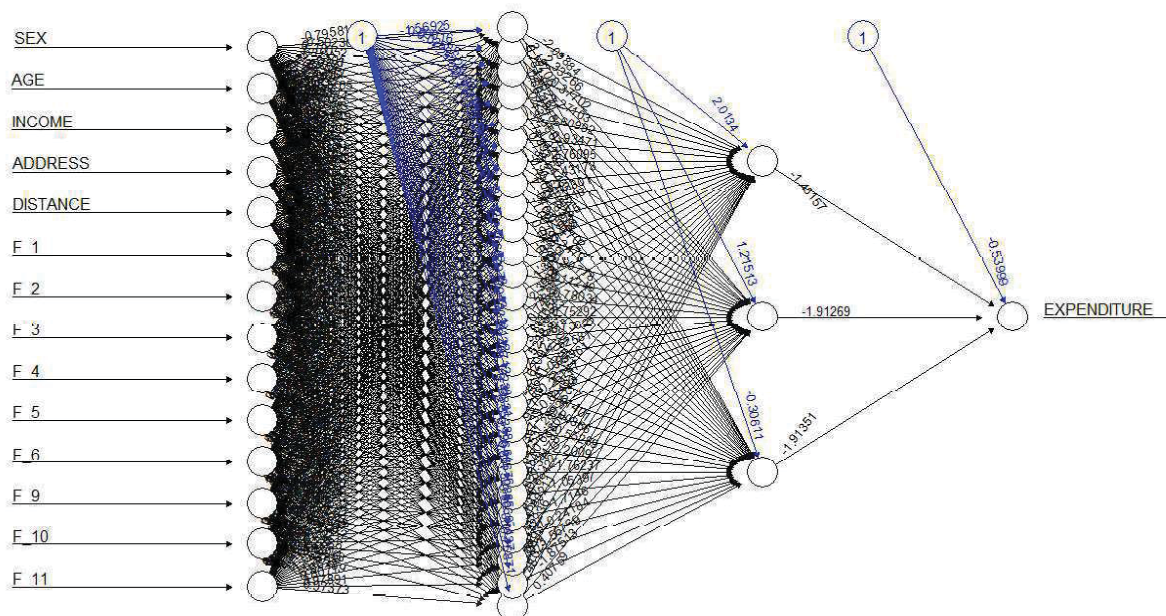


## MATRIZ COM COMPONENTES ROTACIONADOS COM MÉTODO OBLIMIM

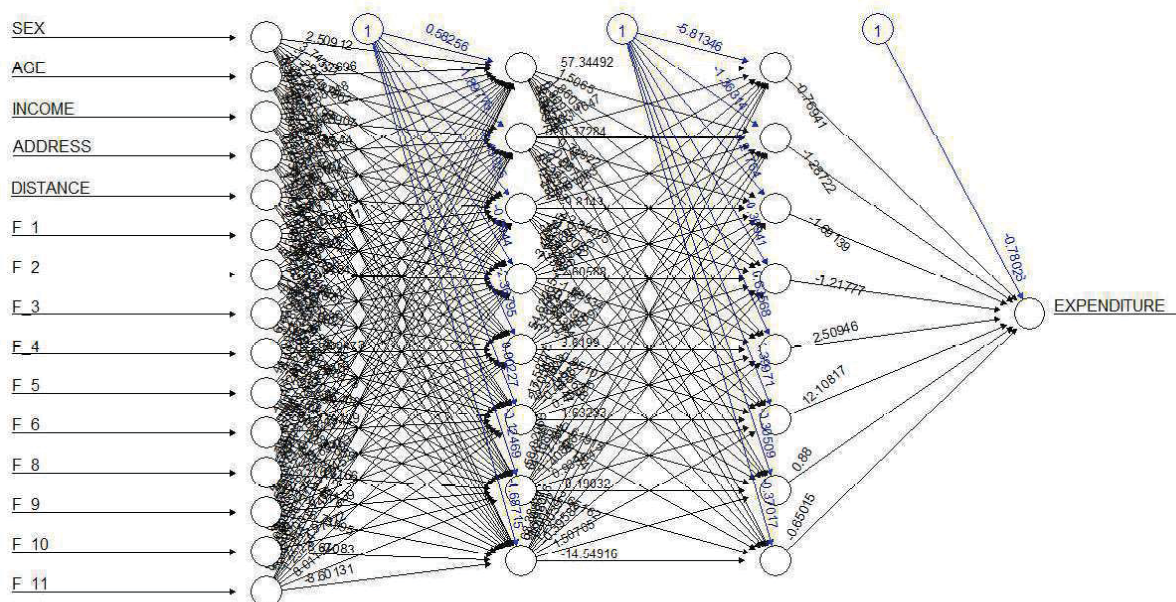
Variáveis	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	Comunidades
C.1.1	0,26	0,06	0,02	0	0,27	0,3	0	0,08	0,1	0,02	0,02	0,32
C.1.2	0,03	0,06	0,13	0,06	0	0,7	-0,08	0,09	-0,14	0,07	-0,03	0,55
C.1.3	0,16	0,26	-0,01	-0,13	0,07	0,36	0,01	-0,02	0,03	-0,05	0	0,27
C.1.4	0,73	-0,07	-0,12	0,05	0,09	0,25	0,04	0,01	0,02	-0,01	0	0,62
C.1.5	0,08	0,09	-0,06	-0,07	0,53	0,15	0,06	-0,21	0,23	-0,07	0,02	0,45
C.1.6	-0,02	-0,03	0,2	0,05	0,7	-0,05	0	0,05	-0,1	0,06	0,01	0,55
C.1.7	0,01	0	0	0,01	0	-0,01	-0,01	0,01	-0,02	0	1	0,99
C.2.1	-0,03	0	0	-0,07	-0,02	0,05	0,05	0,91	0,09	-0,03	0,02	0,84
C.2.2	0,03	-0,01	0,01	0,11	0,03	0,02	0,57	0,1	0,13	0,12	0	0,4
C.2.3	-0,01	0	-0,02	-0,02	0,01	0,02	0,01	-0,03	0,02	0,98	0	0,96
C.2.4	0,02	0,02	0,01	0,01	0	-0,06	0,83	0,03	-0,11	-0,01	-0,02	0,69
C.2.5	0,01	0	0,01	0,07	-0,01	-0,07	-0,06	0,09	0,9	0,04	-0,02	0,82
C.3.1	0,06	0,08	-0,16	0,41	-0,04	0,42	0,02	0,03	-0,03	-0,02	-0,01	0,41
C.3.2	0,09	0,03	-0,34	0,26	0,32	-0,24	-0,16	0,28	-0,24	0,05	-0,07	0,57
C.3.3	-0,05	0,02	0,03	0,71	-0,02	0,11	0,04	-0,06	0,1	-0,03	0,03	0,51
C.3.4	0	-0,04	0,02	0,7	0,01	-0,08	0,03	-0,08	0,04	-0,02	0,01	0,49
C.4.1	0,88	-0,06	0,02	-0,01	-0,01	-0,02	0	0	-0,01	0	0	0,73
C.4.2	0,86	-0,03	-0,01	0,02	-0,03	0	0	0	0	0,01	0	0,72
C.4.3	0,6	0,15	0,15	-0,02	0,17	-0,09	0	-0,01	0,02	0	0,01	0,6
C.4.4	0,76	0,15	-0,04	0	0,1	-0,1	0,01	-0,02	0,03	-0,01	0,01	0,71
C.4.5	0,88	-0,05	0,05	0	-0,09	0,01	-0,01	0,01	-0,02	0,02	0	0,75
C.4.6	0,92	-0,08	-0,09	-0,02	-0,02	0,09	0,01	0	0	0	0,01	0,77
C.4.7	0,84	-0,05	-0,17	0,03	0,07	0,01	0,01	-0,01	0,02	-0,01	0,01	0,66
C.4.8	-0,01	-0,1	0,65	0,02	0,31	0,12	0,02	0,02	0,03	-0,05	-0,01	0,51
C.5.1	0,54	0,15	0,34	0,08	-0,06	0,03	-0,04	0,04	-0,03	0,01	-0,01	0,66
C.5.2	0,5	0,19	0,02	-0,01	0,14	-0,08	0,03	-0,04	0,06	-0,02	0,01	0,41
C.5.3	-0,04	0,76	-0,04	0,04	0	0,04	0,01	0,02	-0,01	0,01	0	0,56
C.5.4	0,49	0,08	0,37	0,08	-0,11	-0,02	-0,05	0,03	-0,03	0,01	-0,01	0,55
C.5.5	0,4	0,18	0,38	0,08	0,04	-0,06	-0,03	0,04	-0,04	0,03	0	0,57
C.5.6	0,67	0,11	0,04	0,01	0,07	-0,04	0,01	-0,01	0,02	-0,01	0	0,57
C.5.7	0,51	0,22	0,14	0,02	0,07	-0,12	-0,01	-0,01	0,02	0	0	0,5
C.5.8	0,11	0,64	0,23	0,06	-0,07	-0,01	-0,03	0,04	-0,03	0,02	0	0,62
C.5.9	0,38	0,34	0,31	0,04	0	-0,1	-0,02	0,02	-0,01	0,01	0	0,6
C.5.10	-0,07	0,74	-0,22	-0,04	0,05	0,06	0,04	-0,03	0,04	-0,01	0	0,51
C.5.11	0,6	0,2	0,16	0,01	-0,06	-0,05	-0,02	0,01	-0,01	0,02	0	0,58
C.5.12	0,74	0,07	0,12	0	-0,03	0	-0,01	0,01	0	0	0	0,65

### ANEXO 3 – REPRESENTAÇÃO DAS REDES NEURAIS PARA AS CLASSES 2 A 5

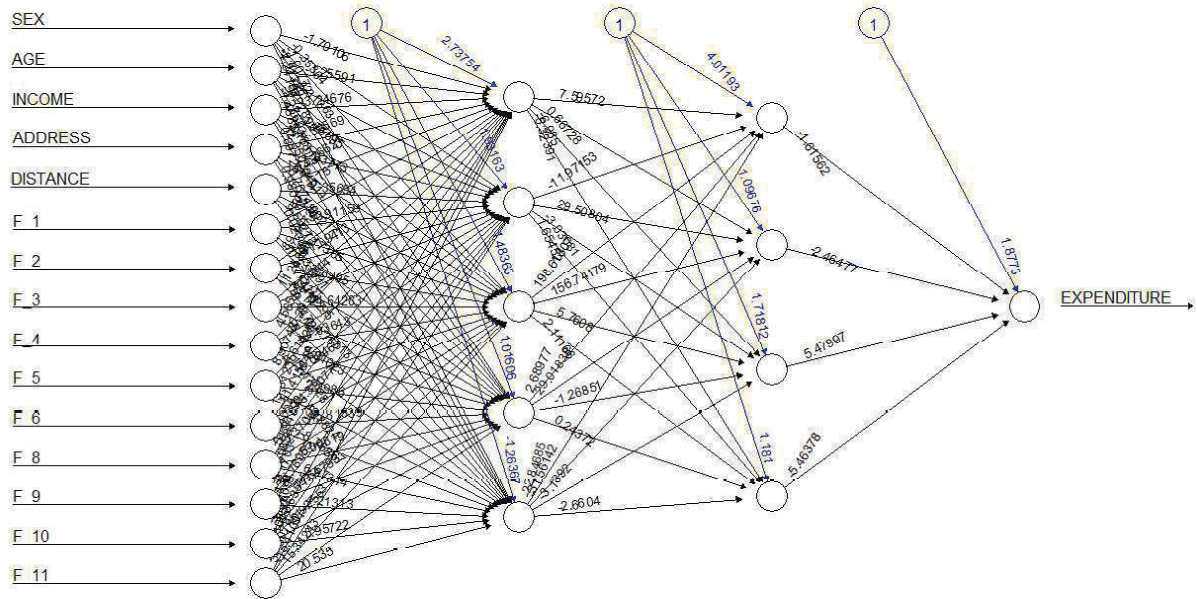
REPRESENTAÇÃO DA REDE 14E27N3N1 PARA CLASSE 2



REPRESENTAÇÃO DA REDE 15E8N8N1 PARA CLASSE 3



REPRESENTAÇÃO DA REDE 15E5N4N1 PARA CLASSE 4



REPRESENTAÇÃO DA REDE 15E3N24N1 PARA CLASSE 5

