

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

ANA CAROLINA NEVES CARNELOSSI

**USO DE APRENDIZAGEM DE MÁQUINA PARA CLASSIFICAÇÃO DE
ITENS DO PORTFÓLIO DE UMA REDE DE SUPERMERCADOS**

Maringá/PR
2022

ANA CAROLINA NEVES CARNELOSSI

**USO DE APRENDIZAGEM DE MÁQUINA PARA CLASSIFICAÇÃO DE
ITENS DO PORTFÓLIO DE UMA REDE DE SUPERMERCADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia de Produção, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção.
Área de concentração: Engenharia de Produção

Orientador(a): Prof. Dr. Edwin V. C. Galdamez
Coorientador(a): Prof. Dr. Rafael H. Palma Lima

Maringá/PR
2022

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

C289u

Carnelossi, Ana Carolina Neves

Uso de aprendizagem de máquina para classificação de itens do portfólio de uma rede de supermercados / Ana Carolina Neves Carnelossi. -- Maringá, PR, 2022.
93 f.: il. color., figs., tabs.

Orientador: Prof. Dr. Edwin Vladimir Cardoza Galdamez.

Coorientador: Prof. Dr. Rafael Henrique Palma Lima.

Dissertação (Mestrado) - Universidade Estadual de Maringá, Centro de Tecnologia, Departamento de Engenharia de Produção, Programa de Pós-Graduação em Engenharia de Produção, 2022.

1. Gestão empresarial. 2. Gestão de suprimentos. 3. Gestão de portfólio de suprimentos. 4. Aprendizagem de máquina. 5. Mineração de dados. I. Galdamez, Edwin Vladimir Cardoza, orient. II. Lima, Rafael Henrique Palma, coorient. III. Universidade Estadual de Maringá. Centro de Tecnologia. Departamento de Engenharia de Produção. Programa de Pós-Graduação em Engenharia de Produção. IV. Título.

CDD 23.ed. 658.5

FOLHA DE APROVAÇÃO

ANA CAROLINA NEVES CARNELOSSI

USO DE APRENDIZAGEM DE MÁQUINA PARA CLASSIFICAÇÃO DE ITENS DO PORTFÓLIO DE UMA REDE DE SUPERMERCADOS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia de Produção, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA

Documento assinado digitalmente



EDWIN VLADIMIR CARDOZA GALDAMEZ

Data: 20/10/2022 09:15:49-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. Edwin V. Cardoza Galdamez

Universidade Estadual de Maringá – DEP/UEM

Documento assinado digitalmente



RAFAEL HENRIQUE PALMA LIMA

Data: 26/10/2022 14:21:41-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. Rafael Henrique Palma Lima

Universidade Estadual de Maringá – DEP/UEM

Coorientador

Prof. Dr. Marco Antonio Ferreira

Universidade Estadual de Maringá – DEP/UEM

Documento assinado digitalmente



BRUNO SAMWAYS DOS SANTOS

Data: 26/10/2022 18:51:57-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. Bruno Samways dos Santos

Universidade Tecnológica Federal do Paraná *campus* Londrina – DEP/UTFPR

Membro 2 da Banca

Aprovada em 08 de Setembro de 2022.

Local da defesa: Sala de Defesa - Bloco C67, campus da Universidade Estadual de Maringá.

AGRADECIMENTOS

Agradeço à minha avó e minha irmã por terem me dado todo o apoio, de que eu nem sabia que precisava, ao longo do período do mestrado. Sem esse apoio, eu não teria conseguido concluir essa etapa de estudos.

Ao Prof. Edwin por ter me aceito como orientanda e ao Prof. Rafael pela orientação construtiva e suporte ao longo de todo o processo. Todos os comentários e feedbacks foram fundamentais para a construção dessa dissertação e me conduziram a aprendizados importantes.

Aos meus colegas do grupo do mestrado por terem compartilhado comigo muito de suas qualidades e generosidade. Ter partilhado dessa experiência fez diferença ao longo do curso.

Aos professores do mestrado, pela dedicação e oportunidades de aprendizado proporcionadas e à secretária do programa, Elisandra, pela presteza de sempre.

Uso de Aprendizagem de Máquina para Classificação de Itens do Portfólio de uma Rede de Supermercados

RESUMO

A gestão de suprimentos constitui uma área de pesquisa do contexto organizacional que tem como principal atribuição disponibilizar recursos externos à empresa, possuindo como variáveis de decisão o que comprar, de quem comprar e sob quais condições mercadológicas e de relacionamento comprar. Nesse contexto, o objetivo da presente pesquisa consistiu em implementar e avaliar modelos de aprendizagem de máquina de classificação para a reprodução das classes de itens do portfólio de uma rede varejista de supermercados quanto à função estratégica que desempenham, de modo a assegurar que a decisão da empresa sobre quais itens comprar atenda à exigência de manter seu portfólio constante em termos da quantidade de SKUs pertencente a cada classe. Para isso, lançou-se mão do processo de extração de conhecimento de base de dados (KDD) delineado por Fayyad, Piatetsky-Shapiro e Smyth (1996), executando-se o passo de mineração de dados via tarefa de classificação e obtendo-se como resultados um destaque para os modelos RF, KNN e DT, em detrimento aos SVM e ANN, assim como da técnica de seleção de variáveis de entrada *Sequential Feature Selector*, SFS, em detrimento às baseadas em *feature importance*, *permutation importance*, *f classification* e *mutual info classification*, bem como do método SHAP para interpretação local e global de modelo. Nesse sentido, a conclusão da pesquisa foi a aplicabilidade de modelos menos complexos para a reprodução de classes no contexto da gestão de portfólio de suprimentos, aliada à técnica de seleção SFS e ao método SHAP para interpretação dos resultados. Por fim, as contribuições da pesquisa foram, sob a perspectiva acadêmica, a colaboração para o preenchimento da lacuna científica sobre gestão de suprimentos apoiada por aprendizagem de máquina e, sob a perspectiva prática, o desenvolvimento de um mecanismo embrionário de classificação automatizada de SKUs capaz de dar replicabilidade, escalabilidade e dinamicidade à classificação, além da construção de uma base de dados com disponibilidade de informações antes desconhecidas pela organização e que podem servir como subsídio para que entenda a representatividade de seus fornecedores e a dependência que possui com relação aos mesmos e tenha condições de melhor delinear as práticas que permeiam seus relacionamentos.

Palavras-chave: *gestão de suprimentos, portfólio, classificação, aprendizagem de máquina, shap.*

Use of Machine Learning to Classification of Supply Portfolio Items in a Supermarket Chain

ABSTRACT

Supply management is an area of research applied to the organizational context whose main attribution is to make resources external to the company available, having as decision variables what to buy, from whom to buy and under what market and relationship conditions to buy. In this context, the objective of the present research was to implement and evaluate classification machine learning models for the reproduction of the classes of items in the portfolio of a supermarket retail company regarding the strategic function they perform, in order to ensure that the decision of the company on which items to buy meets its requirement to keep its portfolio constant in terms of the number of SKUs belonging to each class. For this, the process of knowledge discovery in databases (KDD) outlined by Fayyad, Piatetsky-Shapiro and Smyth (1996) was used, executing the data mining step through the classification task and obtaining as results a highlight for the RF, KNN and DT models, in detriment to the SVM and ANN, for the technique of selection of input variables Sequential Feature Selector, SFS, in detriment to those based on feature importance, permutation importance, f classification and mutual info classification, as well the SHAP method for local and global model interpretation. In this sense, the conclusion of the research was the applicability of less complex models for the reproduction of classes in the context of supply portfolio management, combined with the technique for feature selection SFS and the SHAP method for interpreting the results. Finally, the research contributions were, from the academic perspective, the collaboration to fill the scientific gap on supply management supported by machine learning and, from the practical perspective, the development of an embryonic mechanism of automated classification of SKUs capable of giving replicability, scalability and dynamics to the classification, in addition to building a database with availability of information previously unknown to the company and which can serve as a subsidy to understand the representativeness of its suppliers and the dependence that has in relation to them and is able to better outline the relationship management.

Keywords: *supply management, portfolio, classification, machine learning, shap.*

LISTA DE QUADROS

Quadro 2.1 - Trabalhos que abordaram Classificação no contexto da Gestão de Suprimentos	26
Quadro 2.2 - Trabalhos que abordaram Classificação via AM no contexto da G. Estoques	31

LISTA DE TABELAS

Tabela 2.1 – Métricas de Avaliação de Modelos de Classificação	22
Tabela 3.1 – Características das Bases de Dados Avulsas	37
Tabela 3.2 – Bibliotecas utilizadas	41
Tabela 3.3 – Características das Bases de Dados integradas	43
Tabela 3.4 – Delimitações dos Registros da Base de Dados Integrada	44
Tabela 3.5 – Características da Bases de Dados Integrada com Registros delimitados	44
Tabela 3.6 – Delimitações e Construção de Atributos da Base de Dados Integrada	44
Tabela 3.7 – Características da Base Dados Integrada com Registros/Atributos delimitados	47
Tabela 3.8 – Variáveis de Entrada da Base de Dados objeto desta pesquisa	47
Tabela 3.9 – Correlação entre as Variáveis de Entrada	48
Tabela 4.1 – Modelos e Parâmetros ajustados	52
Tabela 4.2 – Resultados do <i>GridSearchCV</i> para o modelo KNN	53
Tabela 4.3 – Resultados do <i>GridSearchCV</i> para o modelo DT	55
Tabela 4.4 – Resultados do <i>GridSearchCV</i> para o modelo RF	57
Tabela 4.5 – Comparativo das características Pré e Pós SFS dos modelos DT e RF	59
Tabela 4.6 – Resultados do <i>GridSearchCV</i> para o modelo SVM	60
Tabela 4.7 – Resultados do <i>GridSearchCV</i> para o modelo ANN	62
Tabela 4.8 – Técnicas para Seleção de Variáveis de Entrada aplicadas aos Modelos	64
Tabela 4.9 - Relatórios de Avaliação dos modelos Pré redução das variáveis de entrada	65
Tabela 4.10 – <i>Feature Importance</i> dos modelos DT e RF	66
Tabela 4.11 – <i>Permutation Importance</i> dos modelos	67
Tabela 4.12 – Comparativo de Desempenho das Técnicas de Seleção a 9 variáveis	69
Tabela 4.13 – Variáveis selecionadas pelas diferentes técnicas para o modelo RF	71
Tabela 4.14 – Variáveis selecionadas para o modelo RF pela técnica de seleção SFS	72
Tabela 4.15 – Relatórios de Avaliação dos modelos Pós redução das variáveis	72
Tabela 4.16 – Relatório da Avaliação Relativa Pós versus Pré redução das variáveis	74
Tabela 4.17 – Acurácias por classe dos modelos após a redução das variáveis	76
Tabela 4.18 – Exemplos de itens e suas respectivas predições e probabilidades	81

LISTA DE FIGURAS

Figura 3.1 – Características dos itens que compõem a Base de Dados objeto desta pesquisa	38
Figura 3.2 - Atividades da Pesquisa	39
Figura 4.1 - Avaliação dos Parâmetros do Modelo KNN	54
Figura 4.2 – Avaliação dos Parâmetros do Modelo DT	56
Figura 4.3 – Avaliação dos Parâmetros do Modelo RF	58
Figura 4.4 - Avaliação dos Parâmetros do Modelo SVM	61
Figura 4.5 – Avaliação dos Parâmetros do Modelo ANN	63
Figura 4.6 – Desempenho da técnica SFS por modelo e quantidade de variáveis selecionadas	70
Figura 4.7 – <i>Boxplots</i> do Desempenho dos modelos Pré e Pós Seleção das Variáveis	77
Figura 4.8 – Importância das Variáveis de entrada segundo a técnica SHAP	78
Figura 4.9 - Valores Shap para cada Função Estratégica individualmente	80
Figura 4.10 - Valores Shap Cumulativos - Exemplo de um item de cada Função Estratégica	82

LISTA DE ABREVIATURAS E SIGLAS

AHP	<i>Analytical Hierarchy Process</i>
ANN	<i>Artificial Neural Network</i>
ASCII	<i>American Standard Code for Information Interchange</i>
AUC	<i>Area Under the Curve</i>
BLAG	<i>Balanced Bagging</i>
BN	<i>Bayesian Network</i>
CART	<i>Classification and Regression Tree</i>
CSV	<i>Comma-Separated Values</i>
CV	<i>Cross-Validation</i>
DNN	<i>Deep Neural Network</i>
DT	<i>Decision Tree</i>
Gboost	<i>Gradient Boosting</i>
KDD	<i>Knowledge Discovery in Databases</i>
KNN	<i>k- Nearest Neighbor</i>
LGBM	<i>Light Gradient Boosted Machine</i>
LIME	<i>Local Interpretable Model-agnostic Explanations</i>
LRG	<i>Logistic Regression</i>
MLP	<i>Multi Layer Perceptron</i>
NB	<i>Naive Bayes</i>
N-BN	<i>Naive Bayesian Network</i>
NN	<i>Neural Network</i>
PDP	<i>Partial Dependence Plot</i>
RF	<i>Random Forest</i>
RFID	<i>Radio-Frequency Identification</i>
ROC	<i>Receiver Operating Characteristic</i>
SAW	<i>Simple-Additive Weighting</i>
SFS	<i>Sequential Feature Selector</i>
SHAP	<i>Shapley Additive exPlanations</i>
SKU	<i>Stock Keeping Units</i>
SVM	<i>Support Vector Machine</i>
UTF-8	<i>Unicode Transformation Format</i>
XGBoost	<i>Extreme Gradient Boosting</i>

SUMÁRIO

1. INTRODUÇÃO	13
1.1. Contextualização	13
1.2. Objetivos	14
1.3. Justificativa	15
1.4. Estrutura do Trabalho	16
2. REFERENCIAL TEÓRICO	17
2.1. Aprendizagem de Máquina	17
2.1.1. Tipos de Aprendizagem de Máquina e Tarefas	17
2.1.2. Métodos <i>Ensemble</i> de Aprendizagem de Máquina	18
2.1.3. Modelos de Classificação KNN, DT, RF, SVM e ANN	19
2.1.4. Seleção de Variáveis para Modelos de Classificação	21
2.1.5. Avaliação de Modelos de Classificação	22
2.1.6. Interpretação de Modelos de Classificação	24
2.2. Classificação na Gestão de Suprimentos	25
2.3. Classificação na Gestão de Estoques via Aprendizagem de Máquina	30
3. MÉTODO DE PESQUISA	36
3.1. Classificação e Contexto da Pesquisa	36
3.2. Atividades e Recursos da Pesquisa	39
3.3. Etapa 1 – Construção e Preparação da Base de Dados	42
3.3.1. Base de Dados objeto da Pesquisa	47
3.4. Etapa 2 – Implementação da Aprendizagem de Máquina	50
4. RESULTADOS	52
4.1. Parametrização dos Modelos de Aprendizagem de Máquina	52
4.2. Seleção e Importância de Variáveis de Entrada	64
4.3. Avaliação dos Modelos Implementados	72
4.4. Interpretação do melhor Modelo Implementado	78
4.5. Implicações Práticas para a Rede de Supermercados	83
5. CONSIDERAÇÕES FINAIS	85
5.1. Constatações e Conclusão da Pesquisa	85
5.2. Contribuições da Pesquisa	86
5.3. Limitações da Pesquisa	87
5.4. Oportunidades de Pesquisas Futuras	88
REFERÊNCIAS	89

1. INTRODUÇÃO

Este capítulo apresenta a contextualização da pesquisa, seus objetivos e justificativa, além da estrutura do trabalho.

1.1. Contextualização

A gestão de suprimentos constitui uma área do contexto organizacional que tem como principal atribuição disponibilizar recursos externos à empresa, estando relacionada às atividades do processo de aquisição que abrangem desde pesquisas de mercado fornecedor, o pedido e compra em si, até a gestão do relacionamento com os fornecedores. Está inserida no âmbito da gestão da cadeia de suprimentos junto das áreas de gestão de operações, logística e marketing, desempenhando o papel central de constituir um elo entre empresa e fornecedores bem como entre demanda e suprimentos (ELLRAM *et al.*, 2020).

Nesse sentido, perpassam essa área variáveis de decisão como o que comprar, de quem comprar e sob quais condições mercadológicas e de relacionamento comprar, que podem ser apoiadas pela prática de classificação aplicada a itens do portfólio de suprimentos da empresa, fornecedores ou relacionamentos, a exemplo dos trabalhos desenvolvidos por Kraljic (1983), Olsen e Ellram (1997) e Bensaou (1999). Ainda, tais classificações, quando aplicadas aos itens de portfólio, podem ter desdobramentos que extrapolam a gestão de suprimentos em si, alcançando, por exemplo, o escopo da gestão de operações no contexto da gestão de estoques.

A classificação, por sua vez, é uma prática do contexto empresarial que pode ser realizada com base em múltiplos critérios por meio de métodos advindos de disciplinas como a de apoio à decisão multicritério, estatística/econometria via técnicas como análise discriminante e regressão logística, de programação matemática via técnicas como programação linear ou não linear, e de inteligência artificial, via técnicas como lógica *fuzzy*, além das provenientes da aprendizagem de máquina, como modelos de regras ou árvores de decisão e de redes neurais artificiais (DOUMPOS; ZOPOUNIDIS, 2004).

A aprendizagem de máquina, nesse sentido, pode prover a classificação baseada em classes já conhecidas, por meio de sua tarefa de classificação, ou baseada apenas nas características dos itens, sem conhecimento prévio de classes predeterminadas, por meio de sua tarefa de clusterização (GERÓN, 2019). Tanto de uma forma como de outra, quando implementada sobre bases de dados pode constituir o meio implementação da mineração de dados, que é um passo do processo de descoberta de conhecimento de base de dados, KDD, delineado por Fayyad, Piatetsky-Shapiro e Smyth (1996).

Nesse contexto, a presente pesquisa tem como objeto de estudo a gestão de suprimentos apoiada por aprendizagem de máquina por meio da implementação da tarefa de classificação sobre uma base de dados do portfólio de unidades de manutenção de estoque (*Stock Keeping Unit*, SKUs) de uma rede varejista de supermercados a fim de reproduzir as classes desses itens quanto à função estratégica que desempenham para a organização. Essas classes são quatro e compreendem a líder, vice-líder, premium e primeiro-preço, correspondentes a, respectivamente, itens que se diferenciam por serem de uma marca de renome e líder de mercado, de uma marca vice-líder de mercado, itens nobres e caros e itens populares e baratos.

1.2. Objetivos

O objetivo deste trabalho é implementar e avaliar modelos de aprendizagem de máquina de classificação para reproduzir as classes de itens do portfólio de uma rede varejista de supermercados quanto às funções estratégicas que esses itens desempenham para a organização.

Os objetivos específicos para alcançar o objetivo geral são:

- a) Preparar a base de dados de uma rede varejista de supermercados de modo a torná-la passível de ser processada via modelos de aprendizagem de máquina e de vir a constituir o objeto desta pesquisa;
- b) Implementar e avaliar modelos de aprendizagem de máquina para a classificação dos itens da base de dados objeto desta pesquisa quanto às funções estratégicas que desempenham para a organização;
- c) Aplicar e avaliar técnicas de seleção que reduzam os atributos da base de dados àqueles suficientes para constituírem as variáveis de entrada dos modelos e reimplementar/reavaliar esses modelos com a respectiva base de dados reduzida;
- d) Aplicar um método de interpretação sobre o modelo que teve o melhor desempenho quando implementado sobre a respectiva base de dados reduzida segundo a melhor técnica de seleção de variáveis de entrada.

1.3. Justificativa

A presente pesquisa apresenta relevância por ter como objeto de estudo um tema, a gestão de suprimentos apoiada pela aprendizagem de máquina, com escassa literatura científica mas com condições metodológicas de ser amplamente explorado, além de dizer respeito a uma área organizacional com importante papel para o desempenho das empresas e por ser abordado no âmbito de um importante setor de atividade econômica do país.

A escassa literatura sobre gestão de suprimentos apoiada pela aprendizagem de máquina é corroborada por Lajimi e Majidi (2021), que ao realizarem uma revisão sistemática sobre classificação no contexto da gestão de suprimentos não encontraram o emprego de técnicas de classificação via aprendizagem de máquina. Por outro lado, a existência de condições metodológicas para que esse tema seja explorado é embasada pela vasta literatura científica sobre classificação de itens via aprendizagem de máquina, mas aplicada em outro contexto, como o da gestão de estoques.

Nesse contexto da gestão de estoques, Lolli *et al.* (2019) aplicaram a tarefa de classificação para apoiar a política de reposição da empresa e a avaliaram comparativamente a um método de simulação. Mishra e Mohapatro (2020), por sua vez, aplicaram a tarefa para apoiar a classificação multicritério dos itens com base em suas flutuações de demanda e a avaliaram comparativamente a um método de tomada de decisão multicritério. Svoboda e Minner (2021), por sua vez, aplicaram a tarefa também para subsidiar a política de reposição da empresa, mas a avaliaram comparativamente a um método de otimização.

Por sua vez, a área de gestão de suprimentos apoiada por um mecanismo de classificação de itens do portfólio via aprendizagem de máquina pode, além de ter explícito o conhecimento subjetivo envolto a essa prática, ganhar dinamicidade ao realizá-la. Ainda, aliada à extração de outros conhecimentos implícitos na base de dados, como por exemplo o grau de dependência da empresa com o fornecedor, essa classificação pode ser estendida aos relacionamentos e, conseqüentemente, constituir um passo inicial para ser desdobrada em abordagens de suprimentos, políticas de compra e estratégias de relacionamento.

Por fim, o setor do comércio varejista de hipermercados e supermercados apresenta papel relevante em termos de ocupação de pessoal no país. No primeiro trimestre de 2022 quase 20% das pessoas ocupadas pertenciam ao grupamento de atividade econômica do qual o comércio faz parte, tendo este ocupado a primeira posição nesse quesito nos últimos 10 anos, de 2012 a 2021, exceto no segundo semestre de 2020, quando ocupou a segunda posição (IBGE,

2022). Em 2019, o comércio varejista concentrou aproximadamente 3/4 desse total, com participação de quase 1/5 sendo de hipermercados e supermercados (IBGE, 2019).

1.4. Estrutura do Trabalho

O trabalho está estruturado em capítulos de referencial teórico (capítulo 2), de método da pesquisa (capítulo 3), de resultados (capítulo 4) e de considerações finais (capítulo 5).

O capítulo 2 de referencial teórico apresenta os fundamentos da aprendizagem de máquina que alicerçaram esta pesquisa, incluindo tarefas, métodos e modelos de classificação implementados, bem como técnicas de seleção de variáveis de entrada, métricas e mecanismos de avaliação, além de métodos de interpretação de modelos (2.1). Também apresenta trabalhos científicos que abordaram a classificação no contexto da gestão de suprimentos (2.2) e que aplicaram via aprendizagem de máquina no contexto da gestão de estoques (2.3).

O capítulo 3 de método da pesquisa apresenta sua classificação e contexto, caracteriza as bases de dados que constituíram seu ponto de partida (3.1) e dá uma visão geral das atividades e recursos utilizados (3.2). Também descreve a etapa de construção/preparação da base de dados que viera a se tornar a base objeto desta pesquisa e apresenta suas características (3.3) bem como descreve a etapa de implementação da aprendizagem de máquina propriamente dita (3.4).

O capítulo 4 de resultados apresenta os resultados correspondentes à etapa de implementação da aprendizagem de máquina por meio dos modelos de classificação KNN, DT, RF, SVM e ANN. Abrange os resultados do processo de parametrização dos modelos (4.1), da aplicação das técnicas de seleção de variáveis de entrada (4.2), avaliação de desempenho (4.3) e interpretação do melhor modelo sobre a base de dados reduzida segundo melhor técnica (4.4).

O capítulo 5 de considerações finais apresenta as constatações provenientes da pesquisa com relação aos modelos de aprendizagem de máquina implementados, das técnicas de seleção e método de interpretação aplicados, além da conclusão central decorrente de sua realização (5.1). Também apresenta as contribuições científicas e práticas que gerou (5.2), limitações que apresentou (5.3) e as oportunidades de pesquisas futuras relacionadas (5.4).

2. REFERENCIAL TEÓRICO

Este capítulo apresenta um referencial teórico sobre aprendizagem de máquina (2.1) bem como apresenta trabalhos científicos que abordaram a classificação no contexto da gestão de suprimentos (2.2) e no contexto da gestão de estoques via aprendizagem de máquina (2.3).

2.1. Aprendizagem de Máquina

A aprendizagem de máquina pode ser compreendida a partir dos diferentes tipos de aprendizagem, das diferentes finalidades ou tarefas que podem ser desempenhadas e dos diferentes métodos e modelos por meios dos quais essas tarefas podem ser implementadas (GERÓN, 2019). Assim, este subcapítulo discorre sobre cada uma das variáveis que caracterizam a aprendizagem de máquina (2.1.1 a 2.1.3), bem como as atividades que compõem seu fluxo geral de execução (2.1.4 a 2.1.6).

2.1.1. Tipos de Aprendizagem de Máquina e Tarefas

Os tipos de aprendizagem compreendem a supervisionada, semi-supervisionada, não supervisionada e por reforço. A aprendizagem supervisionada é aquela por meio da qual o modelo atribui aos registros valores ou classes predeterminadas, que constituem entradas atreladas aos registros já conhecidos. A aprendizagem não supervisionada, por outro lado, é aquela por meio da qual o modelo cria classes ou descobre padrões presentes nos registros, mas sem que essas classes ou padrões sejam previamente conhecidos (GERÓN, 2019).

Por outro lado, a aprendizagem semi-supervisionada é aquela por meio da qual um modelo atribui valores ou classes predeterminadas que constituem entradas atreladas a apenas uma parte dos registros disponíveis, a registros desconhecidos. Todos os registros são divididos em grupos por um primeiro modelo, aos quais um segundo designa os valores ou classes predeterminadas e as estende aos registros desconhecidos que o compõem. Por fim, a

aprendizagem por reforço, diferentemente das anteriores, é aquela por meio da qual o modelo gera respostas que são aperfeiçoadas via sistema de recompensa/penalização em função dos resultados que gera (GERÓN, 2019).

As principais tarefas de aprendizagem de máquina são regressão, classificação, clusterização e de regras de associação. A regressão e classificação são tarefas que têm como finalidade prever valores ou atribuir classes predeterminadas aos registros desconhecidos, portanto, por meio de aprendizagem supervisionada. Já, a clusterização e de regras de associação têm como finalidade criar classes sob as quais os registros possam ser agrupados ou reconhecer padrões presentes nos mesmos, sem que tais classes ou padrões sejam previamente conhecidos, portanto, por meio de aprendizagem não supervisionada (GERÓN, 2019).

2.1.2. Métodos *Ensemble* de Aprendizagem de Máquina

As tarefas podem ser executadas por meio da implementação individual de determinado modelo ou por meio da combinação de mais de um deles, via métodos denominados *ensemble*. Esses métodos *ensemble* combinam resultados de diferentes modelos de aprendizagem de máquina e geram um resultado final superior aos dos modelos individuais. Essa combinação pode ser paralela ou recursiva, com modelos iguais ou diferentes, e ainda a partir de treinamento realizado sobre os mesmos ou diferentes registros (AGGARWAL, 2015).

O método *boosting*, baseado na hipótese de aumento da acurácia de um algoritmo de “aprendizagem fraca” levantada por Kearns e Valiant (1988, 1989) e respondida por Schapire (1990), consiste na implementação recursiva de modelos de aprendizagem de máquina de modo que uma rodada posterior foca nos registros que a anterior teve dificuldade de realizar a predição, culminando num resultado final de predição com melhor acurácia do que os resultados anteriores individuais.

O método *stacking*, delineado por Wolpert (1992), consiste na combinação de modelos de aprendizagem de máquina individuais realizada por um outro modelo, intermediário, que deduz e corrige os resultados dos modelos individuais em termos de seus vieses e gera um resultado final superior aos que seriam produzidos individualmente por esses modelos que foram combinados.

O método *bagging*, desenvolvido por Breiman (1994), consiste na agregação de modelos de aprendizagem de máquina que gera como resultado final, no caso de modelos de classificação, o resultado correspondente à predição majoritária dos modelos individuais - votação forçada - ou então o correspondente àquele com probabilidade de predição média dos

modelos individuais de maior valor - votação suave (GERÓN, 2019). No caso de modelos de regressão, o resultado corresponde à previsão média dos modelos individuais. Esses resultados individuais são alcançados via *bootstrap*, da onde decorre a expressão "*bootstrap aggregating*" que origina o acrônimo do método.

2.1.3. Modelos de Classificação KNN, DT, RF, SVM e ANN

O modelo de Vizinhos mais Próximos ou k-Vizinhos mais Próximos (*k-Nearest Neighbor*, KNN) consiste num procedimento de classificação não paramétrico que designa a cada registro a ter a classe predita, um vetor-distância com relação aos registros conhecidos e atribui a este a classe correspondente à do registro ou, de forma ponderada ou não pela distância, à classe predominante dos k-registros mais próximos. Esses modelos foram desenvolvidos por Fix e Hodges (1951, 1952) e por Cover e Hart (1967, 1968) (GOLDSTEIN, 1972).

Esse modelo KNN é do tipo baseado em instância, apenas memorizando os registros durante o aprendizado ao invés de construir um modelo classificador a partir dos mesmos, e em seguida designando, com relação ao registro cuja classe se quer atribuir, a classe predominante dos k-registros conhecido(s) mais próximo(s), segundo determinada métrica de distância. Essa métrica de distância pode ser a euclidiana ou outra, como a *manhattan*, *mahalanobis*, *hamming* e *minkowski* (WALTERS-WILLIAMS; LI, 2010).

O modelo de Árvore de Decisão (*Decision Tree*, DT) é um modelo que particiona o espaço dimensional dos atributos dos registros conhecidos, recursivamente, segundo características que discriminam esses registros e suas respectivas classes formando uma estrutura de árvore, o modelo propriamente dito, que pode ser utilizada para classificar registros desconhecidos (GELFAND; RAVISHANKAR; DELP, 1991). Esse modelo foi desenvolvido pela comunidade de aprendizagem de máquina, por Hunt, Marin e Stone (1966), e também pela comunidade de estatística, por Breiman *et al.* (1984) (FÜRNKRANZ, 2010).

Na estrutura de uma árvore de decisão, ao longo de todas as suas camadas, cada atributo é denominado de "nó", sendo o primeiro deles chamado de 'raiz', cada ramificação é denominada de 'ramo' e cada classe é denominada de "folha". O particionamento/ramificação dos nós busca dividi-los em nós "filhos" que sejam os mais "puros" possíveis em termos das diferentes classes que agregam sob a forma dos registros correspondentes. Para isso utiliza um critério que pode ser definido em função de impureza, como o gini, ou de ganho de informação, como a entropia (SAFAVIAN; LANDGREBE, 1991).

O modelo de Floresta Aleatória (*Random Forest*, RF) consiste numa combinação de árvores de decisão construídas a partir de um subconjunto aleatório dos atributos disponíveis, de tamanho predeterminado, e que gera como resultado aquele correspondente ao resultado majoritário das árvores individuais. Esse modelo foi desenvolvido por Breiman (2001) e constitui um método *ensemble* do tipo *bagging*, cuja aleatoriedade também incide sobre a amostragem *bootstrap* dos registros.

O modelo de Máquinas de Vetores de Suporte (*Support Vector Machine*, SVM) é um modelo que, a partir da representação dos registros conhecidos num espaço vetorial, utiliza alguns desses registros como vetores de suporte e constrói um hiperplano, o modelo propriamente dito, que separa os registros em zonas distintas, correspondentes às respectivas classes, e que é então utilizado para classificar novos registros em função da zona do hiperplano a qual se enquadram de acordo com seus atributos. Esse modelo foi introduzido por Cortes e Vapnik (1995) (PONTIL; VERRI, 1998).

A construção do hiperplano do SVM é feita por meio de uma função *kernel*, que pode ser de diferentes tipos, como a gaussiana ou polinomial, e a busca pelo melhor hiperplano, que é aquele que maximiza a margem de separação entre as classes e minimiza os erros de classificação, constitui um problema de otimização que está em função do valor de um parâmetro do modelo, o de regularização (PONTIL; VERRI, 1998).

O modelo de Rede Neural (*Neural Network*, NN ou *Artificial Neural Network*, ANN) sob a arquitetura do Perceptron Multicamadas (*Multi Layer Perceptron*, MLP) é um modelo que constrói uma rede de processamento dos atributos dos registros conhecidos, o modelo propriamente dito, que resulta em valores indicativos de suas respectivas classes, e que é então, utilizada para classificar novos registros de acordo com seus atributos (BAUM, 1988). Esse modelo MLP foi introduzido por Ivaknenko e Lapa (1965), posteriormente à proposição do perceptron por Rosenblatt (1958), e passou a incorporar, a partir de Werbos (1981), o algoritmo de aprendizagem retropropagação (*backpropagation*) (SCHMIDHUBER, 2015).

A estrutura da rede neural do perceptron multicamadas (MLP) compreende uma camada de entrada contendo os atributos dos registros, as camadas intermediárias/ocultas que processam esses atributos e a camada de saída com o resultado do processamento indicando a classe dos registros desconhecidos. O processamento em si se dá via ponderação e compilação dos atributos de entrada e submissão desse compilado a determinada função de ativação, camada a camada da rede, até a de saída. As unidades de compilação e ativação constituem os chamados neurônios, presentes nas camadas da rede, e a ponderação dos atributos de entrada

decorre de um processo de otimização que se dá via algoritmo de retropropagação (*backpropagation*) (BAUM, 1988).

2.1.4. Seleção de Variáveis para Modelos de Classificação

A seleção de variáveis de entradas dos modelos de aprendizagem de máquina, dentre todos os atributos da base de dados, pode ser realizada *i)* a partir da aplicação sobre os atributos de critério matemático que independe do modelo utilizado, *ii)* a partir de informações decorrentes do processamento do próprio modelo implementado, e, *iii)* a partir da aplicação de um algoritmo por meio do qual o modelo de aprendizagem de máquina é implementado, iterativamente, em conjuntos de variáveis de entrada distintos de modo a ter identificado as mais relevantes (AGGARWAL, 2015).

A abordagem filtro corresponde à aplicação, a despeito do modelo de aprendizagem de máquina utilizado, de algum critério matemático sobre os atributos da base de dados, individualmente ou em conjunto, que as ranqueie em termos de seu poder de discriminação dos registros entre as diferentes classes. Esse critério pode ser, dentre outros, a entropia, com vistas à sua minimização e maximização do ganho de informação, ou o índice gini, com vistas à sua minimização (KOHAVI; JOHN, 1997).

A abordagem embutida, por sua vez, corresponde àquela que decorre do peso das variáveis de entrada definido pelo próprio modelo de aprendizagem de máquina e possibilita a remoção das variáveis que são menos utilizadas para realizar a classificação. Essa remoção pode ser realizada de uma única vez ou então recursivamente, quando, de modo iterativo, a cada determinada quantidade de variáveis removidas, o modelo é novamente treinado e o procedimento de remoção reiniciado (AGGARWAL, 2015).

A abordagem *wrapper*, por fim, corresponde à aplicação de um procedimento envolto ao modelo de aprendizagem de máquina que avalia os subconjuntos de variáveis de entrada em termos do desempenho que acarretam ao modelo. Esse procedimento pode ser exaustivo, quando avalia todos os conjuntos possíveis das variáveis, ou pode ser heurístico, via algoritmo ganancioso (*greedy*). O procedimento heurístico pode partir de um conjunto vazio de variáveis e ir adicionando, uma a uma, aquelas que acarretam maior ganho de desempenho ao modelo – *sequential forward* – ou pode partir do conjunto completo de variáveis e ir retirando aquelas que acarretam menor perda de desempenho – *sequential backward* – (KOHAVI; JOHN, 1997).

2.1.5. Avaliação de Modelos de Classificação

O desempenho dos modelos de aprendizagem de máquina pode ser avaliado por diferentes métricas de avaliação (2.1.5.1) e por diferentes mecanismos de validação (0).

2.1.5.1. Métricas de Avaliação de Modelos de Classificação

As principais métricas de avaliação pelas quais o desempenho dos modelos de aprendizagem de máquina pode ser avaliado estão apresentadas na Tabela 2.1.

Tabela 2.1 – Métricas de Avaliação de Modelos de Classificação

Métrica	Fórmula
Acurácia	$\frac{\text{Quantidade de Classificações Corretas}}{\text{Quantidade Total de Classificações}}$
Área sob a curva ROC	<i>Área sob a Curva da Taxa de Verdadeiro x Falso Positivo (Receiver Operating Characteristic, ROC)</i>
Taxa de Falso Positivo (ou Taxa de Alarme Falso)	$\frac{\text{Quantidade de Classificações Positivas Erradas}}{\text{Quantidade de itens Negativos}}$
Taxa de Verdadeiro Positivo (ou Sensibilidade ou <i>Recall</i>)	$\frac{\text{Quantidade de Classificações Positivas Corretas}}{\text{Quantidade de itens Positivos}}$
F1-score	$2 \times \frac{(\text{Sensibilidade} \times \text{Precisão})}{(\text{Sensibilidade} + \text{Precisão})}$
Valor Preditivo Positivo (ou Precisão)	$\frac{\text{Quantidade de Classificações Positivas Corretas}}{\text{Quantidade Total de Classificações Positivas}}$
Valor Preditivo Negativo	$\frac{\text{Quantidade de Classificações Negativas Corretas}}{\text{Quantidade Total de Classificações Negativas}}$
Taxa de Verdadeiro Negativo (ou Especificidade)	$\frac{\text{Quantidade de Classificações Negativas Corretas}}{\text{Quantidade de itens Negativos}}$

Obs. 1: Quantidade de itens positivos corresponde à quantidade de itens que pertencem a determinada classe.

Obs. 2: Quantidade de itens negativos corresponde à quantidade de itens que não pertencem a determinada classe.

Fonte: Adaptado de Sammut e Webb (2010) e Ting (2010)

A Acurácia de um modelo de aprendizagem de máquina, conforme apresentado na Tabela 2.1, consiste na proporção de registros que foram corretamente classificados pelo modelo, dentre todos os registros que o modelo classificou (SAMMUT; WEBB, 2010).

A Área sob a Curva ROC (AUC ROC) consiste na área do gráfico abaixo da curva das taxas de verdadeiro positivo *versus* falso positivo geradas pelo modelo para diferentes limiares de decisão. Por sua vez, a Taxa de Falso Positivo, também chamada de Taxa de Alarme Falso, consiste na proporção de registros negativos que foram erroneamente classificados como positivos, e a Taxa de Verdadeiro Positivo, também denominada de Sensibilidade ou *Recall*, na proporção de registros positivos que foram corretamente classificados (SAMMUT; WEBB,

2010). Já, o F1-score corresponde à média harmônica entre a Taxa de Verdadeiro Positivo e o Valor Preditivo Positivo.

O Valor Preditivo Positivo, também denominado de Precisão, consiste na proporção de registros que foram corretamente classificados como positivos dentre todos os que receberam essa classificação (TING, 2010). De modo análogo, o Valor Preditivo Negativo consiste na proporção de registros que foram corretamente classificados como negativos dentre todos os classificados dessa forma. Por outro lado, a Taxa de Verdadeiro Negativo, também denominada de Especificidade, consiste na proporção de registros corretamente classificados como negativos dentre todos os verdadeiros negativos (SAMMUT; WEBB, 2010).

As métricas de desempenho referentes a um modelo de classificação que abrange mais de duas classes, multiclasse, podem dizer respeito tanto ao desempenho geral do modelo quanto às respectivas classes individuais ou suas médias. A métrica que diz respeito ao desempenho geral do modelo é denominada de média micro e não discrimina o desempenho em termos de alguma classe específica. Já, a métrica que diz respeito ao desempenho médio das classes individuais pode ser a média macro, correspondente à média aritmética do desempenho de cada uma das classes, ou a média ponderada, correspondente à média de cada uma das classes ponderada pela respectiva quantidade de registros correspondentes (KUBAT, 2017).

2.1.5.2. Mecanismos de Validação de Modelos de Classificação

Os diferentes mecanismos de validação por meio do qual as métricas de desempenho dos modelos de aprendizagem de máquinas são mensuradas podem ser via utilização de registros da base de dados exclusivamente alocados ou para treinamento do modelo ou para teste, ou ainda utilizando, de modo não concomitante, os registros para ambas as finalidades, sem ou com a possibilidade de duplicata no conjunto de treinamento.

A primeira maneira, denominada *holdout*, que utiliza registros só para treinamento ou só para teste, divide e aloca os registros de modo aleatório, segundo determinada porcentagem, geralmente próxima à proporção de 2:3 para treinamento e teste, respectivamente, e realiza a avaliação do modelo uma única vez (AGGARWAL, 2015).

A segunda maneira, denominada *cross-validation* (CV) ou *k-fold cross validation*, que utiliza todos os registros para treinamento e também para teste de modo não concomitante e sem a ocorrência de duplicata, divide os registros em k partes iguais, comumente 10, alocando 1 dessas partes para teste e o restante para treinamento, e realiza a avaliação sucessiva do modelo na mesma quantidade de vezes, k , em que os registros foram divididos (GERÓN, 2019).

A terceira maneira, denominada *bootstrap*, que também utiliza todos os registros para treinamento e também para teste, de modo não concomitante mas com a possibilidade de ocorrência de duplicata no conjunto de treinamento, seleciona os registros via amostragem com reposição, o que acarreta na ocorrência de registros duplicados, alocando os selecionados, na mesma quantidade de registros total da base de dados, para treinamento, e os demais para teste, e realiza a avaliação do modelo n vezes (AGGARWAL, 2015).

2.1.6. Interpretação de Modelos de Classificação

A interpretação dos modelos de aprendizagem de máquina consiste no entendimento local e/ou global de como o modelo produziu os resultados gerados e pode ser decorrente da análise da estrutura intrínseca do próprio modelo, como no caso da análise da estrutura de uma árvore de decisão, ou independente dessa estrutura, geralmente via análise dos pares resultado gerado-variável de entrada. Essas análises são expressas em informações quantitativas da estrutura do próprio modelo, como dos atributos e respectivos particionamentos do modelo de árvore de decisão, ou em métricas estatísticas, intrínsecas ou não aos modelos, como o caso da *feature importance* e *permutation importance*, respectivamente (MOLNAR, 2022).

As métricas *feature importance* e *permutation importance* permitem a interpretação global dos modelos por meio da atribuição de valores de importância às variáveis de entrada em termos das contribuições que tiveram para as predições realizadas. A *feature importance* constitui uma métrica inerente aos modelos de árvores de decisão e decorre do quanto os atributos reduziram o índice gini quando dos particionamentos ao longo do espaço dimensional da árvore (MOLNAR, 2022). A *permutation importance*, delineada por Breiman (2001), de modo diferente, é uma métrica extrínseca ao modelo e decorre do quanto os atributos, quando submetidas a ruído, um a um, ocasionaram piora no desempenho do modelo sob análise.

A métrica Shapley, desenvolvida por Shapley (1953) e fundamentada na teoria dos jogos de Neumann e Morgenstern (1944), por outro lado, permite a interpretação local dos modelos por meio da designação, a cada atributo de determinado registro, de uma quantia que corresponderia ao quanto cada um teria incorporado ao valor que agregado àquele que seria o resultado *a priori* do registro, levando em consideração o conjunto completo em que está inserido e desconsiderando as variáveis de entrada, resultaria no valor final da predição.

O método LIME (*Local Interpretable Model-agnostic Explanations*), desenvolvido por Ribeiro, Singh e Guestrin (2016), consiste num algoritmo que também permite a interpretação local dos modelos de aprendizagem de máquina, mas por intermédio de algum outro modelo

que seja interpretável por si só e que, envolvendo o modelo em questão e fazendo uso das variáveis de entrada desejadas ou de combinações das mesmas, não necessariamente todas, possibilita, indiretamente, a interpretação local do modelo sob análise.

O método SHAP (*Shapley Additive exPlanations*), desenvolvido por Lundberg e Lee (2017), por sua vez, permite tanto a interpretação local quanto global do modelo via agregação das explicações locais de cada registro, por meio de uma metodologia desenvolvida a partir de outros métodos, denominados pelos autores de “incorporação de atributo aditivo”, dentre os quais está o anteriormente descrito, o LIME.

Esses métodos de “incorporação de atributo aditivo” são passíveis de serem adaptados aos resultados de unicidade da teoria dos jogos sobre a alocação justa de crédito, segundo a qual a descoberta dos pesos do modelo linear explicativo de variáveis binárias do qual esses métodos comungam, atende a três requisitos relacionados à compatibilidade entre o modelo explicativo e o de aprendizagem de máquina sob análise, sendo esses requisitos a acuracidade local, consistência e omissão (LUNDBERG; LEE, 2017).

Enfim, o método SHAP calcula a métrica Shapley, mas considerando uma função de esperança condicional do modelo de aprendizagem de máquina sob análise, e resulta nos valores relativos a cada registro e atributo, denominados SHAP, que correspondem ao quanto cada atributo teria incorporado ao valor/probabilidade base do registro em questão pertencer àquela classe (LUNDBERG; LEE, 2017).

2.2. Classificação na Gestão de Suprimentos

A literatura sobre classificação realizada no contexto da gestão de suprimentos abrange aplicações que perpassam a gestão de portfólio de itens e a relação entre empresa e fornecedor, sob a perspectiva de um e/ou outro, e inclui pesquisas conceituais, estudos de casos, além de análises ou validações estatísticas realizadas sobre dados de pesquisas *survey*, conforme pode-se verificar pelos trabalhos oriundos da revisão sistemática realizada por Day, Magnan e Moeller (2010) sobre portfólio de compras e classificação de fornecedores, dos quais parte está apresentada no Quadro 2.1.

Quadro 2.1 - Trabalhos que abordaram Classificação no contexto da Gestão de Suprimentos

Fonte	Tipo de Pesquisa	Objeto de Entrada	Atributos de Entrada e Classes	Objeto da Classif.	Atributo de Saída e Classes	Desdobramento
Kraljic (1983)	Conceitual + Estudo de Casos	Itens	Impacto no Lucro (alto e baixo)	Itens	Tipos (estratégicos, gargalo, alavancáveis e não críticos)	Abordagens de Compra
		Itens	Risco de Suprimento (alto e baixo)			
		Itens	Complexidade do Mercado Fornecedor (alto e baixo)	Itens	Estágios de Aquisição (gestão de aquisição, de materiais, de abastecimento e de suprimentos)	Estratégia de Suprimentos
		Itens	Importância da Aquisição (alta e baixa)			
		Empresa-Fornecedor	Força da Empresa (alta, média e baixa)	Empresa	Diretriz Estratégica (diversificar, balancear, explorar vantagem)	Política de Compra
		Fornecedor-Empresa	Força do Fornecedor (alta, média e baixa)			
Olsen e Ellram (1997)	Conceitual	Itens	Importância (alta e baixa)	Itens	Tipos (estratégicos, gargalo, alavancáveis e não críticos)	Estratégia de Relacionamento Ideal Empresa-Fornecedor
		Itens	Dificuldade na Gestão da Aquisição (alta e baixa)			
		Fornecedor	Atratividade (alta, média e baixa)	Empresa/Fornecedor	Relacionamento Real Empresa-Fornecedor (1 a 9)	Ações para a Estratégia de Relacionamento Ideal com o Fornecedor
		Empresa/Fornecedor	Força da Relação (alta, média e baixa)			
Dyer, Cho e Chu (1998)	Estudo de Casos	Itens e Fornecedores	Importância (estratégicos e não estratégicos)	Empresa-Fornecedor	Relacionamento Ideal (comerciais e parcerias estratégicas)	Práticas de Gestão de Relacionamento por parte da Empresa
Tang (1999)	Conceitual + Estudo de Casos	Itens	Importância Estratégica (alta e baixa)	Empresa-Fornecedor	Relacionamento Ideal (vendedor, preferido, exclusivo e parceiro)	Características Operacionais
		Empresa	Poder de Barganha (alto e baixo)			
Nellore e Söderquist (2000)	Estudo de Casos	Itens	Tipos (estratégicos, gargalo, alavancáveis e não críticos)	NA	NA	Designação das respectivas categorias de geradores de especificação, de fornecedor, de especificação e de relação aos diferentes tipos de itens
		Itens	Geradores de Especific. (empresa, fornecedor e ambos)			
		Fornecedor	Categorias (contratual, imaturo, maduro e parceiro)			
		Itens	Categoria de Especific. (padrão, restritiva e colaborativa)			
		Empresa/Fornecedor	Relação no codesenvolvimento (do fornecedor, da empresa e de ambos)			
Kaufman, Wood e	Survey + Análises	Fornecedor	Tecnologia (alta e baixa)	Fornecedor	Categorias (de commodities, especialista)	Estratégias/conduas que os Fornecedores

Fonte	Tipo de Pesquisa	Objeto de Entrada	Atributos de Entrada e Classes	Objeto da Classif.	Atributo de Saída e Classes	Desdobramento
Theyel (2000)	Estatísticas	Empresa-Fornecedor	Colaboração Empresa-Fornecedor (alta e baixa)		em colaboração, em tecnologia e fornecedor de soluções)	podem adotar para competirem no mercado
Svensson (2004)	Survey + Análises Estatísticas	Empresa/Fornecedor	Vulnerabilidade Empresa-Fornecedor (alta e baixa)	Empresa/Fornecedor	Situações de Vulnerabilidade (alta-alta, baixa-baixa, alta-baixa, baixa-alta)	NA (Obs.: O foco foi a construção do constructo de vulnerabilidade)
		Fornecedor/Empresa	Vulnerabilidade Fornecedor-Empresa (alta e baixa)			
Hallikas <i>et al.</i> (2005)	Survey + Análises Estatísticas	Empresa-Fornecedor	Dependência da Empresa (alta e baixa)	Empresa-Fornecedor	Relacionamento Real (estratégica, não estratégica e assimétrica)	Formas de Gestão de Risco e Práticas de Aprendizagem (Colaborativa)
		Fornecedor-Empresa	Dependência do Fornecedor (alta e baixa)			
Caniëls e Gelderman (2007)	Survey + Análises Estatísticas	Itens	Tipos (estratégicos, gargalo, alavancáveis e não críticos)	NA	NA	Designação das categorias de dependência geral e relativa na relação empresa/fornecedor aos diferentes tipos de itens
		Empresa/Fornecedor	Dependência Relativa (domínio da empresa, do fornecedor e equilíbrio)			
		Empresa Fornecedor/	Dependência Geral (alta, moderada e baixa)			

Fonte: A autora (2022)

O escopo e foco desses trabalhos, conforme apresentado no Quadro 2.1, vai desde a classificação dos itens, fornecedores ou de suas relações, até o desdobramento dessas classificações em abordagens de suprimentos, políticas de compra e estratégias de relacionamento, passando pela associação de diferentes classes já existentes na literatura, além da definição de constructos que baseiam a classificação empregada.

Nesse contexto, Kraljic (1983) considerou, com relação aos itens do portfólio de suprimentos, duas classes quanto ao impacto no lucro e ao risco de suprimento para a empresa, alto e baixo. A partir da relação dessas classes gerou o que denominou de classificação do portfólio de compras, conhecida como matriz de Kraljic, segundo a qual os itens são enquadrados dentre estratégicos, gargalo, alavancáveis e não críticos. Em seguida, para cada um desses tipos, descreveu diferentes abordagens de compra, atividades principais, informações requeridas e nível de decisão relacionadas, e atrelou estágios de sofisticação de aquisição.

Ainda, considerou duas classes para o grau de complexidade do mercado fornecedor e para a importância da aquisição do item para a empresa, alta e baixa, e também a partir da relação dessas classes, delineou uma matriz a fim de classificar os diferentes estágios de sofisticação de aquisição, desde gestão de compras (operacional), de materiais, de abastecimento até gestão de suprimentos (estratégico). A seguir, para cada uma dessas categorias de estágio de sofisticação, distinguiu determinado tipo de item, seu nível de

disponibilidade característico, de fornecedores típicos, além do grau de centralização, critérios-chave de desempenho e horizonte de tempo da gestão.

Por fim, considerou três classes para a força da empresa e do fornecedor, alta, média e baixa, e a partir da relação entre as mesmas delineou uma matriz de portfólio de aquisição com a finalidade de atribuir uma diretriz específica a ser adotada pela empresa com relação aos itens do tipo estratégico, dentre diversificar, balancear ou explorar vantagem. Posteriormente, para cada uma dessas diretrizes e com vistas a subsidiar a elaboração de planos de ação por parte da empresa, diferenciou condutas com relação a volume e preço de compra, cobertura contratual, novos fornecedores, níveis de estoque, logística, agregação de valor e produção ou substituição.

Olsen e Ellram (1997), de maneira similar a Kraljic (1983), consideraram duas classes para a importância dos itens do portfólio da empresa e para a dificuldade na gestão da aquisição, alta e baixa, em função, para ambos, da ponderação dos fatores que compõem tais atributos, conforme tabulados pelos autores. A partir da relação dessas classes, delinearam uma matriz de modelo de portfólio a fim de classificar os itens dentre os mesmos tipos já definidos por Kraljic (1983). Em seguida, preconizaram que, idealmente, a empresa deve adotar uma estratégia de relacionamento diferente com o fornecedor de cada tipo de item.

Posteriormente consideraram três classes para atratividade do fornecedor e para força do relacionamento empresa/fornecedor, alta, média e baixa, também em função da ponderação dos fatores que compõem tais atributos, conforme explicitados pelos autores. A partir da relação dessas classes delinearam uma matriz com a finalidade de classificar os relacionamentos reais que a empresa tem com seus fornecedores, dentre 1 dos 9 quadrantes possíveis. Por fim, também com vistas a fomentar a construção de planos de ação por parte da empresa, os autores sugeriram a necessidade de haver um alinhamento entre o tipo de relacionamento que a empresa tem com seu fornecedor e a estratégia de relacionamento que seria ideal, e de coerência entre o nível de recursos alocados para o relacionamento e a atratividade do fornecedor ou força da relação.

Dyer, Cho e Chu (1998) consideraram, no contexto da indústria automotiva, duas categorias de itens, estratégicos e não estratégicos, as quais extrapolaram para os respectivos fornecedores. Diferentemente de Kraljic (1983) e Olsen e Ellram (1997), não relacionaram outras classes a essas categorias, mas, com base em estudos de caso da indústria automotiva estadunidense, japonesa e coreana, atribuíram diretamente para cada um dos tipos de fornecedores duas abordagens de relacionamento por parte da empresa, denominadas de comerciais e parcerias estratégicas. Posteriormente, desdobraram esses tipos de relacionamento em práticas de gestão de fornecedores que a empresa poderia adotar.

Tang (1999) considerou duas classes para a importância estratégica de itens e para o poder de barganha da empresa, alto e baixo, e as relacionou em uma matriz a fim de atribuir diferentes tipos de relacionamentos por parte da empresa com relação aos fornecedores dos itens, dentre vendedor, preferido, exclusivo e parceiro. Em seguida, diferenciou características operacionais para cada um desses tipos de relacionamento em termos de contrato aplicável, duração do contrato, quantidade de fornecedores, tipo de item correspondente, frequência de troca de informações, esquema de preço, frequência de entrega, envolvimento ou não da alta administração e necessidade de desenvolvimento ou não de fornecedor. Por fim, respaldaram a estrutura proposta descrevendo alguns casos reais.

Nellore e Söderquist (2000), consideraram já os tipos de itens de portfólio propostos por Kraljic (1983) e, diferentemente de Kraljic (1983), Olsen e Ellram (1997) e Tang (1999), com vistas a ampliar a abordagem de modelo de portfólio e englobar, para além dos fornecedores e relações envolvidas, também o processo de especificação e desenvolvimento de produto, associaram a esses tipos de itens três possíveis classes de geradores de especificações, dentre empresa, fornecedor e ambos. A partir de estudos de casos da Toyota e outros fabricantes automotivos, estabeleceram, então, as correspondências entre essas classes, designando a cada tipo de item o gerador de especificação adequado.

Posteriormente, consideraram também quatro categorias de fornecedores propostas por Kamath e Liker (1994), contratual, imaturo, maduro e parceiro, três categorias de especificação, padrão, restritiva e colaborativa, além de três classes de relação empresa/fornecedor para o caso em que as especificações partem de ambos, sendo a que o fornecedor codesenvolve produtos com a empresa, a que a empresa codesenvolve com o fornecedor e a que ambos o fazem de modo bilateral, e, da mesma forma, estabeleceram correspondências designando a cada tipo de item, a categoria de fornecedor, de especificação adequada e de relação empresa/fornecedor.

Kaufman, Wood e Theyel (2000) consideraram duas classes de tecnologia e de colaboração empresa-fornecedor, alta e baixa, e, a partir de análises estatísticas realizadas sobre informações levantadas via pesquisa *survey* aplicada a fornecedores de indústrias manufatureiras estadunidenses, e baseada no quesito incerteza de informação, relacionaram as classes em uma matriz/tipologia classificando os fornecedores dentre de *commodities*, especialista em colaboração, em tecnologia e fornecedor de solução de problemas. Em seguida, desdobraram as categorias de fornecedores em estratégias que poderiam adotar para competir no mercado e, a partir de uma análise fatorial implementada sobre resultados da pesquisa *survey*, caracterizaram o tipo fornecedor de solução comparando aos demais com relação à quantidade de funcionários, volume de vendas, rentabilidade, exportação e faixa salarial.

Svensson (2004) considerou duas classes para vulnerabilidade interativa na relação entre empresa e fornecedor, alta e baixa, segundo a perspectiva de cada um, e as relacionou em uma matriz/tipologia classificando situações de vulnerabilidade dentre mutualmente baixa ou alta e predominante para a empresa ou fornecedor. A partir de análises estatísticas realizadas sobre informações levantadas via pesquisa *survey* aplicada a uma indústria automotiva e seus principais fornecedores, e baseada no constructo de vulnerabilidade construído pelo autor em função do grau de dependência de tempo e relacionamento na relação empresa/fornecedor, o autor classificou situações de vulnerabilidade a partir da convergência ou divergência entre as percepções das partes e salientou a necessidade de se ter um planejamento de contingência dinâmico e consciente que considere essa dependência.

Hallikas *et al.* (2005) consideraram duas classes para o risco de dependência da empresa com relação ao fornecedor e do fornecedor com relação à empresa, alta e baixa, e, a partir de uma análise de cluster realizada sobre informações levantadas via pesquisa *survey* aplicada a fornecedores de uma indústria metalúrgica finlandesa, relacionaram as classes em uma matriz/tipologia verificando a ocorrência quanto às quatro categorias de relação empresa-fornecedor propostas por Bensaou (1999), estratégica, não estratégica, assimétrica para a empresa e assimétrica para o fornecedor, das quais constataram as três primeiras. Em seguida, desdobraram essas categorias em formas de gestão de risco e práticas de aprendizagem a partir de uma análise fatorial implementada sobre resultados da pesquisa *survey*.

Caniëls e Gelderman (2007), de modo similar a Nellore e Söderquist (2000), também já consideraram os tipos de itens de portfólio propostos por Kraljic (1983) e associaram a estes diferentes classes relacionadas à dependência relativa na relação empresa/fornecedor, dentre domínio da empresa, do fornecedor e equilíbrio, e à dependência geral entre ambos, dentre alta, moderada e baixa. A partir de resultados de uma análise fatorial realizada sobre informações levantadas via pesquisa *survey* aplicada a membros da associação holandesa de gestão de compras, e baseada nos constructos de dependência da empresa e do fornecedor definidos pelos autores, estabeleceram, então, as correspondências entre essas classes, atribuindo a cada tipo de item a dependência relativa e geral na relação empresa/fornecedor.

2.3. Classificação na Gestão de Estoques via Aprendizagem de Máquina

A literatura sobre classificação via aprendizagem de máquina realizada no contexto da gestão de estoques abrange tanto a reprodução de outros métodos de classificação empregados,

como os métodos de decisão/classificação multicritério, e até de simulação e otimização, quanto a reprodução de classes decorrentes diretamente do histórico de ocorrência de determinada situação, como atraso ou falta do item. Alguns desses trabalhos cuja aplicação da aprendizagem de máquina era evidenciada no próprio título, estão os apresentados no Quadro 2.2.

Quadro 2.2 - Trabalhos que abordaram Classificação via AM no contexto da G. Estoques

Fonte	Método de Classif. Prévia	Base de Dados	Aplicação/ Tipo de Operação	Objeto da Classificação /Classes Atribuídas	Modelos Empregados	Métrica(s) de Avaliação/ Resultado do Melhor Modelo	Desdobramento	
Kartal <i>et al.</i> (2016)	SAW AHP VIKOR	Real 715 itens	Centro de Distribuição/ Indústria Automotiva	Itens/ 'A', 'B', 'C'	SVM NB BN ANN	Acurácia Precisão Sensibilidade F1-score AUC ROC	94,7% 95% 94% 94% 99%	Classif. Multi- critério
Lajili, Ladhari e Babai (2016)	AHP R, ZF, Ng, H e Peer	Real 47 itens	Hospital	Itens/ 'A', 'B', 'C'	SVM KNN ANN DT N-BN	Acurácia	98%	Classif. Multi- critério
Mishra e Mohapatro (2020)	SAW	Real 500 itens	Centro de Distribuição	Itens/ 'A', 'B', 'C'	SVM NB KNN	Acurácia Precisão Sensibilidade F1-score	84,82% 85% 85% 85%	Classif. Multi- critério
Lolli <i>et al.</i> (2019)	Simulação	2 Sintéticas: 28.350 e 39.690 séries temporais	Não especificado	Itens/ '3', '5', '7' períodos para revisão	SVM DNN	Acurácia Matriz de Confusão	> 97%	Política de Reposição
Svoboda J., Minner S. (2021)	Otimização	3 Reais	Indústria e Atacadista	Itens/ K Política de Reposição	DT	NA	NA	Política de Reposição
Santis, Aguiar e Goliatt (2018)	NA	Real 8 semanas	Não especificado	Itens/ 'propenso' ou 'não propenso' a atraso/falta	DT LRG RF GBoost BLAG	AUC ROC AUC PRC	94,82% 30,7%	Política de Reposição
Ntakolia C., Kokkotis C., Karlsson P., Moustakidis S. (2021)	NA	Real 1.048.574 itens	Não especificado	Itens/ 'propenso' ou 'não propenso' a atraso/falta	ANN LRG SVM KNN RF LGBM XGBoost BLAG	Acurácia Precisão Sensibilidade F1-score AUC ROC	88% 88% 90% 89% 95%	Política de Reposição
Rahim <i>et al.</i> (2018)	NA	Real Não especif.	Não especificado	Itens/ 'alta', 'média' ou 'baixa' fornecimentos	DT	Erro	Não especif.	Previsão de Demanda
Yang <i>et al.</i> (2014)	NA	Real 1 ano	Empresa de e-commerce	Pedidos/ 'malicioso' ou 'normal'	LRG	Acurácia Sensibilidade	91% 93%	Gestão de Pedidos

Fonte: A autora (2022)

O foco desses trabalhos, conforme apresentado no Quadro 2.2, vai desde a obtenção de classes de itens segundo a classificação ABC, passando pela atribuição de classes referentes a políticas de estoque ou períodos de reposição, classificação do item quanto à propensão de vir a sofrer atraso/falta ou quanto ao seu nível de demanda, até a classificação de pedidos quanto a serem normais ou maliciosos.

Nesse contexto, Kartal *et al.* (2016) aplicaram essa técnica sobre 715 itens de um armazém de uma indústria automotiva da Turquia a partir de variáveis brutas relacionadas à criticidade, demanda e suprimento e previamente processadas pelos métodos de decisão multicritério Ponderação Aditiva Simples (*Simple-Additive Weighting*, SAW), Processo Analítico Hierárquico (*Analytical Hierarchy Process*, AHP) e VIKOR para priorização e classificação dos itens em três classes, A, B e C, a fim de apoiar a gestão da demanda e requisitos de estoque desses itens.

Os modelos de aprendizagem de máquina implementados para realizar essa classificação, tanto considerando os dados originalmente desbalanceados quanto dados balanceados por subamostragem, foram o SVM, NB, BN e ANN, gerando resultados que chegaram à acurácia média de 94,70% com relação ao modelo mais eficaz, ANN, e validando, segundo os autores, o fato de modelos de aprendizagem de máquina serem aplicáveis a problemas de classificação de estoque.

Da mesma forma, Lajili, Ladhari e Babai (2016) também aplicaram essa técnica, mas sob uma abordagem mais teórica, sobre 47 itens de um hospital de uma base de dados desbalanceada fornecida no trabalho de Flores, Olson e Dorai (1992) e proveniente do trabalho de R. A. Reid (1987), a partir das variáveis custo unitário médio, uso anual de dólar e *lead-time*, e previamente processados por seis técnicas de classificação multicritério – AHP e os modelos de otimização ponderados R, ZF, Ng, H e Peer – para priorização e classificação dos itens em A, B e C.

Os modelos de aprendizagem de máquina implementados para realizar essa classificação, tanto com adaptação prévia para lidar com dados desbalanceados por algoritmo desenvolvido pelos autores, quanto sem adaptação, foram o SVM, KNN, ANN, DT e N-BN, gerando resultados que chegaram a 98% de acurácia com relação aos modelos adaptados mais eficazes, DT, SVM e ANN, além de menores custos totais de estoque, considerando custo de manutenção e de aquisição, mas, por outro lado, segundo os autores, não atenderam à distribuição de classes preconizada pelo princípio da classificação ABC.

De maneira similar, Mishra e Mohapatro (2020) aplicaram a técnica de classificação sobre 500 itens de um centro de distribuição a partir de variáveis relacionadas à movimentação

de entrada e saída obtidas por um sistema de Identificação por Radiofrequência (*Radio-Frequency Identification*, RFID) e previamente processadas pelo método de decisão multicritério Ponderação Aditiva Simples (*Simple-Additive Weighting*, SAW) para priorização e classificação dos itens quanto à sua flutuação de demanda, em três classes, A, B e C, a fim de melhor lidar com as mudanças rápidas de demanda a que esses itens estão sujeitos.

Os modelos implementados para realizar essa classificação foram SVM, NB e KNN, gerando resultados que chegaram a 84,8% de acurácia na com relação ao modelo mais eficaz, SVM, atendendo à distribuição de classes preconizada pelo princípio da classificação ABC, e certificando, segundo os autores, que o sistema proposto pode ser utilizado para classificação de itens de estoque sob flutuações reais de demanda, além de poder ser extrapolado e adaptado também para flutuações de suprimento.

Lolli *et al.* (2019), por outro lado, aplicaram essa técnica sobre amostras de duas bases de dados geradas randomicamente, e processadas previamente via simulação a partir de variáveis referentes à demanda, tempo de reposição, custo de estoque (do pedido, de aquisição e de manutenção) e nível máximo de pedido, para realizar a classificação dos itens em A, B e C em termos do intervalo de revisão do estoque (3, 5 ou 7 períodos, respectivamente) aos quais deveriam ser submetidos a fim de gerarem menor custo de estoque.

Os modelos implementados para realizar essa classificação, a partir de atributos intrínsecos aos itens além das variáveis já utilizadas no processo de simulação, foram o SVM e DNN, gerando resultados que chegaram a mais de 97% de acurácia para ambos os modelos, e sugerindo, segundo os autores, que modelos de aprendizagem de máquina podem ser usados em sistemas de classificação de estoques.

Svoboda e Minner (2021) aplicaram essa técnica sobre três bases de dados diferentes, uma relativa a uma indústria, com 218 períodos de demanda mensal, outra relativa a um atacadista, com 1826 períodos de demanda diária, e outra relativa a uma indústria de peças de reposição aeroespacial com 84 períodos de demanda mensal, a partir de atributos dos itens como tamanhos de lote, custos de manutenção, custos de penalidade e informações de demanda, também processados previamente, mas via otimização por programação matemática, para realizar a classificação dos itens em uma das k políticas de reposição aos quais deveriam ser submetidos a fim de gerarem menor custo total de estoque considerando custo de manutenção, de reposição e de penalidade por pedidos não atendidos.

O modelo implementado para realizar essa classificação foi o DT, mas não com o tradicional algoritmo de treinamento CART e sim via algoritmo genético com vistas à minimização do custo, o que prescindiu de registros previamente classificados, gerando

resultados em termos do aumento do custo de estoque comparativamente ao método de otimização previamente empregado de na média 1,85% e 7,68%, respectivamente, sem e com restrição geral de serviço, considerados pelos autores de desempenho satisfatório.

Santis, Aguiar e Goliatt (2018), diferentemente dos autores anteriores que empregaram métodos preliminares para gerar as classificações prévias dos itens, aplicaram a técnica sobre dados históricos de 8 semanas de uma base de dados real desbalanceada obtida da seção de competição do repositório *kaggle*, que já provia a classificação dos itens em termos de terem ou não sofrido atraso/falta, a fim de predizerem a propensão ou não dos mesmos com relação a essa ocorrência a partir de seus atributos, 21, relacionados a nível de estoque atual e ótimo, tempo e quantidade em ressuprimento e em atraso, histórico do fornecedor, risco geral e previsão de vendas.

Os modelos implementados para realizar essa classificação, tanto considerando os dados originalmente desbalanceados quanto dados balanceados por subamostragem aleatória e sobreamostragem minoritária sintética, foram o DT, por meio do algoritmo CART, o LRG, e também os *ensemble* RF, GBoost e BLAG, gerando resultados que chegaram à AUC ROC de 94,82% com relação ao Gboost e à AUC precisão-sensibilidade de 30,7% com relação ao BLAG e indicando, segundo os autores, um sistema de diagnóstico de atrasos com potencial para aumentar o nível de serviço num sistema real de gestão de estoques.

Ntakolia *et al.* (2021) aplicaram a técnica de classificação sobre a mesma base de dados que Santis, Aguiar e Goliatt (2018), também a fim de predizerem a propensão ou não dos itens com relação à ocorrência de falta/atraso, mas a partir das variáveis de entrada previamente selecionadas via abordagem *wrapper* pelo método *Boost-ARoota*, totalizando 17 e sendo também relacionadas a demanda, suprimentos e nível de estoque.

Os modelos implementados para realizar essa classificação, considerando os registros previamente balanceados por subamostragem, foram os ANN, LRG, SVM, KNN, RF, LGBM, XGBoost e BLAG, gerando resultados que chegaram à acurácia de 88% com relação aos RF e BLAG. Esses autores ainda extrapolaram a avaliação dos modelos e lançaram mão do método SHAP de interpretabilidade, mensurando as contribuições de cada variável para as classificações realizadas e apresentando graficamente, para os itens classificados como propensos a falta/atraso, o valor dessas contribuições junto da indicação do valor do atributo quanto a ser alto ou baixo.

Rahim *et al.* (2018), de modo similar a Santis, Aguiar e Goliatt (2018) e Ntakolia *et al.* (2021), aplicaram a técnica sobre 41 itens de uma base de dados já previamente classificados em termos de suas quantidades de fornecimento em alto, médio ou baixo, a fim de predizerem o

nível de consumo desses itens. O modelo de aprendizagem de máquina implementado para prever essa classificação a partir dos atributos categoria e tipo dos itens foi o DT, por meio do algoritmo C4.5, gerando, segundo os autores, sem especificar métricas de avaliação, um resultado satisfatório no que tange à capacidade de aplicar essa classificação para prever os níveis de consumos dos itens.

Yang *et al.* (2014), diferentemente dos autores anteriores, aplicaram essa técnica sobre pedidos de compra, ao invés de itens, especificamente sobre 23.704.063 pedidos relativos ao período de um ano de uma base de dados da *Dangdang*, uma grande empresa de *e-commerce* da China, e também já previamente classificados quanto à ocorrência de fraude em não malicioso e malicioso, a fim de prever e prevenir a ocorrência dos últimos. O modelo de aprendizagem de máquina implementado para realizar essa classificação, a partir de atributos relacionadas a características dos pedidos em si e de seus usuários, foi o LRG, gerando resultados que chegaram à acurácia de 91% e sensibilidade de 93%, e indicando, segundo os autores que, a metodologia proposta pode ser utilizada por outras empresas.

3. MÉTODO DE PESQUISA

Este capítulo apresenta a classificação e contexto da pesquisa (3.1), suas atividades e recursos utilizados (3.2), as etapas da construção/preparação da base de dados objeto de estudo e suas características (3.3) e descreve a implementação da aprendizagem de máquina (3.4).

3.1. Classificação e Contexto da Pesquisa

A pesquisa realizada teve finalidade aplicada, caráter exploratório e abordagem quanti-qualitativa. De acordo com a classificação formalizada por Gupta, Verma e Victorino (2006), caracterizou-se como empírica, tendo tido o propósito de apoiar a gestão de portfólio no âmbito da gestão de suprimentos de uma rede varejista de supermercados por meio da implementação da tarefa de aprendizagem de máquina de classificação como meio de realizar a mineração de dados do processo de descoberta de conhecimento de base de dados (KDD) delineado por Fayyad, Piatetsky-Shapiro e Smyth (1996).

A rede de supermercados em questão realiza a classificação dos itens do seu portfólio de suprimentos quanto ao que denominou de função estratégica para assegurar que sua gestão atenda à exigência de que seu conjunto de itens se mantenha constante quanto à quantidade de itens correspondente a cada classe da função estratégica, apesar da modificação constante que esse conjunto sofre visto que, semanalmente, à medida que novos itens passam a compô-lo, outros, em mesma quantidade, têm a aquisição e comercialização descontinuada.

Atualmente essa classificação é realizada de forma subjetiva e contempla quatro classes de função estratégica denominadas pela organização de líder, vice-líder, premium e primeiro-preço. Os itens da função líder são aqueles que se diferenciam por serem de uma marca de renome e líder de mercado, assim como os da função vice-líder, mas estes com a marca considerada vice-líder de mercado. Por outro lado, os itens da função premium são aqueles mais nobres e caros, enquanto os da função primeiro-preço, mais populares e baratos.

Essas classes de função estratégica, no âmbito desta pesquisa, correspondem às classes que vieram a constituir o atributo correspondente a variável de saída dos modelos implementados e cuja classificação foi reproduzida a partir dos demais atributos dos itens, que então constituíram as variáveis de entrada dos modelos. Tais variáveis de entrada e saída dos modelos formaram a base de dados construída e preparada para se tornar objeto desta pesquisa.

Nesse contexto, o ponto de partida da pesquisa consistiu na manipulação de bases de dados avulsas relativas à organização e que posteriormente deram origem à base de dados objeto da pesquisa, sobre a qual, então, os modelos de classificação foram implementados e aplicadas técnicas de seleção de atributos e de interpretação do modelo com melhor desempenho quando implementado sobre a respectiva base de dados reduzida pela melhor técnica de seleção.

As bases de dados avulsas que deram origem à base de dados objeto desta pesquisa possuíam as características apresentadas na Tabela 3.1.

Tabela 3.1 – Características das Bases de Dados Avulsas

Bases de Dados Avulsas	Conteúdo (relativo aos itens)	Quantidade de Itens	Quantidade de Atributos	Extensão do Arquivo
Base 1	Características	138.888	*140	.xlsx
Base 2		14.546	°15	.csv
Base 3	Situação Ativa/Bloqueada	145.298	1	.xlsx
Base 4	Valor divergente de estoque	10.829	1	.csv
Base 5	Quantidade Recebida	▼ 12.668	1	.csv
Base 6	Quantidade Vendida	▼ 15.627	1	.csv

Obs.: *Engloba 8 variáveis que foram expandidas de 2, desconsideradas, com informações aglutinadas.

Obs.: °Quantidade de variáveis redundantes com a Base 1: 13.

Obs.: Quantidade de registros de recebimento e venda desagregados por período: ▼ 63.241 e ▼ 139.844.

Fonte: A autora (2022)

As bases de dados avulsas, conforme apresentado na Tabela 3.1, totalizavam 6, sendo duas com atributos relativos a características dos itens, uma com indicação da situação quanto a estarem ativos ou bloqueados no sistema da organização e três relativas ao período de um ano-calendário, 2020, contendo o valor de divergência de estoque que esses itens apresentaram em um dos inventários semestrais realizados, além de, respectivamente, as quantidades recebida e vendida desagregadas por período. Essas bases estavam em formatos .xlsx ou .csv e nenhuma possuía dicionário de dados que informasse de que se tratava cada um de seus atributos.

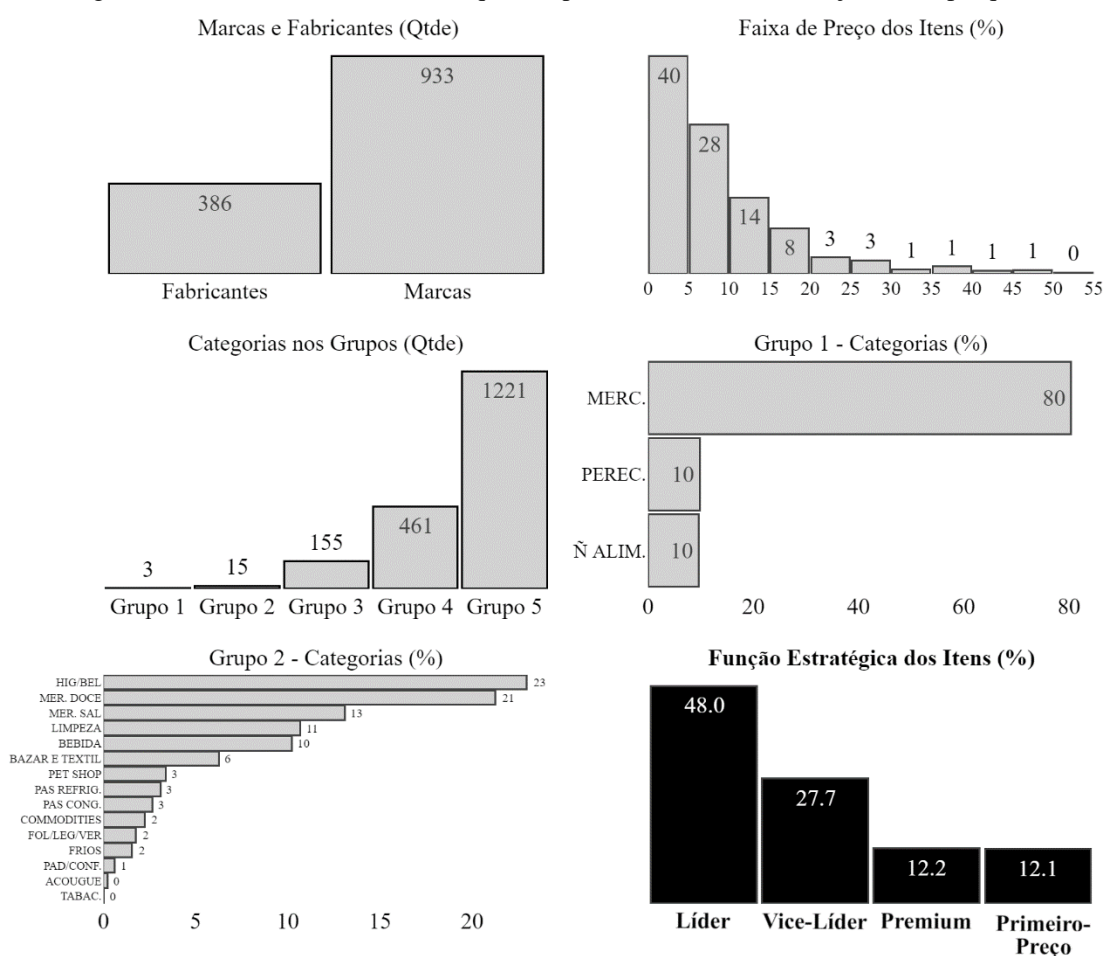
As bases de dados 1 e 2, juntas, possuíam informações dos itens referentes a fabricante, marca, preço de venda, custo médio, grupos em diferentes níveis de detalhamento variando desde categorias mais amplas até a descrição completa do item, tempo de reposição de estoque e tempo de entrega, quantidade de camadas de itens quando transportados em palete, tamanho do lastro e de sua embalagem individual. Também, informações referentes a custo de aquisição

e quantidade relativas à última compra, nome do fornecedor vendedor, forma de aquisição (se pela unidade de supermercado ou pela rede) versus armazenamento (intermediado pelo centro de distribuição da rede ou não), quantitativo disponível em estoque, custos operacionais/tributários e códigos fiscais.

Ainda, informações referentes a datas de cadastro/alterações realizadas no sistema e da última aquisição do item, nome de usuário que tenha realizado alguma edição ou movimentação, diferentes tipos de identificação do item, descrição, peso, dimensões, formato de etiqueta e prazo de validade, além de informações qualitativas quanto à localização de armazenamento no centro de distribuição, quanto a ser sazonal ou não, passível de troca ou não, sua abrangência territorial dentre nacional ou regional e, por fim, a classe de função estratégica que desempenha para a organização.

Os itens correspondentes aos registros da base de dados objeto desta pesquisa possuíam as características apresentadas na Figura 3.1.

Figura 3.1 – Características dos itens que compõem a Base de Dados objeto desta pesquisa



Fonte: A autora (2022)

Esses itens, conforme apresentado na Figura 3.1, eram de 386 fabricantes e 933 marcas distintas e tinham preços que variam de centavos até cerca de trezentos reais, com aproximadamente 70% sendo até R\$9,99, cerca de 90% até R\$19,99 e cerca de 10%, acima desse valor. Os itens eram categorizados pela rede de supermercados em 5 grupos diferentes com níveis de detalhamento de categorias aumentando no sentido do primeiro para o quinto.

O grupo 1 era composto por três categorias, mercearia, perecíveis e não alimentar, representadas por, respectivamente, cerca de 80%, 10% e 10% dos itens. O grupo 2 era composto por 15 categorias das quais as 5 primeiras - higiene/beleza, mercearia doce, mercearia salgada, limpeza e bebidas - compreendiam cerca de 80% dos itens e as 10 restantes cerca de 20%. Os grupos 3, 4 e 5, por sua vez, eram compostos por 155, 461 e 1221 categorias.

As funções estratégicas que esses itens desempenham para a rede de supermercados eram representadas 48% por itens da classe líder, 27,7% da vice-líder, 12,2% premium e 12,1% primeiro-preço. Esses valores indicam maior representatividade de itens líder e vice-líder, com quase 3/4 da frequência relativa total da base de dados, e uma menor representatividade de itens premium e primeiro-preço, com pouco menos de 1/4 da frequência relativa total.

3.2. Atividades e Recursos da Pesquisa

As atividades da pesquisa foram baseadas no processo de extração de conhecimento de base de dados delineado por Fayyad, Piatetsky-Shapiro e Smyth (1996) e foram desenvolvidas em duas etapas, conforme apresentado na Figura 3.2.

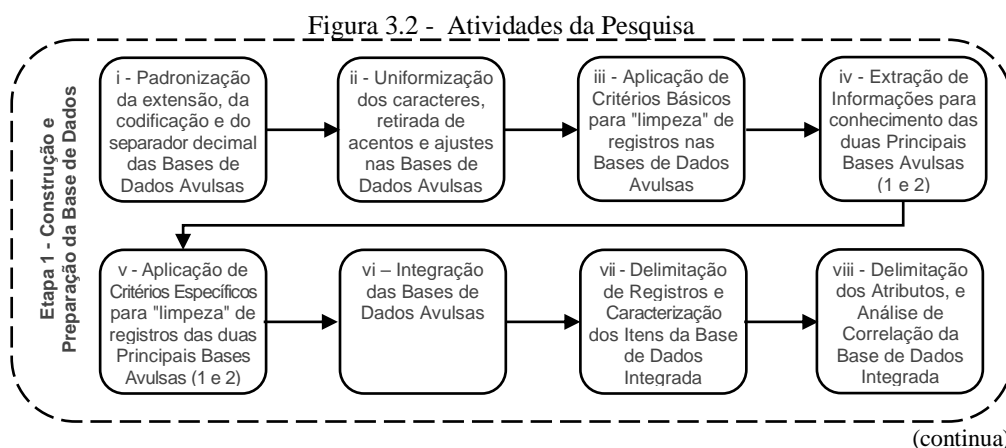
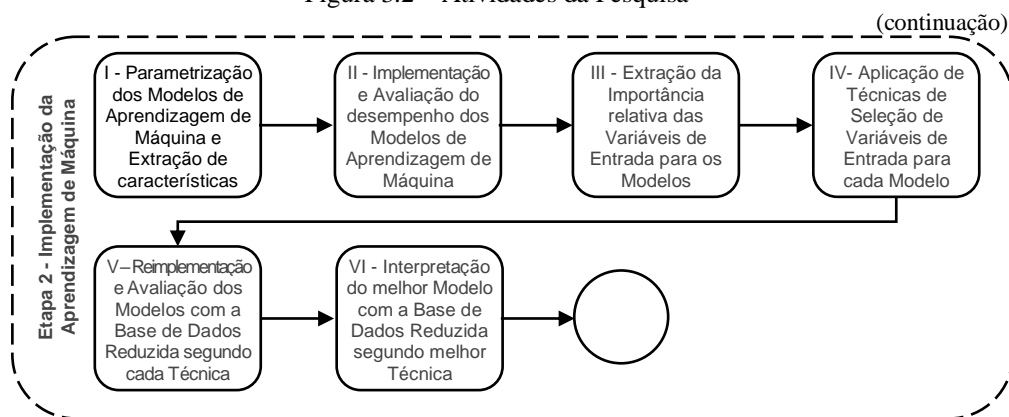


Figura 3.2 – Atividades da Pesquisa



Fonte: A autora (2022)

As etapas da pesquisa, conforme apresentado Figura 3.2, foram duas: a de construção e preparação da base de dados para posterior submissão aos modelos de aprendizagem de máquina escolhidos para esse fim (etapa 1) e a etapa de implementação da aprendizagem de máquina propriamente dita (etapa 2) por meio dos modelos de classificação de k-vizinhos mais próximos (KNN), árvore de decisão (DT), floresta aleatória (RF), máquinas de vetores de suporte (SVM) e rede neural artificial (ANN) sob arquitetura do perceptron multicamadas.

A etapa 1, de modo análogo ao processo de descoberta de conhecimento delineado por Fayyad, Piatetsky-Shapiro e Smyth (1996), compreendeu os passos de criação da base de dados, correspondente à atividade de integração das bases de dados avulsas (atividade *vi*), limpeza e pré-processamento de dados, correspondente à atividade de delimitação de registros (atividade *vii*) e, por fim, redução de dados, correspondente à atividade de delimitação dos atributos da base de dados (atividade *viii*). As atividades *i* a *v* foram preparatórias para as posteriores devido à pesquisa ter partido de mais de uma base de dados, todas avulsas mas relacionadas entre si.

A etapa 2 compreendeu os passos de parametrização dos modelos, correspondente à atividade de mesmo nome (atividade I), mineração de dados, correspondente à atividade de implementação dos modelos de aprendizagem de máquina (atividade II) e, por fim, interpretação dos padrões minerados, correspondente à atividade de interpretação do melhor modelo (atividade VI). As atividades III, IV e V, de extração de importância, aplicação de técnicas de seleção de variáveis de entrada e reimplementação/avaliação dos modelos, respectivamente, foram voltadas à viabilização da última atividade, de interpretação do modelo.

Essas atividades foram realizadas via linguagem de programação *Python 3.8.8* implementada no ambiente de desenvolvimento integrado (*Integrated Development Environment, IDE*) *Jupyter* com uso do *plugin* da extensão *jupyter_contrib_nbextensions* para congelamento de células, executado no navegador *Microsoft Edge* em um computador com

sistema operacional de 64 bits *Windows 10 Home Single Language* versão 21H2, processador Intel(R) Core(TM) i7-7500U com núcleos de frequência 2.70 GHz-2.90 GHz e 16,0 GB de memória RAM instalada.

As bibliotecas utilizadas em cada uma das atividades da pesquisa foram a *chardet* 4.0.0, *numpy* 1.21.5, *pandas* 1.2.4, *scikit-learn* 1.0.1, *plotly* 5.6.0, *mlxtend* 0.19.0, *shap* 0.40.0 e *matplotlib* 3.5.1, conforme apresentado na Tabela 3.2.

Tabela 3.2 – Bibliotecas utilizadas

Atividade	Biblioteca					
	<i>Chardet</i> 4.0.0	<i>Numpy</i> 1.21.5	<i>Pandas</i> 1.2.4	<i>Scikit-learn</i> 1.0.1	<i>Plotly</i> 5.6.0	<i>Mlxtend</i> 0.19.0 <i>Shap</i> 0.40.0 <i>Matplotlib</i> 3.5.1
i-Padronização da extensão, da codificação e do separador decimal das Bases de Dados Avulsas	x	x	x			
ii-Uniformização dos caracteres, retirada de acentos e ajustes nas Bases de Dados Avulsas		x	x			
iii-Aplicação de Critérios Básicos para "limpeza" de registros nas Bases de Dados Avulsas			x			
iv- Extração de Informações para conhecimento das duas principais Bases Avulsas (1 e 2)			x			
v- Aplicação de Critérios Específicos para "limpeza" de registros das principais Bases Avulsas (1 e 2)			x			
vi- Integração das Bases de Dados Avulsas			x			
vii- Delimitação de Registros e Caracterização dos itens da Base de Dados Integrada			x		x	
viii- Delimitação dos Atributos, e Análise de Correlação da Base de Dados Integrada			x	x		
I-Parametrização dos Modelos de Aprendizagem de Máquina e extração de características		x	x	x	x	
II-Implementação e Avaliação do desempenho dos Modelos de Aprendizagem de Máquina		x	x	x	x	
III-Extração da Importância relativa das Variáveis de Entrada para os Modelos			x	x		
IV-Aplicação de Técnicas de Seleção de Variáveis de Entrada para cada Modelo			x	x	x	
V-Reimplementação e Avaliação dos Modelos com a Base de Dados Reduzida segundo cada Técnica		x	x	x	x	
VI-Interpretação do melhor Modelo com a Base de Dados Reduzida segundo melhor Técnica		x	x			x x

Fonte: A autora (2022)

A biblioteca inicial utilizada, conforme apresentado na Tabela 3.2, foi a *chardet* para identificação da codificação de caracteres das bases de dados avulsas. Em seguida, a *numpy* e *pandas* para manipulação das bases em formato de matrizes ou tabelas, além de para conversão de números em formato exponencial para decimal, no caso do *numpy*. O *matplotlib* e *plotly* foram utilizados para produção dos gráficos que caracterizaram os itens da base de dados objeto desta pesquisa e dos que subsidiaram a configuração dos parâmetros, avaliação e interpretação dos modelos de aprendizagem de máquina. O *scikit-learn*, *mlxtend* e *shap*, por fim, foram utilizados, respectivamente, para desenvolvimento e avaliação dos modelos, seleção de variáveis de entrada via *Sequential Feature Selector* (SFS) e interpretação do melhor modelo.

Os tempos de processamento das atividades variaram de milissegundos, passando por alguns minutos a até algumas horas, tendo sido as atividades que consumiram maior tempo as de busca pela melhor configuração dos parâmetros dos modelos SVM e ANN, gastando em torno de 4h e 3h, respectivamente. Em seguida, as atividades de extração da *Permutation*

Importance das variáveis de entrada dos modelos KNN e SVM, levando cerca de 1,5h e 3h, nessa ordem, e, por fim, a atividade de aplicação da técnica de seleção de variáveis de entrada *Sequential Feature Selector* (SFS) para os modelos RF e SVM, consumindo aproximadamente 1h e 3h, respectivamente.

3.3. Etapa 1 – Construção e Preparação da Base de Dados

A primeira etapa da pesquisa, que precedeu a implementação da aprendizagem de máquina propriamente dita, consistiu na realização das atividades de *i)* padronização da extensão, da codificação e do separador decimal das 6 bases de dados avulsas, *ii)* uniformização dos caracteres, retirada de acentos e pequenos ajustes ortográficos, *iii)* aplicação de critérios básicos para limpeza de registros, *iv)* extração de informações para conhecimento das duas principais bases avulsas (1 e 2), *v)* aplicação de critérios específicos para limpeza de registros das duas bases principais, *vi)* integração das bases de dados avulsas, *vii)* delimitação de registros e caracterização dos itens da base de dados integrada e, por fim, *viii)* delimitação dos atributos e análise de correlação da base já sob a forma do objeto desta pesquisa.

A padronização da extensão das bases de dados avulsas consistiu na conversão para o formato *.csv* (*comma-separated values*) daquelas bases cujos arquivos estavam originalmente no formato *.xlsx* de planilha do Microsoft Excel. Por sua vez, a padronização da codificação de caracteres consistiu na conversão para *utf-8* daquelas bases cujos arquivos traziam a codificação *latin-1* (bases 2 e 4) e *ascii* (bases 5 e 6), e, por fim, a padronização do separador decimal consistiu no emprego do ponto como marcador decimal dos números, ao invés da vírgula.

A uniformização dos caracteres consistiu na retirada dos acentos, na transformação de todas as letras para maiúsculo e na conversão para o formato decimal dos números que estavam em formato exponencial. Ajustes realizados nessas bases compreenderam a retirada de caracteres iniciais duplicados ou errados das *strings*, como ‘Zz’, ‘Zc’ ‘C’, ‘0’, entre outros. Também, especificamente para a base avulsa 1, na expansão de dois atributos na forma de outros 8, por conterem informações aglutinadas e, especificamente para as bases 5 e 6 de quantidade recebida e vendida, na agregação dos registros por item ao invés de data, como estava.

A aplicação de critérios básicos para limpeza de registros inadequados dessas bases consistiu na eliminação de registros vazios, duplicados ou identificados por códigos de barras também duplicados ou ausentes, além daqueles que constassem com o respectivo item tendo sido vendido ou recebido no ano em questão em quantidades negativas ou nulas. Essas

eliminações resultaram no descarte de 14.178 registros, sendo 106, 0, 17, 4, 681 e 1456 das bases avulsas 1 a 6, respectivamente. As bases de quantidade recebida e vendida, especificamente, passaram a dizer respeito, juntas, a 14.957 registros/itens.

A extração de informações das duas principais bases avulsas (bases 1 e 2) consistiu no levantamento, com relação a cada um de seus atributos, do formato correspondente (*string*, *datetime*, *int* ou *float*), da quantidade de registros não nulos, da quantidade de registros únicos, do registro predominante e sua frequência relativa, da quantidade de registros ‘0’ ou ‘1’, dos tamanhos das *strings*, além da explicitação dos registros únicos. Em seguida, foram identificados os atributos vazios ou quase vazios (com 99% ou mais de registros vazios), e dos apenas com um registro único.

A aplicação de critérios específicos para limpeza de registros das duas bases avulsas principais (bases 1 e 2) consistiu na eliminação daqueles que tivessem discrepância com relação ao padrão de alguma variável em termos dos tipos de caracteres (alfabético, numérico, data, pontuação e espaço) e dos tamanhos de suas *strings*. Essa atividade e a anterior foram executadas iterativamente até que os critérios adotados contemplassem todos os necessários para a limpeza completa da base. Essa eliminação resultou no descarte de 572 e 0 registros/itens das bases avulsas 1 e 2, respectivamente.

A integração das bases de dados avulsas consistiu na junção dessas bases, com mescla dos atributos redundantes entre as bases 1 e 2, e na realização de ajustes de registros cuja informação relativa aos atributos redundantes apresentasse divergência em suas *strings* ou cuja informação constasse como ‘*null*’, mas devendo ser ‘0’ ou ‘1’. Essa base de dados integrada, ainda sem nenhuma delimitação, passou a ter as características apresentadas na Tabela 3.3.

Tabela 3.3 – Características das Bases de Dados integradas

Bases de Dados	Conteúdo (relativo aos itens)	Quantidade de Itens/ Registros	Quantidade de Atributos
Bases 1 a 6 integradas sem qualquer delimitação		°323.819	*146

Obs.: *146 = 140 + 2 + 1 + 1 + 1 + 1 (das bases de dados avulsas 1 a 6, respectivamente)

Obs.: °323.819 = 138.210 U 15.546 U 145.281 U 10.825 U 11.987 U 14.171 registros

Fonte: A autora (2022)

A delimitação de registros dessa base de dados integrada, com 323.819 itens e 146 atributos, conforme apresentado na Tabela 3.3, consistiu na eliminação (-8.294) daqueles referentes a itens que tivessem algum atributo faltando e que passaram a se tornar duplicados quando da delimitação dos atributos da base, além dos que não tivessem tido alguma venda ou recebimento no ano em questão e dos que tivessem determinadas características, conforme detalhado na Tabela 3.4.

Tabela 3.4 – Delimitações dos Registros da Base de Dados Integrada

Registros eliminados da Base de Dados Integrada	Quantidade	Porc. Corresp.
Registros com algum atributo com informação faltante	9	0,1%
Registros duplicados	2	0%
Itens que não tivessem tido alguma venda ou recebimento no ano em questão	3.756	45%
Itens do tipo pesáveis	402	5%
Itens de categoria de materiais, serviços, insumos ou não informada	255	3%
Itens com preço ou custo com valor zero	1.621	20%
Itens sem função estratégica definida	1.943	23%
Itens com função estratégica redundante com algum atributo da base	306	4%
Total de Registros eliminados		-8.294

Fonte: A autora (2022)

A eliminação de registros, conforme apresentado na Tabela 3.4, compreendeu aqueles correspondentes a itens que tivessem algum atributo com informação faltante (9), que vieram a se tornar duplicados quando da delimitação dos atributos da base (2), que não tivessem tido alguma venda ou recebimento no ano em questão (3.756), que fossem do tipo pesáveis (402), de categoria do grupo 1 correspondente a materiais, serviços, insumos ou não informada (255), que tivessem preço ou custo com valor zero (1.621), além daqueles sem função estratégica definida (1.943) ou fosse redundante com algum atributo da base (306).

Posteriormente a essa delimitação de registros foi realizada a caracterização dos itens em termos de seus atributos básicos e também de suas frequências com relação a cada uma das quatro classes de função estratégica que desempenhavam para a rede de supermercados, conforme apresentado na Figura 3.1 do subcapítulo 3.1 deste trabalho.

A base de dados integrada já com os registros delimitados passou a ter as características apresentadas na Tabela 3.5.

Tabela 3.5 – Características da Bases de Dados Integrada com Registros delimitados

Bases de Dados	Conteúdo (relativo aos itens)	Quantidade de Itens/ Registros	Quantidade de Atributos
Bases 1 a 6 integradas com registros delimitados		°6.663	*146

Obs.: *146 = 140 + 2 + 1 + 1 + 1 + 1 (das bases de dados avulsas 1 a 6, respectivamente)

Obs.: °6.663 = 14.957 – 8294

Fonte: A autora (2022)

A delimitação dos atributos da base de dados integrada com os registros já delimitados, 6.663, e os 146 atributos originais das bases de dados avulsas, conforme apresentado na Tabela 3.5, compreendeu a incorporação de novos atributos construídos a partir dos já existentes (+28) e a limitação àqueles significativos/compreensíveis e considerados viáveis/relevantes para submissão aos modelos de aprendizagem de máquina (-135), conforme detalhado na Tabela 3.6.

Tabela 3.6 – Delimitações e Construção de Atributos da Base de Dados Integrada

Atributos Incorporados/Eliminados da Base de Dados Integrada	Quantidade	Porc. Corresp.
Atributos Incorporados		+28

Lucro Unitário e Lucro Percentual	2	7%
Quantidade de itens similares nos grupos 3, 4 e 5	3	11%
De Participações Relativas da marca e fabricante no portfólio da empresa	23	82%
Atributos Eliminados		-135
<u>Atributos considerados não significativos ou não compreendidos</u>	<u>-69</u>	<u>-51%</u>
Contendo apenas um registro único	25	19%
Completamente vazios	15	11%
Com mais de 99% de registros vazios	3	2%
Redundantes com os que foram mantidos na base objeto desta pesquisa	10	7%
Não compreendidos	16	12%
<u>Atributos considerados não viáveis</u>	<u>-12</u>	<u>-9%</u>
Cuja <i>dummização</i> geraria 3.167 variáveis	7	5%
Redundantes com os que gerariam elevada quantidade de variáveis	5	4%
<u>Atributos considerados não relevantes</u>	<u>-54</u>	<u>-40%</u>
Custo e quantidade relativos à última aquisição pela empresa	4	3%
Forma de aquisição	1	1%
Fornecedor vendedor	2	1%
Quantitativo em estoque	4	3%
Custos operacionais e tributários	7	5%
Códigos fiscais	4	3%
Quantidade vendidas no período de um ano-calendário	1	1%
Quantidade recebidas no período de um ano-calendário	1	1%
Divergência de estoque no período de um ano-calendário	1	1%
Situação dos itens no sistema da empresa	1	1%
Datas de cadastros/alterações e de última aquisição	7	5%
Usuário que tenha realizado alguma edição ou movimentação	5	4%
Identificação	3	2%
Descrição	5	4%
Dimensão	3	2%
Formato de etiqueta	1	1%
Localização física no centro de distribuição	1	1%
Sazonal ou não	1	1%
Passível de troca ou não	1	1%
Nacional ou Regional	1	1%

Fonte: A autora (2022)

A construção de novos atributos a partir dos já existentes na base de dados integrada abrangeu, conforme apresentado na Tabela 3.6, além do lucro unitário e percentual, cujas informações já constavam da base mas com valores diferentes dos calculados, a quantidade de similares ao item em cada um dos grupos 3 a 5 e as participações relativas dos fornecedores nos papéis de fabricante e marca com relação ao portfólio da rede de supermercados.

A quantidade de itens similares indica em valores absolutos quanto determinado item, conforme seu detalhamento no respectivo grupo (3, 4 ou 5) possui de outros similares a si. Por exemplo, para um item identificado pelo código de barras hipotético 7890000000000 e descrito nos grupos 3, 4 e 5 como “cerveja”, “cerveja tradicional” e “cerveja tradicional pilsen”, respectivamente, seria o quanto há de cerveja no portfólio geral da rede de supermercados independentemente de fabricante e marca (grupo 3); o quanto há de cerveja tradicional (grupo 4) e, finalmente, o quanto há de cerveja tradicional pilsen (grupo 5).

Já, as participações relativas da marca e fabricante compreendem a quantidade de marcas que o fabricante do item possui no portfólio da rede de supermercados, a quantidade

absoluta de itens que tanto um quanto outro, marca e fabricante, possuem, além das participações relativas que têm em cada grupo (1 a 5), em termos de itens e de categorias. Por exemplo, com relação ao item citado anteriormente, hipoteticamente da fabricante Ambev e da marca Antarctica, seria quantas marcas diferentes a Ambev possui; quantos itens a Ambev e a Antarctica têm na organização e quantos porcentos a Ambev e a Antarctica têm de participação no grupo 1 em termos de itens e também de categorias, assim como para os demais grupos.

A limitação dos atributos aos considerados significativos/compreensíveis acarretou na eliminação dos que tinham apenas um registro único (25), dos completamente vazios (15), com mais de 99% de registros vazios (3), redundantes com os que foram mantidos (10) e que não puderam ser compreendidos (16). Já, a limitação dos atributos aos considerados viáveis acarretou na eliminação dos qualitativos correspondentes a marca, fabricante e grupos 1 a 5, que quando *dummizados* gerariam elevada quantidade de outras variáveis, 3.167, (7), além dos atributos redundantes (5) a esses.

Por fim, a limitação aos atributos considerados relevantes acarretou na eliminação dos correspondentes às informações de última aquisição pela organização (4), forma de aquisição (1), fornecedor vendedor do item (2), quantitativo em estoque (4), custos operacionais/tributários (7), códigos fiscais (4), quantidades vendidas (1), recebidas (1) e de divergência de estoque (1), além dos com informações da situação do item no sistema (1), datas (7), usuários (5), identificação (3), descrição (5), dimensão de frente, largura e profundidade (3), formato de etiqueta (1), localização de armazenamento no centro de distribuição (1), sazonalidade (1), possibilidade de troca (1) e abrangência nacional ou regional (1) do item.

Posteriormente a essa delimitação foi realizada a padronização dos atributos mantidos via método *StandardScaler* da biblioteca *scikit-learn*, que dimensiona os valores dos atributos de modo a tornar suas médias nulas e as variâncias unitárias, e também a análise das correlações existentes entre os atributos segundo o coeficiente de *Pearson*, pela biblioteca *pandas*. A *dummização* dos atributos não foi realizada visto que estes passaram a ser todos quantitativos.

A base de dados integrada já com os registros e também os atributos delimitados passou a ter as características apresentadas na Tabela 3.7.

Tabela 3.7 – Características da Base Dados Integrada com Registros/Atributos delimitados

Bases de Dados	Conteúdo (relativo aos itens)	Quantidade de Itens/ Registros	Quantidade de Atributos
	Base Objeto desta Pesquisa	6.663	*38+1

Obs.: *39 = 146 + 28 – 135

Fonte: A autora (2022)

A base de dados objeto desta pesquisa, com 6.663 registros/itens e 38 variáveis de entrada e 1 variável de saída, conforme apresentado na Tabela 3.7, correspondente ao resultado da delimitação da base de dados integrada tanto em termos de registros quanto em termos de atributos, tem os itens identificados pelos códigos de barras sendo caracterizados pelos 38 atributos de entrada e classificados quanto à variável de saída em termos de uma das quatro classes correspondentes à função estratégica que desempenham para a rede de supermercados.

3.3.1. Base de Dados objeto da Pesquisa

As 38 variáveis de entrada da base de dados objeto desta pesquisa, referentes aos 6.663 registros/itens que possui, estão detalhadas na Tabela 3.8.

Tabela 3.8 – Variáveis de Entrada da Base de Dados objeto desta pesquisa

Dimensão	Variável de Entrada	Descrição	Unidade
Características Intrínsecas aos itens	1. Preço	Preço Unitário	monetário
	2. Custo	Custo Unitário	monetário
	*3. Lucro	Lucro Unitário	monetário
	*4. Lucro_%	Lucro Percentual	percentual
	5. Peso	Peso líquido	kg
	6. Altura	Altura	cm
	7. Validade	Prazo de Validade	dias
Características Extrínsecas aos itens	8. Reposição	Tempo de Reposição	dias
	9. Entrega	Tempo de Entrega	dias
	10. Lastro	Tamanho do Lastro	unitário
	11. Camadas	Quantidade de Camadas	unitário
	12. Tamanho da Embalagem	Tamanho da Embalagem	unitário
	*13. Similares_3	Quantidade Itens similares no grupo correspondente	unitário
*14. Similares_4	unitário		
*15. Similares_5	unitário		
Participações Relativas da Marca/ Fabricante em termos de quantidade de itens ou categorias	*16. Fabricante_Marca	Quantidade de Marcas que o Fabricante possui	unitário
	*17. Marca_Prop_Geral	Participação da Marca em termos de itens no geral	percentual
	*18. Fabricante_Prop_Geral	Participação do Fabricante em termos de itens no geral	percentual
	*19. G1_Marca	Participação da Marca/Fabricante em termos de itens no grupo correspondente	percentual
	*20. G1_Fabricante		percentual
	*21. G2_Marca		percentual
	*22. G2_Fabricante		percentual
	*23. G3_Marca		percentual
*24. G3_Fabricante	percentual		
*25. G4_Marca	percentual		

termos de itens nos grupos 1 e 2. Também possuem correlação negativa as quantidades de similares e a participação da marca e do fabricante em termos da quantidade de itens nos grupos 3 a 5.

3.4. Etapa 2 – Implementação da Aprendizagem de Máquina

A segunda etapa da pesquisa, que consistiu na implementação dos modelos de aprendizagem de máquina propriamente dita, compreendeu as atividades de *I)* busca pela melhor configuração dos hiperparâmetros dos modelos e extração de características inerentes aos DT e RF, *II)* implementação e avaliação de desempenho dos modelos, *III)* extração da importância relativa das variáveis de entrada dos modelos, *IV)* aplicação de técnicas de seleção de variáveis de entrada para cada modelo, *V)* reimplementação e reavaliação dos modelos com a base de dados reduzida segundo cada técnica de seleção aplicada e *VI)* interpretação do melhor modelo com a base de dados reduzida segundo melhor técnica.

A parametrização dos modelos de aprendizagem de máquina implementados levou em conta os hiperparâmetros de cada um e foi realizada via método *GridSearchCV* da biblioteca *scikit-learn* considerando a base de dados sem qualquer redução de variável de entrada e por validação cruzada com embaralhamento prévio dos registros e 10 partes com mesma proporção de itens de cada classe. Os resultados de cada combinação das configurações dos hiperparâmetros disseram respeito às cinco métricas de avaliação utilizadas, acurácia, AUC ROC sob estratégia *one versus rest (ovr)*, sensibilidade, precisão e F1-score, e foram plotados em gráficos de modo a apoiar a identificação da combinação mais vantajosa com base na acurácia geral e AUC ROC média, nessa ordem, ponderadas com relação à complexidade dos modelos em termos dos recursos computacionais demandados e potencial de *overfitting*.

A implementação e avaliação do desempenho dos modelos foram realizadas, assim como para a busca da melhor configuração dos seus hiperparâmetros, via validação cruzada (*Cross Validation, CV*) e também com relação às cinco métricas de avaliação adotadas. Os resultados dessa avaliação foram posteriormente utilizados para posterior comparação com os resultados dos modelos implementados sobre as respectivas bases com variáveis de entrada reduzidas, tendo tido os modelos DT, RF e ANN os parâmetros relacionados à suas dimensões, profundidade e quantidade de árvore(s) no caso do DT e RF e a quantidade de neurônios no caso do ANN, arbitrariamente diminuídos antes dessa redução. O algoritmo de treinamento utilizado pelo modelo DT, padrão da biblioteca *scikit-learn* (2022), foi, segundo a própria, uma versão otimizada do *Classification and Regression Trees (CART)*.

A extração da importância relativa de cada variável de entrada do modelo foi realizada via *feature importance* para o DT e RF, exclusivamente, devido a essa importância ser uma característica inerente desses modelos e, conseqüentemente, passível de extração diretamente dos mesmos. Já, a extração da importância via *permutation importance* foi realizada para todos os modelos. Para ambos os casos essas importâncias foram extraídas somente quando dos modelos implementados com a base de dados completa com todas as variáveis de entrada.

A aplicação das técnicas de seleção de variáveis foi realizada e avaliada para todos os modelos e abrangeu, além do resultado decorrente das importâncias relativas das variáveis de entrada via *Feature Importance e Permutation Importance*, também o resultado da aplicação das técnicas *F Classification, Mutual Info Classification* e da abordagem *wrapper Sequential Feature Selector (SFS) backward*, além da consideração das variáveis selecionadas pela combinação do modelo e técnica de seleção que geraram o melhor desempenho (*RF+SFS*) para todos os outros modelos.

A reimplementação e reavaliação do desempenho dos modelos implementados sobre as correspondentes bases de dados reduzidas via respectiva técnica de seleção que gerou os melhores resultados, foi realizada também por validação cruzada (*Cross Validation, CV*) em termos do quanto cada modelo classificou os itens originalmente pertencentes a determinada classe da função estratégica como sendo pertencentes à própria classe ou às demais, dos resultados relativos a cada uma das cinco métricas de avaliação utilizadas, do comparativo desses resultados com relação aos modelos implementados sobre a base de dados sem redução de nenhuma variável e, por fim, tanto para os modelos implementados sobre a base de dados completa quanto reduzida, em termos da distribuição dos resultados de cada uma das 10 avaliações do *cross validation*, para cada métrica de avaliação.

A interpretação do modelo, realizada com relação àquele que teve o melhor desempenho quando implementado sobre a respectiva base de dados reduzida segundo a melhor técnica de seleção, foi realizada via método SHAP considerando todos os registros da base de dados. Por meio desse método foi mensurada a importância relativa das variáveis de entrada para a classificação geral e com relação a cada classe individual. Também foi apresentado graficamente, para cada classe, os valores SHAP de cada registro representando a probabilidade que cada um de seus atributos incorporou à sua probabilidade base (*a priori*) de pertencer àquela classe de função estratégica. Por fim, foi apresentado graficamente para quatro registros tomados como exemplo, os valores SHAP indicando a probabilidade que cada um de seus atributos incorporaram à sua probabilidade final de pertencer àquela classe atribuída.

4. RESULTADOS

Este capítulo apresenta os resultados da etapa de implementação da aprendizagem de máquina, contemplando desde a parametrização dos modelos de classificação implementados (4.1), passando pela seleção de variáveis de entrada (4.2) até a avaliação dos modelos (4.3) e interpretação do melhor deles implementado sobre a base de dados reduzida segundo melhor técnica de seleção (4.4).

4.1. Parametrização dos Modelos de Aprendizagem de Máquina

Os modelos de aprendizagem de máquina implementados estão apresentados na Tabela 4.1, junto dos respectivos parâmetros e variações correspondentes avaliadas.

Tabela 4.1 – Modelos e Parâmetros ajustados

Modelo	Parâmetro	Variações
1. KNN	<i>n_neighbors</i>	1, 2, 3 , 4, ... 100
	<i>metric</i>	'euclidean', ' manhattan '
	<i>weights</i>	'uniform', ' distance '
2. DT	<i>max_depth</i>	5, 10, 15, * 21 , 26, 31
	<i>criterion</i>	'gini', ' entropy '
	<i>class_weight</i>	None , 'balanced'
3. RF	<i>max_depth</i>	* 21
	<i>n_estimators</i>	5, 10, 15, 21, 26, 31, 36, 42, 84, ▽ 126 , 168, 252, 336
	<i>criterion</i>	'gini', ' entropy '
4. SVM	<i>class_weight</i>	None , 'balanced'
	<i>decision_function_shape</i>	ovr
	<i>C (Regularization)</i>	0.01, 0.1, 1.0, 10, 100 , 250, 500
5. ANN	<i>kernel</i>	'linear', 'poly', ' rbf ', 'sigmoid'
	<i>n_layers</i>	2
	<i>hidden_layer_sizes</i>	42, 84, ▽ 168 e 252
	<i>solver</i>	'lbfgs', 'sgd', ' adam '
	<i>activation</i>	'identity', 'logistic', ' tanh ', 'relu'

Obs. 1: Configurações que ficaram definidas estão em negrito.

Obs. 2: Valores redefinidos para e pós aplicação das técnicas de seleção de variáveis de entrada: *19, ▽ 52 e ▽ 65.

Fonte: A autora (2022)

O modelo KNN, conforme apresentado na Tabela 4.1, teve os parâmetros *n_neighbors*, *metric* e *weights* variados. As variações do *n_neighbors* compreenderam o intervalo de '1' a

‘100’, as da *metric* compreenderam as distâncias ‘*euclidean*’ e ‘*manhattan*’, e do *weights* compreenderam as ponderações ‘*uniform*’ e ‘*distance*’. As combinações das diferentes configurações desses parâmetros foram avaliadas com relação às métricas de desempenho e resultaram em valores dentre os apresentados na Tabela 4.2.

Tabela 4.2 – Resultados do *GridSearchCV* para o modelo KNN

Hiperparâmetros			Métricas de Avaliação (médias macro)					
metric	weights	n_neighbors	Acurácia	AUC ROC	Precisão	Sensibilidade	F1-score	
<u>Melhores resultados com relação à Acurácia</u>								
1°.	manhattan	distance	2	89,49±1,33	94,52±1,23	87,78±1,93	88,29±1,81	87,98±1,83
2°.	manhattan	uniform	1	89,45±1,30	92,19±1,13	87,74±1,90	88,27±1,79	87,95±1,80
3°.	manhattan	distance	1	89,45±1,30	92,19±1,13	87,74±1,90	88,27±1,79	87,95±1,80
4°.	manhattan	distance	3	89,30±1,44	95,49±1,02	87,56±1,67	87,82±1,97	87,66±1,79
5°.	manhattan	distance	4	88,92±1,57	95,94±1,13	87,36±1,87	87,24±2,00	87,25±1,88
<u>Melhores resultados com relação à AUC ROC</u>								
1°.	manhattan	distance	14	87,18±2,30	96,66±0,80	86,34±2,53	84,27±3,14	85,16±2,81
2°.	manhattan	distance	15	86,67±2,34	96,66±0,82	85,91±2,53	83,63±3,14	84,62±2,79
3°.	manhattan	distance	11	87,56±1,98	96,65±0,87	86,68±2,20	84,93±2,82	85,68±2,50
4°.	manhattan	distance	13	87,38±2,13	96,63±0,86	86,50±2,39	84,62±2,87	85,43±2,60
5°.	manhattan	distance	17	86,33±2,20	96,62±0,86	85,86±2,37	82,99±3,23	84,21±2,81
<u>Piores resultados com relação à Acurácia</u>								
399°.	euclidean	uniform	98	57,81±1,36	79,95±1,40	58,82±3,85	41,54±2,22	42,54±3,13
400°.	euclidean	uniform	100	57,68±1,55	79,81±1,37	59,57±4,84	41,30±2,30	42,27±3,36
<u>Piores resultados com relação à AUC ROC</u>								
399°.	euclidean	uniform	99	57,89±1,42	79,88±1,39	59,42±4,43	41,61±2,31	42,65±3,30
400°.	euclidean	uniform	100	57,68±1,55	79,81±1,37	59,57±4,84	41,30±2,30	42,27±3,36

Obs. 1: Valores dos resultados das métricas de avaliação em porcentagem.

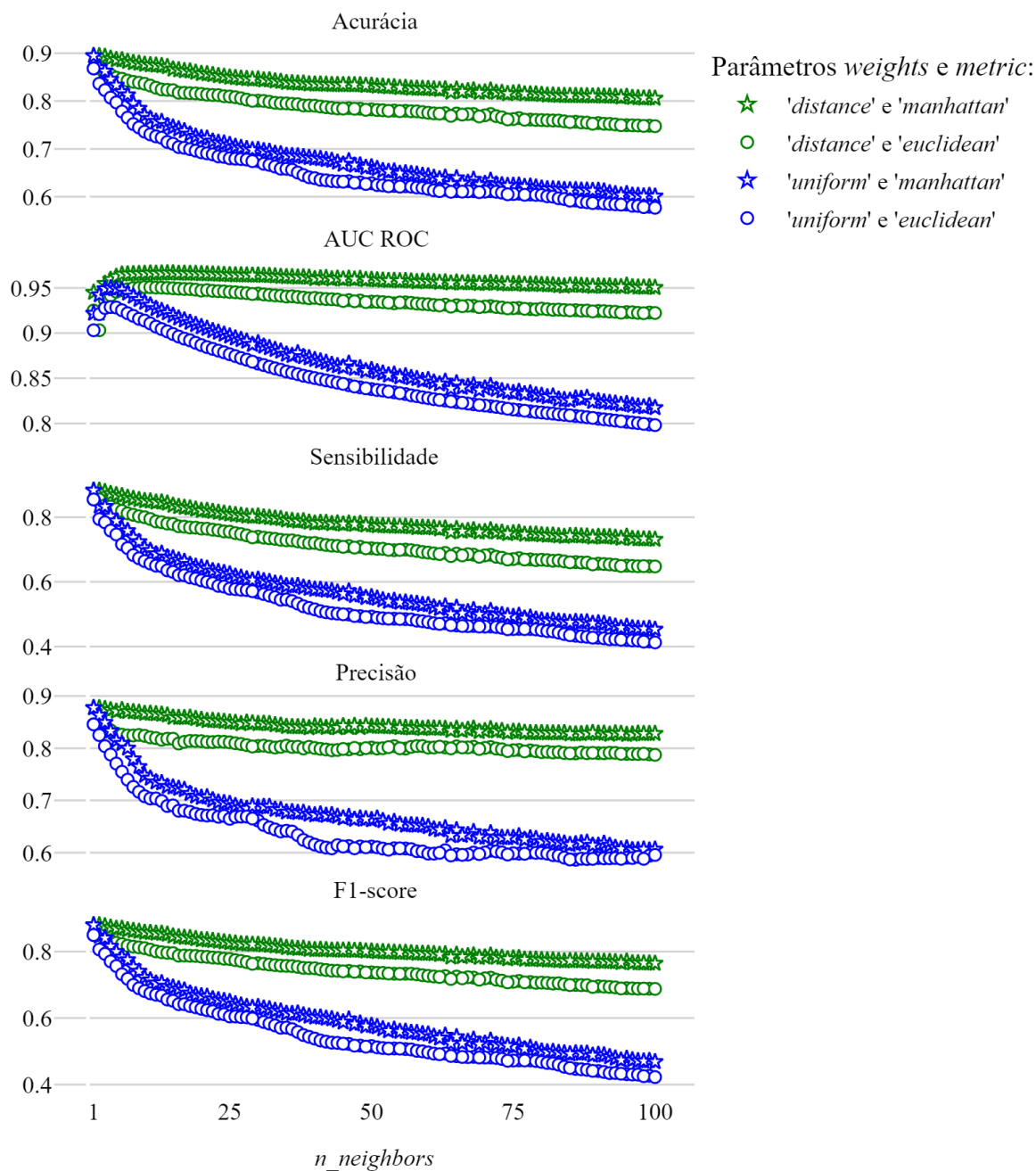
Obs. 2: Configurações que ficaram definidas estão em negrito.

Fonte: A autora (2022)

O melhor desempenho do modelo com relação à acurácia geral, conforme apresentado na Tabela 4.2, foi para os parâmetros *metric* como ‘*manhattan*’, *weights* como ‘*distance*’ e *n_neighbors* com a segunda menor quantidade experimentada, 2 (89,49%±1,33). Por outro lado, com relação a AUC ROC média, foi também para ‘*manhattan*’ e ‘*distance*’, mas com *n_neighbors* de 14 (96,66%±0,8). O pior desempenho com relação a ambas as métricas foi para *metric* como ‘*euclidean*’, *weights* como ‘*uniform*’ e para a maior quantidade de *n_neighbors* experimentada, 100 (57,68%±1,55 de acurácia e 79,81%±1,37 de AUC ROC). Por fim, levando em consideração tais resultados, a combinação escolhida foi ‘*manhattan*’, ‘*distance*’ e ‘3’.

A representação das métricas de desempenho para todas as combinações possíveis das configurações dos parâmetros variados para o modelo KNN está apresentada na Figura 4.1.

Figura 4.1 - Avaliação dos Parâmetros do Modelo KNN



Fonte: A autora (2022)

O parâmetro qualitativo que mais impactou o modelo KNN, conforme pode ser visualizado na Figura 4.1, foi o *weights*, tendo a configuração '*distance*' apresentado resultados superiores ao '*uniform*' e variando com relação à acurácia geral, à medida do aumento da *n_neighbors*, na ordem de 2,5% para a *metric* '*manhattan*' e 3,2% para a '*euclidean*', até 20,5% e 17,1%, e, respectivamente.

Ainda, o segundo parâmetro qualitativo que mais impactou o modelo, depois do *weights*, foi o parâmetro *metric*, tendo a configuração '*manhattan*' apresentado resultados um

pouco superiores ao 'euclidean', variando com relação à acurácia geral também à medida do aumento da $n_neighbors$, na ordem de 1,4% para o $weights$ 'uniforme' e 2,6% para 'distance' até 4,8% e 6%, respectivamente.

Por fim, com relação ao parâmetro quantitativo $n_neighbors$ e, independentemente da configuração dos demais parâmetros, mas especialmente para o $metric$ como 'uniform', seu aumento gerou resultados decrescentes para as métricas acurácia geral, sensibilidade média, precisão média e F1-score médio, mas até o valor aproximado de '10', gerou resultados ligeiramente crescentes, na média de 0,44%, para a métrica AUC ROC média.

O modelo DT, por sua vez, teve os parâmetros max_depth , $criterion$ e $class_weight$ variados. As variações do max_depth compreenderam as opções '5', '10', '15', '21', '26' e '31', as do $criterion$ compreenderam as opções 'gini' e 'entropy', e as do $class_weight$ compreenderam as ponderações 'balanced' ou 'None'. As combinações das diferentes configurações desses parâmetros foram avaliadas e resultaram em valores dentre os apresentados na Tabela 4.3.

Tabela 4.3 – Resultados do *GridSearchCV* para o modelo DT

Hiperparâmetros			Métricas de Avaliação (médias macro)				
criterion	class_weight	max_depth	Acurácia	AUC ROC	Precisão	Sensibilidade	F1-score
			<u>Melhores resultados com relação à Acurácia</u>				
1°. entropy	None	21	82,80±1,65	86,99±1,64	79,83±2,14	79,83±2,37	79,74±2,15
2°. entropy	None	31	82,44±1,70	86,51±1,49	79,47±1,87	79,50±2,30	79,41±2,01
3°. entropy	None	26	82,44±1,70	86,51±1,49	79,47±1,87	79,50±2,30	79,41±2,01
4°. entropy	balanced	21	82,00±2,24	86,51±1,71	78,66±2,46	79,31±2,52	78,92±2,45
5°. entropy	balanced	26	81,96±2,46	86,31±1,82	78,69±2,72	79,18±2,76	78,87±2,71
			<u>Melhores resultados com relação à AUC ROC</u>				
1°. entropy	None	15	80,07±1,87	88,11±1,68	76,87±2,40	76,36±2,51	76,51±2,44
2°. entropy	balanced	10	67,75±3,57	87,44±1,67	63,65±3,24	70,39±3,49	64,99±3,43
3°. entropy	None	10	70,31±1,99	87,39±1,11	68,31±4,43	62,75±1,59	64,36±2,27
4°. gini	balanced	15	79,53±1,43	87,24±1,12	75,47±1,96	77,91±2,12	76,50±2,02
5°. entropy	balanced	15	79,77±3,08	87,21±1,66	75,74±3,43	78,08±3,01	76,72±3,22
			<u>Piores resultados com relação à Acurácia e AUC ROC</u>				
23°. gini	balanced	5	45,91±2,27	73,9±1,61	46,71±1,95	50,91±2,03	44,25±2,09
24°. entropy	balanced	5	43,60±2,18	73,09±1,30	45,18±1,01	49,89±1,61	42,48±1,73

Obs. 1: Valores dos resultados das métricas de avaliação em porcentagem.

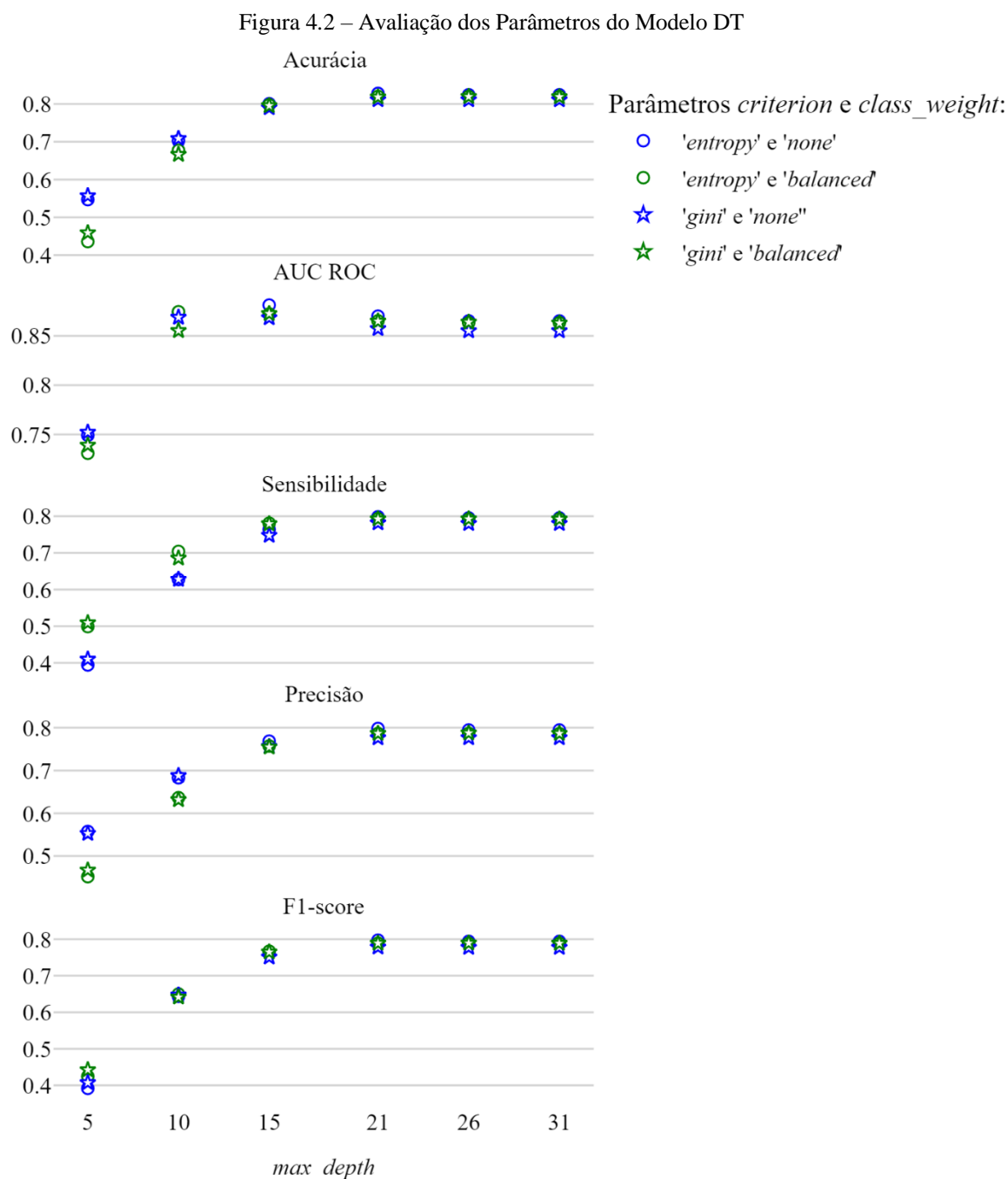
Obs. 2: Configurações que ficaram definidas estão em negrito.

Fonte: A autora (2022)

O melhor desempenho com relação à acurácia geral, conforme apresentado na Tabela 4.3, foi para os parâmetros $criterion$ como 'entropy', $class_weight$ como 'None' e max_depth de '21' (82,80%±1,65). Por outro lado, com relação a AUC ROC média, foi também para 'entropy' e 'None', mas com max_depth de '15' (88,11%±1,68). O pior desempenho com relação a ambas as métricas foi para $criterion$ como 'entropy', $class_weight$ como 'balanced' e para o menor valor experimentado de max_depth , '5' (43,60%±2,18 de acurácia e 73,09%±1,30

de AUC ROC). Por fim, a combinação escolhida foi ‘*entropy*’, ‘*None*’ e ‘*21*’, tendo este último sido arbitrariamente diminuído para ‘*19*’ antes da aplicação das técnicas de seleção de variáveis.

A representação dos resultados para todas as combinações possíveis dos parâmetros variados para o modelo DT está apresentada na Figura 4.2.



Fonte: A autora (2022)

O parâmetro que mais impactou o modelo DT, conforme pode ser visualizado na Figura 4.2, foi o parâmetro quantitativo *max_depth*, visto que à medida do seu aumento, as diferentes configurações dos demais parâmetros, *class_weight* como ‘*None*’ ou ‘*balanced*’ e *criterion*

como ‘entropy’ ou ‘gini’, foram se tornando mais irrelevantes e os resultados das diferentes métricas de avaliação foram convergindo entre si.

Entretanto, a partir do *max_depth* configurado para ‘21’, o parâmetro *class_weight* como ‘None’ e o *criterion* como ‘entropy’ passaram a gerar resultados ligeiramente superiores para as métricas de avaliação no geral, com diferença com relação à combinação ‘gini’ e ‘None’ de 1,7% para a acurácia geral e sensibilidade média, de 1,3% para a AUC ROC média e de 2,2% para a precisão média.

O modelo RF, assim como o DT, teve os parâmetros *criterion* e *class_weight* variados, além do *n_estimators*. As variações do *criterion* e *class_weight* foram as mesmas do modelo DT, ‘gini’ ou ‘entropy’ e ‘None’ ou ‘balanced’, respectivamente, e as do *n_estimators* foram as opções ‘5’, ‘10’, ‘15’, ‘21’, ‘26’, ‘31’, ‘36’, ‘42’, ‘84’, ‘126’, ‘168’, ‘252’ e ‘336’. Já, o parâmetro *max_depth* foi arbitrariamente fixado em ‘21’. As combinações das diferentes configurações foram avaliadas e resultaram em valores dentre os apresentados na Tabela 4.4.

Tabela 4.4 – Resultados do *GridSearchCV* para o modelo RF

Hiperparâmetros			Métricas de Avaliação (médias macro)					
criterion	class_weight	n_estimators	Acurácia	AUC ROC	Precisão	Sensibilidade	F1-score	
<u>Melhores resultados com relação à Acurácia</u>								
1°.	entropy	None	336	90,50±1,54	98,10±0,54	90,23±1,69	87,84±2,40	88,91±1,95
2°.	entropy	None	252	90,30±1,36	98,09±0,54	89,99±1,60	87,45±2,21	88,60±1,82
3°.	gini	None	336	90,27±1,61	98,07±0,55	90,09±1,73	87,47±2,47	88,65±2,03
4°.	entropy	balanced	336	90,23±1,48	98,11±0,54	90,07±1,62	87,50±2,18	88,66±1,82
11°.	entropy	None	126	90,05±1,53	98,05±0,53	89,62±1,86	87,30±2,47	88,34±2,09
<u>Melhores resultados com relação à AUC ROC</u>								
1°.	entropy	balanced	336	90,23±1,48	98,11±0,54	90,07±1,62	87,5±2,18	88,66±1,82
2°.	gini	balanced	336	90,12±1,64	98,11±0,52	90,14±2,01	87,20±2,47	88,51±2,13
3°.	entropy	balanced	252	90,12±1,52	98,10±0,56	89,95±1,72	87,43±2,13	88,57±1,86
4°.	entropy	None	336	90,50±1,54	98,10±0,54	90,23±1,69	87,84±2,40	88,91±1,95
11°.	entropy	balanced	126	89,93±1,43	98,06±0,58	89,98±1,93	87,08±1,92	88,38±1,76
<u>Piores resultado com relação à Acurácia</u>								
51°.	gini	balanced	5	84,29±1,99	94,39±1,03	83,63±2,43	80,42±2,46	81,81±2,29
52°.	gini	None	5	84,14±1,59	94,58±0,81	82,79±1,83	80,07±2,54	81,25±2,15
<u>Piores resultado com relação à AUC ROC</u>								
51°.	entropy	balanced	5	84,74±1,79	94,43±1,29	83,81±2,06	80,87±3,04	82,16±2,54
52°.	gini	balanced	5	84,29±1,99	94,39±1,03	83,63±2,43	80,42±2,46	81,81±2,29

Obs. 1: Valores dos resultados das métricas de avaliação em porcentagem.

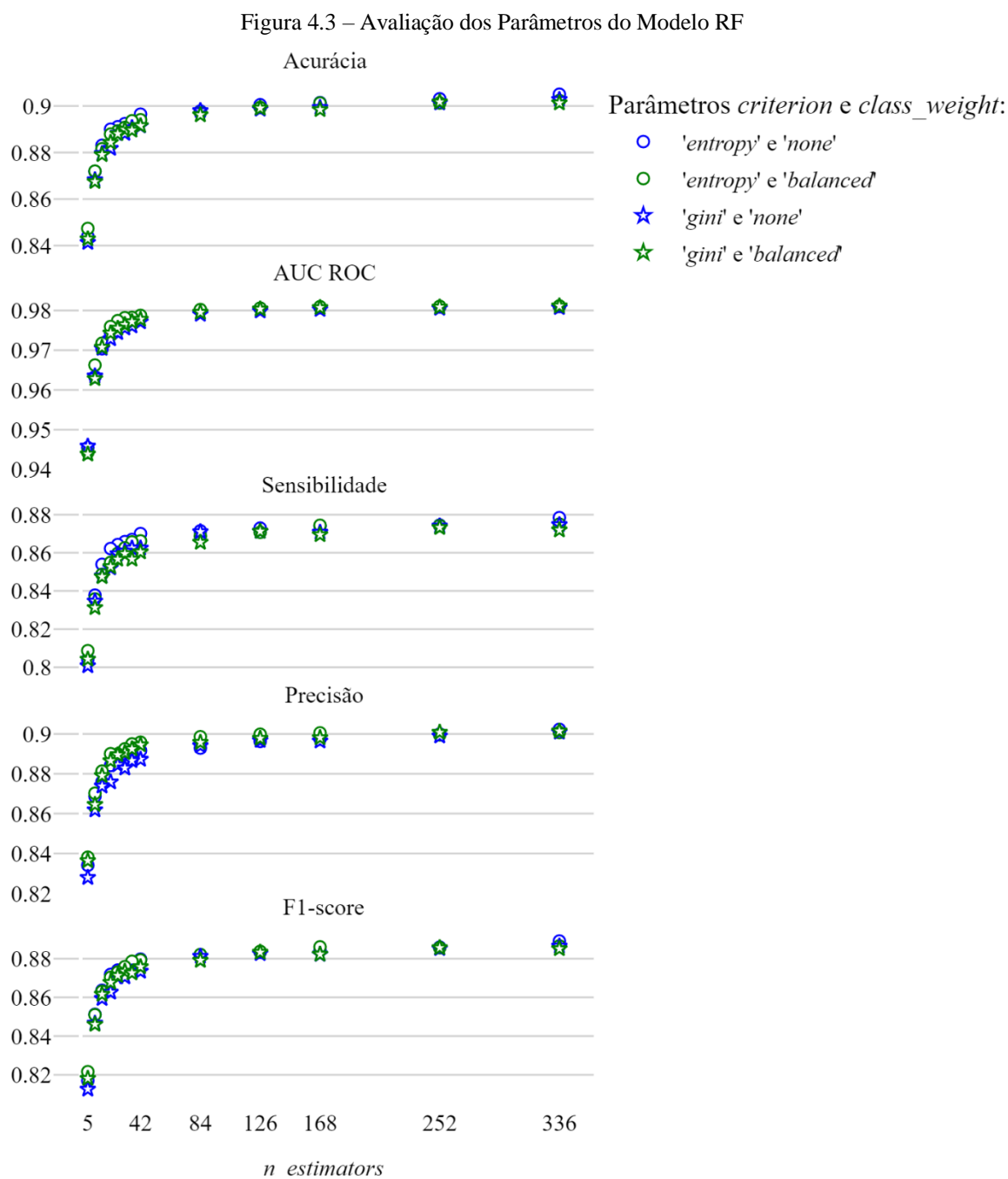
Obs. 2: Configurações que ficaram definidas estão em negrito.

Fonte: A autora (2022)

O melhor desempenho com relação à acurácia geral, conforme apresentado na Tabela 4.4, foi para os parâmetros *criterion* como ‘entropy’, *class_weight* como ‘None’, assim como ocorreu com o DT, e *n_estimators* com o maior valor experimentado, ‘336’ (90,50%±1,54). Por outro lado, com relação a AUC ROC média, foi também para ‘entropy’ e *n_estimators* de ‘336’, mas para *class_weight* como ‘balanced’ (98,11%±0,54). O pior desempenho para ambas as métricas foi para *criterion* como ‘gini’ e para o menor valor experimentado de *n_estimators*,

'5', e, especificamente com relação à acurácia geral, para *class_weight* como 'None' (84,14%±1,59), e à AUC ROC média, para *class_weight* como 'balanced' (94,39%±1,03). Por fim, a combinação escolhida foi 'entropy', 'None' e '126', tendo este último valor sido arbitrariamente diminuído para '52' antes da aplicação das técnicas de seleção de variáveis.

A representação dos resultados para todas as combinações dos parâmetros variados para o modelo RF está apresentada na Figura 4.3.



Fonte: A autora (2022)

O parâmetro que mais impactou o modelo RF, conforme pode ser visualizado na Figura 4.3, foi o parâmetro quantitativo *n_estimators*, visto que à medida do seu aumento, assim como ocorreu com o modelo DT com relação ao parâmetro *max_depth*, as diferentes configurações dos demais parâmetros, *class_weight* como 'None' ou 'balanced' e *criterion* como 'entropy' ou 'gini', foram se tornando mais irrelevantes e os resultados das diferentes métricas de avaliação foram convergindo entre si. Ainda, até que esse parâmetro *n_estimators* atingisse o valor '42', houve uma significativa melhora em todas as métricas de avaliação, na ordem de 5% para acurácia geral, 3% para AUC ROC média e 6% para as demais.

Entretanto, a partir de *n_estimators* de '126', o impacto em todas as métricas não ultrapassou 0,5% e, especificamente para '126', o parâmetro *criterion* como 'entropy' apresentou resultados superiores para as métricas acurácia geral e sensibilidade média quando combinado com o *class_weight* como 'None', e para as métricas precisão e F1-score médios, quando combinado com *class_weight* como 'balanced'. Essa diferença, entretanto, não chegou a 0,5% para essas métricas e, especialmente para a AUC ROC média, não chegou a 0,02%.

As características das árvores de decisão dos modelos DT e RF implementados antes e após a redução das variáveis de entrada, estão apresentadas na Tabela 4.5.

Tabela 4.5 – Comparativo das características Pré e Pós SFS dos modelos DT e RF

Características	DT		RF	
	Pré SFS	Pós SFS	Pré SFS	Pós SFS
max_depth	21	19	18 a 21	19
node_count	1765	1963	1753 a 2095	1867 a 2301
n_leaves	883	982	877 a 1048	934 a 1151

Fonte: A autora (2022)

A profundidade, quantidade de nós (*node_count*) e folhas (*n_leaves*) da árvore de decisão constituinte do modelo DT antes da redução das variáveis de entrada, conforme apresentado na Tabela 4.5, eram de 21, 1765 e 883, respectivamente. Após a redução das variáveis, a árvore do modelo passou a ter profundidade menor (*max_depth*), de 19, e quantidade de nós e folhas maiores, de 1963 e 982, respectivamente.

Já, a profundidade das árvores de decisão do modelo RF antes da redução das variáveis de entrada variava entre 18 e 21, sendo aproximadamente 90% com profundidade 21 e apenas uma com profundidade 18, e após a redução passou a ser de 19 para todas as árvores. Ainda, da mesma forma que para o DT, também a quantidade de nós e folhas aumentou nas árvores do modelo após a redução das variáveis, passando de árvores que tinham entre 1753 e 2095 nós, para árvores que passaram a ter entre 1867 e 2301 nós, e de 877 a 1048, para 934 e 1151 folhas.

O modelo SVM teve os parâmetros *C*, *kernel* e *class_weight* variados. As variações do *C* compreenderam ‘0.01’, ‘0.1’, ‘1.0’, ‘10’, ‘100’, ‘250’, ‘500’, as do *kernel* compreenderam as opções ‘linear’, ‘poly’, ‘rbf’ e ‘sigmoid’, e do *class_weight*, assim como para os modelos DT e RF, as ponderações ‘None’ ou ‘balanced’. Já, o parâmetro *decision_function_shape* foi arbitrariamente fixado como ‘ovr’ (*one versus rest*). As combinações das diferentes configurações foram avaliadas e resultaram em valores dentre os apresentados na Tabela 4.6.

Tabela 4.6 – Resultados do *GridSearchCV* para o modelo SVM

Hiperparâmetros			Métricas de Avaliação (médias macro)					
class_weight	kernel	C	Acurácia	AUC ROC	Precisão	Sensibilidade	F1-score	
<u>Melhores resultados com relação à Acurácia</u>								
1°.	None	rbf	250	85,22±1,52		83,69±2,03	82,76±1,54	83,15±1,72
2°.	balanced	rbf	500	85,16±1,13		83,34±1,51	82,85±1,26	83,04±1,28
3°.	None	rbf	100	85,11±1,29		83,86±1,76	82,46±1,40	83,07±1,47
4°.	None	rbf	500	85,10±1,21		83,53±1,76	82,56±1,33	82,98±1,43
5°.	balanced	rbf	250	85,10±1,38		83,11±1,79	83,09±1,52	83,04±1,57
<u>Piores resultado com relação à Acurácia</u>								
55°.	balanced	sigmoid	100	27,07±1,67		29,66±1,72	30,52±1,97	26,89±1,61
56°.	balanced	poly	0,01	22,98±1,45		51,96±3,28	34,62±1,70	21,87±1,97

Obs. 1: Valores dos resultados das métricas de avaliação em porcentagem.

Obs. 2: Configurações que ficaram definidas estão em negrito.

Fonte: A autora (2022)

O melhor desempenho com relação à acurácia geral, conforme apresentado na Tabela 4.6, foi para os parâmetros *class_weight* como ‘None’, *kernel* como ‘rbf’ e *C* com o segundo maior valor experimentado, de ‘250’ (85,22%±1,52). O pior desempenho com relação à acurácia geral foi para *class_weight* como ‘balanced’, *kernel* como ‘poly’ e *C* com o menor valor experimentado, de ‘0,01’ (22,98%±1,45). Por outro lado, o penúltimo pior desempenho foi para ‘balanced’, ‘sigmoid’ e *C* de ‘100’ (27,07%±1,67). Com relação a AUC ROC média o desempenho não foi mensurado por requerer elevado tempo de processamento (>xh). Por fim, a combinação escolhida foi ‘None’, ‘rbf’ e ‘100’.

A representação dos resultados para todas as combinações dos parâmetros variados para o modelo SVM está apresentada na Figura 4.4.

Figura 4.4 - Avaliação dos Parâmetros do Modelo SVM



Fonte: A autora (2022)

O parâmetro que mais impactou o modelo SVM, conforme pode ser visualizado na Figura 4.4, foi o parâmetro qualitativo *kernel*, tendo a configuração '*rbf*' ou '*poly*', quando combinada à configuração do *C* com valor maior ou igual a '1' e *class_weight* como '*balanced*', gerado os melhores resultados, com superioridade média de acurácia geral de cerca de 42% com relação ao '*sigmoid*' e 27% com relação a '*linear*', e de 38% e 24% também com relação a '*sigmoid*' e '*linear*'.

Já, quando do parâmetro *class_weight* como ‘None’ ao invés de ‘balanced’, esses valores foram respectivamente, em média, 1% e 4% maiores para as métricas acurácia geral e precisão média e 1% e 4% menores para as métricas sensibilidade e F1-score médios. Ainda, com o parâmetro *kernel* ajustado com uma dessas configurações, houve uma melhora média das métricas de avaliação de 13% e 5% à medida que o parâmetro *C* passou de ‘1’ para ‘10’ e de ‘10’ para ‘100’, mas, a partir do ‘100’ o impacto sobre as métricas não chegou a 0,6%.

O modelo ANN, por fim, teve os parâmetros *hidden_layer_sizes*, *solver* e *activation* variados. As variações do *hidden_layer_sizes* compreenderam os valores de ‘42’, ‘84’, ‘168’ e ‘252’, as do *solver* compreenderam as opções ‘lbfgs’, ‘sgd’ e ‘adam’ e as da *activation*, as opções ‘identity’, ‘logistic’, ‘tanh’ e ‘relu’. Já, o parâmetro correspondente à quantidade de camadas ocultas no modelo, *n_layers*, foi arbitrariamente fixado em ‘2’. As combinações das diferentes configurações foram avaliadas e resultaram em valores dentre os apresentados na Tabela 4.7.

Tabela 4.7 – Resultados do *GridSearchCV* para o modelo ANN

Hiperparâmetros				Métricas de Avaliação (médias macro)				
rank	solver	activation	hidden_layer_sizes	Acurácia	AUC ROC	Precisão	Sensibilidade	F1-score
Melhores resultados com relação à Acurácia								
1°.	adam	tanh	252	85,71±1,71	95,78±0,92	84,17±2,08	83,51±2,43	83,76±2,21
2°.	adam	relu	252	85,68±1,29	95,52±0,86	83,93±1,82	83,53±1,79	83,62±1,57
3°.	adam	tanh	168	85,28±1,14	95,22±0,84	83,59±1,71	83,22±1,72	83,36±1,64
4°.	sgd	relu	252	84,81±1,65	95,29±1,04	82,68±2,39	82,21±2,24	82,36±2,23
5°.	lbfgs	tanh	252	84,77±0,87	94,72±0,80	82,53±0,82	82,64±1,48	82,53±1,09
Melhores resultados com relação à AUC ROC								
1°.	adam	tanh	252	85,71±1,71	95,78±0,92	84,17±2,08	83,51±2,43	83,76±2,21
2°.	adam	relu	252	85,68±1,29	95,52±0,86	83,93±1,82	83,53±1,79	83,62±1,57
3°.	sgd	relu	252	84,81±1,65	95,29±1,04	82,68±2,39	82,21±2,24	82,36±2,23
4°.	adam	relu	168	84,74±1,34	95,23±0,73	82,54±1,60	82,84±1,98	82,59±1,70
5°.	adam	tanh	168	85,28±1,14	95,22±0,84	83,59±1,71	83,22±1,72	83,36±1,64
Piores resultados com relação à Acurácia								
47°.	sgd	logistic	42	54,12±1,60	73,62±1,81	53,69±9,50	38,31±1,73	37,99±2,07
48°.	adam	identity	42	54,07±2,32	75,18±1,66	48,33±3,37	41,63±2,63	43,13±2,90
Piores resultados com relação à AUC ROC								
47°.	sgd	logistic	42	54,12±1,60	73,62±1,81	53,69±9,50	38,31±1,73	37,99±2,07
48°.	sgd	logistic	84	54,52±1,51	73,25±1,80	51,64±5,99	38,37±1,67	38,29±2,05

Obs. 1: Valores dos resultados das métricas de avaliação em porcentagem.

Obs. 2: Configurações que ficaram definidas estão em negrito.

Fonte: A autora (2022)

O melhor desempenho com relação à acurácia geral e à AUC ROC média, conforme apresentado na Tabela 4.7, foi para os parâmetros *solver* como ‘adam’, *activation* como ‘tanh’ e *hidden_layer_sizes* com a maior quantidade experimentada de neurônios, ‘252’ (85,71%±1,71 de acurácia e 95,78%±0,92 de AUC ROC). Já, o pior desempenho com relação à acurácia geral foi para *solver* como ‘adam’, *activation* como ‘identity’ e para a menor quantidade de *hidden_layer_sizes* experimentada, ‘42’ (54,07%±2,32). Já, o pior desempenho com relação à AUC ROC média, foi para *solver* como ‘sgd’, *activation* como ‘logistic’ e também para a

menor quantidade de *hidden_layer_sizes*, de '42' (73,25%±1,80). Por fim, a combinação escolhida foi 'adam', 'tanh', e '168', tendo este último sido arbitrariamente diminuído para '65' antes da aplicação das técnicas de seleção de variáveis.

A representação dos resultados para todas as combinações dos parâmetros variados para o modelo ANN está apresentada na Figura 4.5.



Fonte: A autora (2022)

O parâmetro que mais impactou o modelo ANN, conforme pode ser visualizado na Figura 4.5, foi o parâmetro qualitativo *activation*, tendo as configurações 'tanh' e 'relu' gerado

resultados em média 30% superiores do que a ‘*identity*’, independentemente da configuração do *solver*, e do que a ‘*logistic*’, quando do *solver* ‘*sgd*’. Já, para o *solver* configurado como ‘*adam*’ a diferença média foi de 2% e como ‘*lbfgs*’ não houve diferença significativa.

Com relação ao parâmetro *hidden_layer_sizes*, quanto mais alto foi seu valor, melhor foi o resultado com relação a todas as métricas de avaliação, embora essa melhora tenha sido relativamente pequena, na ordem de 2% em média para o aumento de ‘42’ a ‘84’ neurônios e de ‘84’ a ‘168’, respectivamente, e de 1% para o aumento de ‘168’ a ‘252’, para os parâmetros *activation* como ‘*tanh*’ ou ‘*relu*’ e o *solver* como ‘*adam*’.

4.2. Seleção e Importância de Variáveis de Entrada

As técnicas de seleção de variáveis de entrada aplicadas aos modelos de aprendizagem de máquina estão apresentadas na Tabela 4.8.

Tabela 4.8 – Técnicas para Seleção de Variáveis de Entrada aplicadas aos Modelos

Modelo	<i>Feature Import.</i>	<i>Permut. Import.</i>	<i>F Classif.</i>	<i>Mutual Info Classif.</i>	<i>Seq. Feature Selector (SFS)</i>	" <i>RF+SFS</i> "
KNN		x	x	x	x	x
DT	x	x	x	x	x	x
RF	x	x	x	x	x	
SVM		x	x	x	x	x
ANN		x	x	x		x

Obs.: Técnica de Seleção adotada está em negrito

Fonte: A autora (2022)

A técnica de seleção baseada na *Feature Importance* como uma característica inerente dos modelos baseados em árvores de decisão, conforme apresentado na Tabela 4.8, foi aplicada aos DT e RF. Já, a técnica de Permutation Importance foi aplicada a todos os modelos, assim como as *F Classification* e *Mutual Info Classification*. Por fim, a técnica *Sequential Feature Selector* (SFS) foi aplicada a todos os modelos com exceção do ANN, por requerer tempo de processamento computacional superior a xh. Para o ANN, assim como para os demais modelos, foi aplicada também a seleção baseada nas variáveis de entrada que já haviam sido selecionadas pela combinação do modelo e técnica que tinham gerado o melhor desempenho, o RF e a SFS.

O desempenho geral dos modelos antes aplicação das técnicas de seleção de variáveis de entrada está apresentado na Tabela 4.9.

Tabela 4.9 - Relatórios de Avaliação dos modelos Pré redução das variáveis de entrada

	RF	KNN	DT	ANN	SVM
Acurácia					
Geral	90,05%	89,30%	82,80%	85,28%	85,11%
AUC ROC					
Líder	97,78%	95,65%	87,84%	94,66%	94,30%
Vice-Líder	97,82%	95,73%	87,98%	94,77%	94,35%
Primeiro-Preço	98,61%	95,32%	86,67%	96,39%	96,11%
Premium	98,00%	95,25%	85,48%	95,05%	94,28%
Média macro	98,05%	95,49%	86,99%	95,22%	94,76%
Precisão					
Líder	90,55%	92,22%	87,31%	88,54%	87,08%
Vice-Líder	90,28%	88,27%	82,31%	83,25%	84,48%
Primeiro-Preço	88,12%	84,12%	76,57%	82,24%	83,03%
Premium	89,24%	85,48%	72,57%	79,95%	80,33%
Média macro	89,55%	87,52%	79,69%	83,50%	83,73%
Sensibilidade					
Líder	94,60%	91,41%	87,04%	88,78%	89,88%
Vice-Líder	88,17%	89,04%	82,31%	83,61%	82,42%
Primeiro-Preço	83,96%	86,32%	75,62%	81,22%	80,35%
Premium	82,45%	84,54%	74,36%	79,26%	77,18%
Média macro	87,29%	87,83%	79,83%	83,22%	82,46%
F1-score					
Líder	92,53%	91,81%	87,17%	88,66%	88,46%
Vice-Líder	89,21%	88,65%	82,31%	83,43%	83,44%
Primeiro-Preço	85,99%	85,21%	76,10%	81,73%	81,67%
Premium	85,71%	85,01%	73,45%	79,61%	78,72%
Média macro	88,36%	87,67%	79,76%	83,36%	83,07%

Fonte: A autora (2022)

A acurácia geral antes da redução das variáveis de entrada, conforme apresentado na Tabela 4.9 foi de cerca de 90% e 89% para os modelo RF e KNN respectivamente, enquanto para os modelos ANN, SVM e DT foi menor, de cerca de 85%, 85% e 83%, nessa ordem.

A área sob as curvas ROC do modelo RF foi na média de 98%, enquanto as dos modelos KNN, ANN e SVM foram inferiores, cerca de 95%, e a do DT foi a menor, de 87%. Para todos os modelos, a diferença de área sob a curvas ROC que alguma classe específica da função estratégica apresentou com relação à média não passou de 1%, exceto com relação à classe premium para o caso do DT, cujo desempenho ficou 1,5% abaixo da média do modelo.

As precisões do modelo RF foram de aproximadamente 91%, 90%, 89% e 88%, para as classes líder, vice-líder, premium e primeiro-preço, respectivamente. A precisão média foi de cerca de 90% para esse modelo RF, seguida do modelo KNN, de 88%, e posteriormente dos modelos SVM, ANN e DT, de cerca de 84%, 84% e 80%, respectivamente. Para todos os modelos, a classe que teve desempenho superior ao da média foi a classe líder, mas com menor diferença para o modelo RF, de 1%.

As sensibilidades do modelo RF foram de aproximadamente 95% e 88% para as classes líder e vice-líder, respectivamente, e de 84% e 82% para as classes primeiro-preço e premium.

A sensibilidade média foi de cerca de 87% para esse modelo RF, superada pelo modelo KNN, com cerca de 88%, e seguida dos modelos ANN, SVM e DT, com cerca de 83%, 82% e 80%, respectivamente. Para todos os modelos, assim como ocorreu com a precisão, a classe que teve desempenho superior ao da média foi a líder, mas com menor diferença para o KNN, de 3,6%.

Os F1-scores do modelo RF foram de aproximadamente 93% e 89% para as classes líder e vice-líder, respectivamente, e de 86% para as classes primeiro-preço e premium. O F1-score médio foi de cerca de 88% para esse modelo RF, seguido dos KNN, ANN, SVM e DT, com cerca de 88%, 83%, 83% e 80%, respectivamente. Para todos os modelos, assim como ocorreu com a precisão e sensibilidade, também a classe que teve desempenho superior ao da média foi a líder, com maior diferença para o modelo DT, de 7,4%.

As importâncias das variáveis de entrada segundo a *Feature Importance*, para o DT e RF, estão apresentadas na Tabela 4.10.

Tabela 4.10 – *Feature Importance* dos modelos DT e RF

<i>Feature Importance para o DT</i>			<i>Feature Importance para o RF</i>		
	%	% Acum.		%	% Acum.
G3_Fabricante	7,2%	7,2%	G3_Fabricante	5,3%	5,3%
Similares_3x'	6,4%	13,7%	G3_Marca	5,1%	10,4%
Similares_4	5,9%	19,5%	G4_Marca	4,6%	15,0%
G3_Marca	4,9%	24,4%	Similares_3	4,4%	19,4%
Similares_5	4,4%	28,8%	Custo	4,3%	23,6%
Custo	4,2%	33,0%	G4_Fabricante	4,2%	27,9%
Fabricante_Marca	4,1%	37,1%	Similares_4	4,1%	31,9%
G4_Marca	4,0%	41,2%	Preço	4,1%	36,0%
Peso	3,8%	45,0%	G5_Marca	4,0%	40,0%
G5_Marca	3,7%	48,6%	Similares_5	3,9%	43,9%
G2_Fabricante	3,6%	52,3%	Lucro	3,7%	47,6%
G4_Fabricante	3,6%	55,8%	Peso	3,7%	51,3%
Preço	3,5%	59,3%	G2_Fabricante	3,3%	54,6%
Lucro_%	3,1%	62,4%	G5_Fabricante	3,3%	57,8%
G2_Marca	3,0%	65,4%	Lucro_%	3,2%	61,0%
Tam. Emb.	2,8%	68,2%	Fabricante_Marca	3,0%	64,0%
Lastro	2,6%	70,8%	Tam. Emb.	2,8%	66,8%
Altura	2,6%	73,5%	Lastro	2,8%	69,6%
Validade	2,5%	76,0%	Altura	2,6%	72,2%
Lucro	2,5%	78,5%	G2_Marca	2,5%	74,7%
Reposição	2,4%	80,9%	Validade	2,5%	77,2%
G5_Fabricante	2,1%	83,0%	C3_Fabricante	2,3%	79,5%
G1_Marca	1,7%	84,7%	Reposição	2,1%	81,7%
Camadas	1,7%	86,4%	Camadas	2,0%	83,7%
C3_Marca	1,6%	88,1%	C4_Fabricante	1,9%	85,5%
C3_Fabricante	1,6%	89,7%	G1_Fabricante	1,8%	87,3%
C5_Fabricante	1,5%	91,2%	C3_Marca	1,7%	89,0%
C4_Marca	1,5%	92,7%	C4_Marca	1,7%	90,7%
Entrega	1,1%	93,7%	C5_Fabricante	1,4%	92,1%
C4_Fabricante	1,0%	94,7%	G1_Marca	1,4%	93,5%
C5_Marca	0,9%	95,7%	C2_Fabricante	1,3%	94,9%
C2_Fabricante	0,9%	96,6%	C5_Marca	1,0%	95,9%
G1_Fabricante	0,9%	97,4%	Entrega	1,0%	96,8%
C2_Marca	0,8%	98,2%	Fabricante_Prop_Geral	0,9%	97,8%
Marca_Prop_Geral	0,6%	98,8%	C2_Marca	0,9%	98,6%

C1_Fabricante	0,5%	99,3%	C1_Fabricante	0,6%	99,2%
Fabricante_Prop_Geral	0,5%	99,8%	Marca_Prop_Geral	0,6%	99,8%
C1_Marca	0,2%	100,0%	C1_Marca	0,2%	100,0%

Fonte: A autora (2022)

Essas importâncias via *feature importance*, conforme apresentado na Tabela 4.10, indicaram, para ambos os modelos DT e RF, uma maior relevância da participação do fabricante em termos de itens no grupo 3, assim como da marca nos grupos 3 e 4 e da quantidade de similares nesses mesmos grupos 3 e 4. Especificamente para o modelo DT, também se sobressaíram a quantidade de similares no grupo 5 e a quantidade de marcas que o fabricante possui, além do custo do item. Para o modelo RF, se destacaram também a participação do fabricante em termos de itens no grupo 4, além do custo e preço do item. Essas oito variáveis representaram, respectivamente, uma importância cumulativa de 41% e 36% para os modelos DT e RF, respectivamente.

As variáveis com relevância intermediária, entre 0,04 e 0,025, foram o lucro unitário, percentual, validade, altura, tamanho da embalagem, do lastro e participação do fabricante e marca em termos de itens no grupo 2 e da marca no grupo 5, além de, para o caso do DT, das variáveis de participação do fabricante em termos de itens no grupo 4, peso, preço e tempo de reposição do item, bem como, para o RF, da participação do fabricante em termos de itens no grupo 5, da quantidade de similares nesse mesmo grupo e da quantidade de marcas que o fabricante possui.

As variáveis com menor relevância, inferiores a 0,025, por outro lado, foram, para ambos os modelos, as correspondentes à participação da marca e fabricante em termos de categorias em quaisquer grupos, em termos de itens no geral ou no grupo 1 e do tempo de entrega dos itens e da quantidade de camadas quando dispostos em palete, além de, para o caso do DT, da participação do fabricante em termos de itens no grupo 5. Tais variáveis, 17, acumularam importância relativa de cerca de 20%.

As importâncias das variáveis de entrada segundo a *Permutation Importance*, para todos os modelos, estão apresentadas na Tabela 4.11.

Tabela 4.11 – *Permutation Importance* dos modelos

KNN		DT		RF		SVM		ANN	
Entrega	0,9	G3_Fab	35,4	G3_Fab	0,9	C5_Fab	12,2	Similares_3	12,5
Validade	0,8	G5_Marca	16,5	G5_Marca	0,8	C4_Marca	11,6	G1_Fab	11,2
Reposição	0,8	G3_Marca	15,9	G3_Marca	0,6	C3_Fab	9,9	C4_Marca	10,5
Similares_4	0,7	Similares_4	15,7	G4_Marca	0,5	Similares_3	9,7	Fab_Marca	10,3
Similares_3	0,6	G2_Fab	13,7	G4_Fab	0,4	Marca_Prop_Geral	9,6	G2_Fab	10,0
Similares_5	0,5	Fab_Marca	13,6	Similares_4	0,2	G2_Fab	9,1	C1_Fab	10,0
Marca_Prop_Geral	0,5	Similares_3	12,9	G2_Fab	0,2	C5_Marca	8,9	Reposição	9,6
G5_Marca	0,5	G4_Marca	11,7	Similares_3	0,2	Fab_Marca	8,8	Marca_Prop_Geral	9,1

Altura	0,4	G4_Fab	10,6	Similares_5	0,2	G3_Fab	8,7	Validade	8,5
Lastro	0,4	Similares_5	10,4	Preço	0,2	C2_Fab	8,4	G1_Marca	8,2
Tam. Emb.	0,4	Preço	9,4	Custo	0,2	G4_Marca	8,3	G3_Marca	7,8
Fab_Marca	0,4	Custo	7,7	Fab_Marca	0,1	Similares_4	8,1	C2_Marca	7,7
Fab_Prop_Geral	0,4	G2_Marca	7,3	G2_Marca	0,1	Similares_5	8,1	C3_Fab	7,6
G2_Fab	0,4	G1_Marca	6,4	G1_Marca	0,1	G5_Marca	8,0	Similares_4	7,5
G3_Marca	0,4	C5_Fab	6,3	Peso	0,1	G2_Marca	8,0	C5_Marca	7,3
G4_Marca	0,4	Peso	6,3	Validade	0,1	G4_Fab	7,9	G3_Fab	7,3
G4_Fab	0,4	Validade	4,8	G5_Fab	0,1	G1_Fab	7,9	G4_Marca	7,1
G5_Fab	0,4	G5_Fab	4,3	C3_Fab	0,1	G3_Marca	7,7	C3_Marca	6,2
C1_Marca	0,4	C3_Fab	4,3	Reposição	0,1	Validade	7,7	Entrega	6,2
C2_Marca	0,4	Reposição	4,0	Tam. Emb.	0,1	C1_Fab	7,5	C5_Fab	6,1
C4_Marca	0,4	Tam. Emb.	3,9	Lucro	0,1	Reposição	7,4	C2_Fab	6,1
C4_Fab	0,4	Lucro	3,8	Lucro_%	0,1	C4_Fab	7,1	G5_Fab	6,0
C5_Fab	0,4	Lucro_%	3,5	Altura	0,1	C2_Marca	7,1	Similares_5	5,9
Preço	0,3	C3_Marca	3,5	Lastro	0,1	C3_Marca	7,0	G5_Marca	5,9
Custo	0,3	Altura	3,4	Entrega	0,1	G1_Marca	6,9	G2_Marca	5,8
Lucro	0,3	Lastro	3,3	G1_Fab	0,1	G5_Fab	6,9	Custo	5,6
Lucro_%	0,3	C4_Marca	3,1	C5_Fab	0,0	Fab_Prop_Geral	5,9	G4_Fab	5,3
Peso	0,3	Entrega	2,7	C3_Marca	0,0	C1_Marca	4,6	Fab_Prop_Geral	5,1
G1_Marca	0,3	C2_Fab	2,2	C4_Marca	0,0	Custo	4,3	Peso	4,9
G1_Fab	0,3	Camadas	2,2	C2_Fab	0,0	Entrega	4,3	Altura	4,6
G2_Marca	0,3	G1_Fab	2,1	Camadas	0,0	Lucro	4,1	C1_Marca	4,4
G3_Fab	0,3	C4_Fab	2,0	C4_Fab	0,0	Peso	3,7	Lucro_%	4,3
C1_Fab	0,3	C5_Marca	1,8	C5_Marca	0,0	Tam. Emb.	2,8	Tam. Emb.	4,0
C2_Fab	0,3	Marca_Prop_Geral	1,3	Marca_Prop_Geral	0,0	Preço	2,7	C4_Fab	3,5
C3_Marca	0,3	Fab_Prop_Geral	1,1	Fab_Prop_Geral	0,0	Altura	2,7	Lucro	3,4
C3_Fab	0,3	C2_Marca	1,0	C2_Marca	0,0	Lucro_%	2,1	Preço	2,8
C5_Marca	0,3	C1_Fab	1,0	C1_Fab	0,0	Lastro	1,5	Lastro	1,8
Camadas	0,1	C1_Marca	0,2	C1_Marca	0,0	Camadas	0,5	Camadas	0,8

Obs.: Valores multiplicados por 100

Fonte: A autora (2022)

Essas importâncias via *permutation importance*, conforme apresentado na Tabela 4.11, indicaram que não houve uma convergência entre os modelos quanto a essa importância. Para o KNN e RF, entretanto, as variáveis de entrada que se destacaram foram as relativas às participações do fabricante no grupo 4, e da marca nos grupos 4 e 5, especialmente para o RF, além da quantidade de itens similares nos grupos 3, 4 e 5, especialmente para o KNN.

Em contrapartida, as variáveis que menos se destacaram para ambos os modelos RF e KNN, foram as participações do fabricante e da marca em termos da quantidade de categorias em todos os grupos, seguidas das variáveis peso e lucro, tanto unitário quanto percentual, tamanho do lastro, quantidade de camadas e tamanho da embalagem, e, posteriormente, da quantidade de marcas que o fabricante possui, de sua participação em termos da quantidade de itens no geral e no grupo 1 bem como da marca nos grupos 1 e 2.

Por fim, as variáveis altura, prazo de validade, tempo de reposição e de entrega, tiveram resultados relevantes para o modelo KNN, mas não para o modelo RF, e a participação em termos de itens do fabricante no grupo 4 e da marca nos grupos 4 e 5 tiveram resultados relevantes para o modelo RF, mas não para o KNN.

O desempenho de cada modelo versus cada técnica de seleção de variáveis que foi aplicada para redução da base a nove variáveis de entrada está apresentado na Tabela 4.12.

Tabela 4.12 – Comparativo de Desempenho das Técnicas de Seleção a 9 variáveis

	<i>Feature Import.</i>	<i>Permut. Import.</i>	<i>F Classif.</i>	<i>Mutual Info Classif.</i>	<i>Seq. Feature Selector (SFS)</i>	<i>"RF+SFS"</i>
KNN						
Acurácia	NA	77,70%	83,72%	83,75%	88,56%	88,80%
AUC ROC	NA	88,37%	91,85%	92,46%	94,55%	94,80%
Precisão	NA	74,86%	80,21%	80,82%	86,48%	87,26%
Sensibilidade	NA	73,40%	80,44%	80,58%	86,38%	86,95%
F1-score	NA	74,01%	80,26%	80,64%	86,37%	87,06%
DT						
Acurácia	83,58%	85,13%	80,31%	80,34%	87,83%	85,10%
AUC ROC	88,07%	90,66%	85,94%	85,37%	92,90%	88,89%
Precisão	81,01%	82,17%	76,08%	76,62%	85,94%	82,67%
Sensibilidade	80,86%	81,35%	76,47%	76,64%	84,75%	82,59%
F1-score	80,81%	81,68%	76,21%	76,53%	85,28%	82,56%
RF						
Acurácia	86,58%	85,67%	85,64%	85,32%	90,09%	NA
AUC ROC	96,90%	96,65%	96,37%	96,53%	97,84%	NA
Precisão	85,72%	84,34%	83,73%	84,34%	89,28%	NA
Sensibilidade	83,38%	81,80%	81,58%	81,32%	87,71%	NA
F1-score	84,40%	82,91%	82,54%	82,63%	88,42%	NA
SVM						
Acurácia	NA	71,66%	65,68%	64,49%	77,35%	76,15%
AUC ROC	NA	NA	NA	NA	NA	NA
Precisão	NA	73,29%	61,01%	63,44%	75,16%	75,61%
Sensibilidade	NA	61,11%	53,64%	50,35%	71,71%	69,60%
F1-score	NA	64,96%	55,49%	52,39%	73,11%	72,02%
ANN						
Acurácia	NA	79,72%	76,99%	76,20%	NA	84,60%
AUC ROC	NA	92,42%	90,83%	89,40%	NA	94,75%
Precisão	NA	77,52%	72,18%	72,27%	NA	82,40%
Sensibilidade	NA	74,40%	71,19%	71,47%	NA	81,63%
F1-score	NA	75,72%	71,61%	71,72%	NA	81,92%

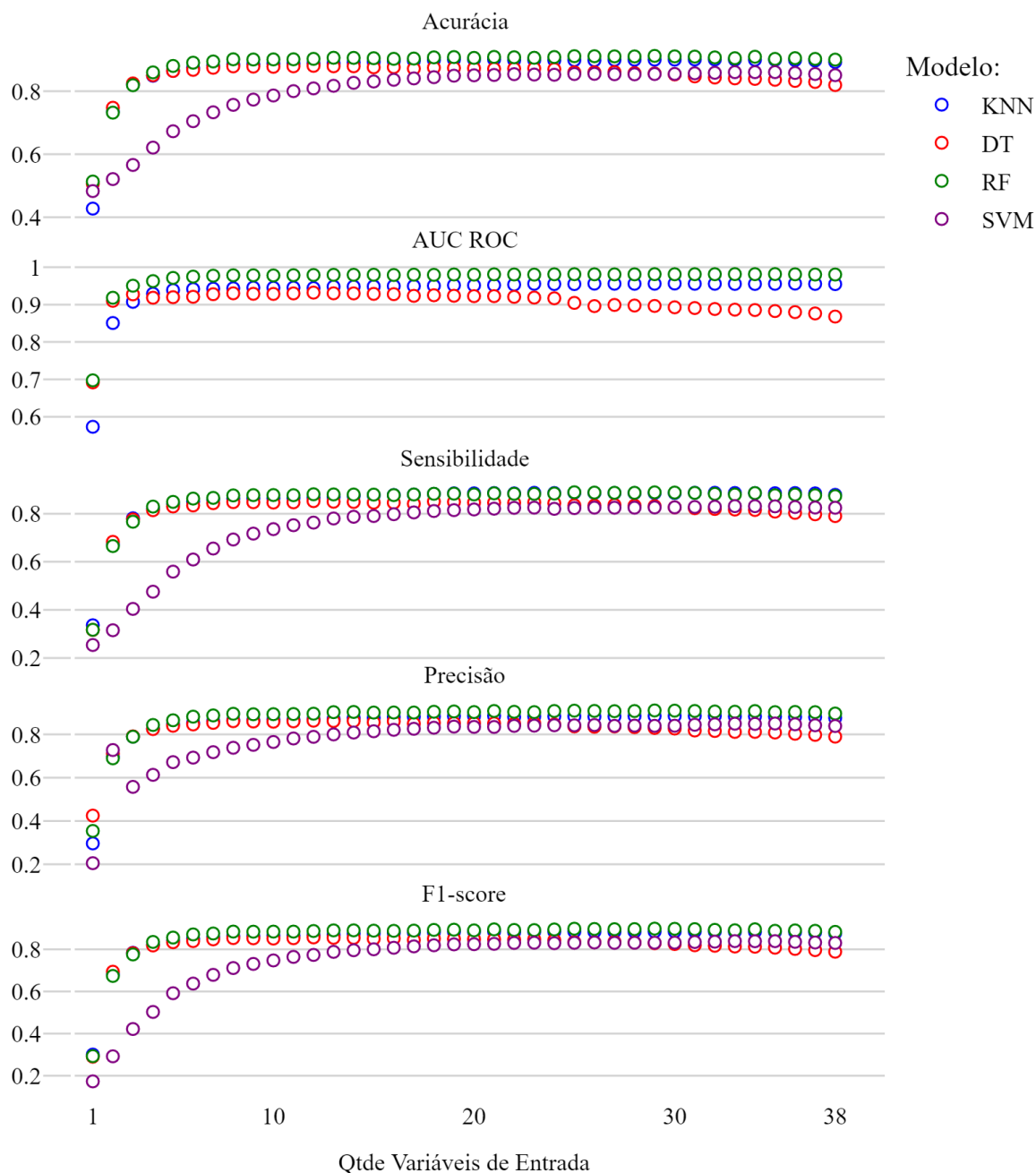
Fonte: A autora (2022)

A técnica de seleção que gerou os melhores resultados, conforme apresentado na Tabela 4.12, foi a SFS para os modelos DT, RF e SVM, e a que consistiu na aplicação do resultado das variáveis selecionadas para o modelo RF via técnica SFS, mas aplicada ao modelo ANN e também para KNN, tendo, nesse último caso, essa “técnica” apresentado superioridade média de 0,5% quando comparada à SFS.

Ainda, as técnicas *Feature Importance* e *Permutation Importance* se sobressaíram com relação a *F Classification* e *Mutual Info Classification* para o modelo DT, especialmente a *Permutation Importance* para a AUC ROC média, e também para o SVM e ANN no caso da *Permutation Importance*, enquanto que a *F Classification* e *Mutual Info Classification* se destacaram para o KNN. Para o RF, os resultados dessas quatro técnicas tiveram pouca diferença média, com máxima de 1,3% entre a técnica *Feature Importance* e *F Classification*.

O desempenho dos modelos para as diferentes quantidades de variáveis possíveis, 1 a 38, selecionadas via técnica SFS está apresentada na Figura 4.6.

Figura 4.6 – Desempenho da técnica SFS por modelo e quantidade de variáveis selecionadas



Fonte: A autora (2022)

Para a métrica acurácia geral, conforme apresentado na Figura 4.6, os modelos KNN e RF geraram resultados muito parecidos para as quantidades a partir de duas de variáveis de entrada, com superioridade máxima de 2% do RF, tendo ambos atingido um platô a partir de 8 variáveis. O modelo DT, por sua vez, começou a gerar resultados decrescentes a partir de 24 variáveis, e o modelo SVM só atingiu patamares próximos aos dos demais a partir de 16

variáveis, embora ainda mais baixos, pelo menos até a quantidade de 28, a partir de quando alcançou e ultrapassou, em no máximo 3%, o DT.

Já, para a métrica AUC ROC média, os modelos KNN e RF tiveram tendências de crescimento semelhantes à medida do aumento da quantidade de variáveis de entrada, tendo o RF gerado resultados em média 3% superiores a partir da quantidade de 3 variáveis de entrada. Da mesma forma que para a acurácia geral, o modelo DT começou a gerar resultados decrescentes a partir de 24 variáveis, e inferiores aos dos demais modelos.

Por fim, para as métricas precisão, sensibilidade e F1-score médios, o comportamento dos modelos com relação às métricas anteriores se repetiu, tendo os resultados dos modelos RF e KNN ficado ainda um pouco mais próximos para a sensibilidade, com nenhuma diferença na média, e diferença máxima de 1% com relação à precisão e F1-scores para o RF.

As variáveis selecionadas para o modelo RF pelas diferentes técnicas de seleção preconfiguradas para reduzir a base a nove variáveis estão apresentadas na Tabela 4.13.

Tabela 4.13 – Variáveis selecionadas pelas diferentes técnicas para o modelo RF

Variáveis de Entrada	<i>Seq. Feature Selector (SFS)</i>	<i>Feature Import.</i>	<i>Permut. Import.</i>	<i>F Classif.</i>	<i>Mutual Info Classif.</i>
Preço	x	x	x		x
Custo		x	x		x
Lucro				x	x
Validade	x				
Similares_3	x	x			x
Similares_4		x	x		
Similares_5			x		
Fabricante_Marca	x				
G1_Marca	x			x	
G2_Marca				x	
G2_Fabricante	x				
G3_Marca	x	x	x	x	x
G3_Fabricante		x	x	x	x
G4_Marca	x	x	x	x	x
G4_Fabricante		x	x	x	x
G5_Marca		x	x	x	x
G5_Fabricante				x	
C3_Marca	x				

Fonte: A autora (2022)

Das nove variáveis de entrada que foram selecionadas pela técnica SFS, conforme apresentado na Tabela 4.13, quatro foram selecionadas somente por essa técnica, sendo essas o prazo de validade do item, quantidade de marcas que o fabricante possui, participação deste em termos de itens no grupo 2 e da marca em termos de categorias no grupo 3. Apenas duas foram selecionadas também por todas as demais técnicas, participação da marca em termos de itens nos grupos 3 e 4, e outras três foram selecionadas por somente algumas das técnicas.

As variáveis de entrada selecionadas pelas técnicas *F Classification* e *Mutual Info Classification*, independentemente do modelo, foram as participações do fabricante e da marca em termos de itens nos grupos 3, 4 e 5, além do lucro do item. Especificamente pela técnica *F Classification*, também o preço e custo do item, além da quantidade de similares no grupo 3. Já, pela técnica *Mutual Info Classification*, também foram selecionadas a participação do fabricante em termos de itens no grupo 5 e da marca nos grupos 1 e 2.

A combinação do modelo RF e da técnica de seleção SFS reduziu as variáveis de entrada às apresentadas na Tabela 4.14.

Tabela 4.14 – Variáveis selecionadas para o modelo RF pela técnica de seleção SFS

Dimensão	Variáveis de Entrada	Descrição
C. Intrínsecas	I. Preço	Preço Unitário (R\$)
	II. Validade	Prazo de Validade (dias)
C. Extrínsecas	III. Similares_3	Itens similares no grupo correspondente (quantidade)
	IV. Fabricante_Marca	Marcas que o Fabricante possui (quantidade)
Participações da Marca/ Fabricante	V. G1_Marca	Participação da Marca/Fabricante em termos da quantidade de itens ou categorias no grupo correspondente
	VI. G2_Fabricante	
	VII. G3_Marca	
	VIII. G4_Marca	
	IX. C3_Marca	

Fonte: A autora (2022)

Dessas nove variáveis de entrada selecionadas, conforme apresentado na Tabela 4.14, duas dizem respeito a características intrínsecas aos itens, o preço e a validade, uma a características extrínsecas, quantidade de itens similares no grupo 3, e, por fim, seis dizem respeito a participações da marca e/ou fabricante em termos da quantidade de marcas que o fabricante possui e da quantidade de itens ou categorias nos diferentes grupos, sendo em termos de itens nos grupos 1, 3 e 4 e de categorias no grupo 3 com relação à marca e, finalmente, com relação ao fabricante, em termos de itens no grupo 2.

4.3. Avaliação dos Modelos Implementados

O desempenho geral dos modelos após a redução das variáveis de entrada em termos das diferentes métricas de avaliação adotadas está apresentado na Tabela 4.15.

Tabela 4.15 – Relatórios de Avaliação dos modelos Pós redução das variáveis

	RF	KNN	DT	ANN	SVM
	Acurácia				
Geral	90,05%	88,56%	87,83%	84,60%	77,35%
	AUC ROC				
Líder	97,58%	94,98%	93,65%	94,15%	90,37%
Vice-Líder	97,50%	95,46%	93,42%	94,17%	90,35%
Primeiro-Preço	98,01%	94,50%	92,43%	96,03%	92,80%

Premium	97,31%	93,25%	92,09%	94,63%	91,13%
Média macro	97,60%	94,55%	92,90%	94,75%	91,16%
Precisão					
Líder	90,91%	91,34%	90,27%	87,73%	80,23%
Vice-Líder	90,03%	89,09%	87,97%	84,23%	76,43%
Primeiro-Preço	88,96%	82,82%	84,65%	78,47%	71,45%
Premium	87,48%	82,14%	80,60%	78,76%	71,72%
Média macro	89,35%	86,35%	85,87%	82,30%	74,96%
Sensibilidade					
Líder	93,75%	91,32%	92,41%	89,13%	86,19%
Vice-Líder	89,20%	89,04%	86,87%	83,45%	74,61%
Primeiro-Preço	84,20%	83,33%	80,22%	78,86%	65,05%
Premium	83,19%	81,84%	79,51%	75,09%	60,98%
Média macro	87,59%	86,38%	84,75%	81,63%	71,71%
F1-score					
Líder	92,31%	91,33%	91,32%	88,42%	83,10%
Vice-Líder	89,62%	89,06%	87,41%	83,84%	75,51%
Primeiro-Preço	86,52%	83,08%	82,38%	78,66%	68,10%
Premium	85,28%	81,99%	80,05%	76,88%	65,92%
Média macro	88,43%	86,36%	85,29%	81,95%	73,16%

Fonte: A autora (2022)

A acurácia geral após a redução das variáveis de entrada, conforme apresentado na Tabela 4.15, do modelo RF foi de cerca de 90%, enquanto a dos modelos KNN e DT foram de, respectivamente, 89% e 88%, e as dos ANN e SVM, de cerca de 85% e 77%, nessa ordem.

As áreas sob as curvas ROC do modelo RF, foram de cerca de 98% para todas as classes, exceto para a premium, que foi de 97%. As áreas médias com desempenhos subsequentes foram as dos modelos KNN e ANN, com cerca de 95%, e dos modelos DT e SVM, com valores de aproximadamente 93% e 91%, respectivamente. Para todos os modelos, os menores valores foram para a classe premium, apesar de com pouca diferença, cerca de 0,3%, para a classe primeiro-preço, no caso do DT.

As precisões do modelo RF foram de aproximadamente 91% e 90% para as classes líder e vice-líder, respectivamente, e de 89% e 87% para as classes primeiro-preço e premium. A precisão média foi de cerca de 89% para esse modelo RF, e as subsequentes foram as dos modelos KNN e DT, com cerca de 86%, e depois as dos ANN e SVM, com cerca de 82% e 75%, respectivamente. Os modelos DT e KNN, assim como o RF, tiveram pior desempenho para a classes premium, enquanto os ANN e SVM, para a classe primeiro-preço.

As sensibilidades do modelo RF foram de aproximadamente 94% e 89% para as classes líder e vice-líder, respectivamente, e de 84% e 83% para as classes primeiro-preço e premium. A sensibilidade média foi de cerca de 88% para esse modelo RF, e as subsequentes foram as dos demais modelos, KNN, DT, ANN e SVM, nessa ordem, com valores de 86%, 85%, 82% e 72%. Para todos os modelos, assim como ocorreu com as áreas sob a curva ROC, o pior desempenho

foi com relação à classe premium, apesar de, para o modelo DT, também com pouca diferença com relação à primeiro-preço, cerca de 0,7%.

Os F1-scores do modelo RF foram de aproximadamente 92% e 90% para as classes líder e vice-líder, respectivamente, e de 87% e 85% para as classes primeiro-preço e premium. O F1-score médio foi de cerca de 88% para esse modelo RF, e os subsequentes foram os do KNN, DT, ANN e SVM, respectivamente, com valores de 86%, 85%, 82% e 73%. Para todos os modelos, assim como ocorreu com as áreas sob as curvas ROC e com a sensibilidade, os piores resultados foram com relação à classe premium.

No geral, os modelos RF, DT e SVM apresentaram precisão maior do que sensibilidade para as classes premium, primeiro-preço e vice-líder, e menor para a classe líder, enquanto o modelo ANN apresentou maior precisão para as classes premium e vice-líder, e menor para líder e primeiro-preço, e o KNN quase não apresentou diferença, com discrepância na ordem de 0,03%.

O impacto geral que a redução das variáveis de entrada teve sobre o desempenho dos modelos implementados está apresentado na Tabela 4.16.

Tabela 4.16 – Relatório da Avaliação Relativa Pós versus Pré redução das variáveis

	RF	KNN	DT	ANN	SVM
Acurácia					
Geral	0,00%	-0,82%	+6,07%	-0,79%	-9,12%
AUC ROC					
Líder	-0,20%	-0,70%	+6,61%	-0,53%	-4,17%
Vice-Líder	-0,71%	-2,09%	+7,73%	-0,45%	-3,33%
Primeiro-Preço	-0,61%	-0,86%	+6,64%	-0,37%	-3,44%
Premium	-0,33%	-0,28%	+6,18%	-0,64%	-4,24%
Média macro	-0,46%	-0,98%	+6,79%	-0,50%	-3,79%
Precisão					
Líder	+0,40%	-0,95%	+3,39%	-0,91%	-7,87%
Vice-Líder	-1,97%	-3,91%	+11,05%	-1,48%	-10,72%
Primeiro-Preço	+0,96%	-1,55%	+10,54%	-4,59%	-13,95%
Premium	-0,27%	+0,92%	+6,87%	+1,17%	-9,53%
Média macro	-0,22%	-1,34%	+7,75%	-1,43%	-10,48%
Sensibilidade					
Líder	-0,89%	-0,10%	+6,17%	+0,39%	-4,10%
Vice-Líder	+0,89%	-3,19%	+6,93%	-5,26%	-20,99%
Primeiro-Preço	+0,30%	-3,46%	+6,09%	-2,91%	-19,04%
Premium	+1,17%	0,00%	+5,54%	-0,19%	-9,48%
Média macro	+0,34%	-1,64%	+6,17%	-1,91%	-13,04%
F1-score					
Líder	-0,24%	-0,52%	+4,76%	-0,27%	-6,05%
Vice-Líder	-0,50%	-3,55%	+8,98%	-3,42%	-16,27%
Primeiro-Preço	+0,62%	-2,50%	+8,25%	-3,75%	-16,62%
Premium	+0,45%	+0,46%	+6,20%	+0,49%	-9,50%
Média macro	+0,08%	-1,49%	+6,94%	-1,69%	-11,94%

Fonte: A autora (2022)

O modelo RF, conforme apresentado na Tabela 4.16, não apresentou diferença com relação à acurácia geral, mas apresentou uma pequena piora com relação à AUC ROC e precisão médias, de 0,46% e 0,22%, respectivamente, e uma pequena melhora com relação à sensibilidade e do F1-score médios, de 0,34% e 0,08%, respectivamente. Esse impacto sobre as médias das métricas de avaliação não ocorreu, entretanto, especificamente para a precisão no caso das funções estratégicas líder e primeiro-preço, que tiveram melhora na ordem de 0,40% e 0,96%, respectivamente, para a sensibilidade no caso da função estratégica líder, que teve piora na ordem de 0,89%, e para o F1-score no caso das funções estratégicas líder e premium, que tiveram piora na de 0,24% e 0,50%, respectivamente.

Já, o modelo KNN apresentou uma pequena piora com relação à média dos resultados das métricas de avaliação, com diminuição de 0,82% para a acurácia geral, 0,98% para a AUC ROC média, 1,34% para a precisão média, 1,64% para a sensibilidade média e 1,49% para o F1-score médio. Essa piora não ocorreu, entretanto, especificamente para a função estratégica vice-líder, tanto no caso da precisão, que teve melhora na ordem de 0,92%, da sensibilidade, que não apresentou diferença, quanto do F1-score, que teve melhora na ordem de 0,46%.

O modelo DT, por sua vez, apresentou melhora decorrente da redução das variáveis de entrada para todas as métricas de avaliação e para todas as funções estratégicas. Essa melhora foi, na média, de 6,07% para a acurácia geral, 6,79% para a AUC ROC média, 7,75% para a precisão média, 6,17% para a sensibilidade média e 6,94% para o F1-score médio. Como destaque dessas melhoras foi a precisão das funções estratégicas primeiro-preço e premium, com melhora de 10,54% e 11,05%, respectivamente.

O modelo ANN apresentou uma pequena piora com relação à média de todas as métricas de avaliação, na ordem de 0,79% para a acurácia geral, 0,50% para a AUC ROC média, 1,43% para a precisão média, 1,91% para a sensibilidade média e 1,69% para o F1-score médio, mas, especificamente com relação à função estratégica vice-líder, uma melhora de 1,17% e 0,49% para as métricas precisão e F1-score, e, com relação à líder, de 0,39% com relação à sensibilidade.

Por fim, o modelo SVM, apresentou piora com relação à média de todas as métricas de avaliação, sem qualquer exceção com relação a alguma função estratégia específica. Essa piora foi de 9,12% para a acurácia geral, 3,79% para a AUC ROC média, 10,48% para a precisão média, 13,04% para a sensibilidade média e 11,94% para o F1-score médio. As funções estratégicas líder e vice-líder apresentaram as maiores pioras com relação a AUC ROC, e as primeiro-preço e premium para a precisão, sensibilidade e F1-score.

O desempenho dos modelos após a redução das variáveis de entrada em termos da acurácia por classe da função estratégica dos itens está apresentado na Tabela 4.17.

Tabela 4.17 – Acurácias por classe dos modelos após a redução das variáveis

Classes Atribuídas → Originais ↓	Líder	Premium	Primeiro-Preço	Vice-Líder
KNN				
Líder	91,32%	2,75%	2,03%	3,91%
Premium	10,67%	81,84%	3,68%	3,80%
Primeiro-Preço	8,08%	2,99%	83,33%	5,60%
Vice-Líder	6,78%	1,79%	2,39%	89,04%
DT				
Líder	92,41%	2,34%	1,44%	3,81%
Premium	11,66%	79,51%	2,94%	5,89%
Primeiro-Preço	9,95%	3,73%	80,22%	6,09%
Vice-Líder	7,81%	2,77%	2,55%	86,87%
RF				
Líder	93,75%	1,69%	1,19%	3,37%
Premium	11,41%	83,19%	1,96%	3,44%
Primeiro-Preço	8,21%	1,87%	84,20%	5,72%
Vice-Líder	7,65%	1,52%	1,63%	89,20%
SVM				
Líder	86,19%	3,19%	2,56%	8,06%
Premium	23,80%	60,98%	4,54%	10,67%
Primeiro-Preço	21,77%	3,36%	65,05%	9,83%
Vice-Líder	16,87%	3,64%	4,88%	74,61%
ANN				
Líder	89,13%	2,94%	2,34%	5,59%
Premium	15,09%	75,09%	4,05%	5,77%
Primeiro-Preço	10,32%	3,11%	78,86%	7,71%
Vice-Líder	10,47%	2,50%	3,58%	83,45%

Obs.: Porcentagens acertadas estão salientadas em negrito.

Fonte: A autora (2022)

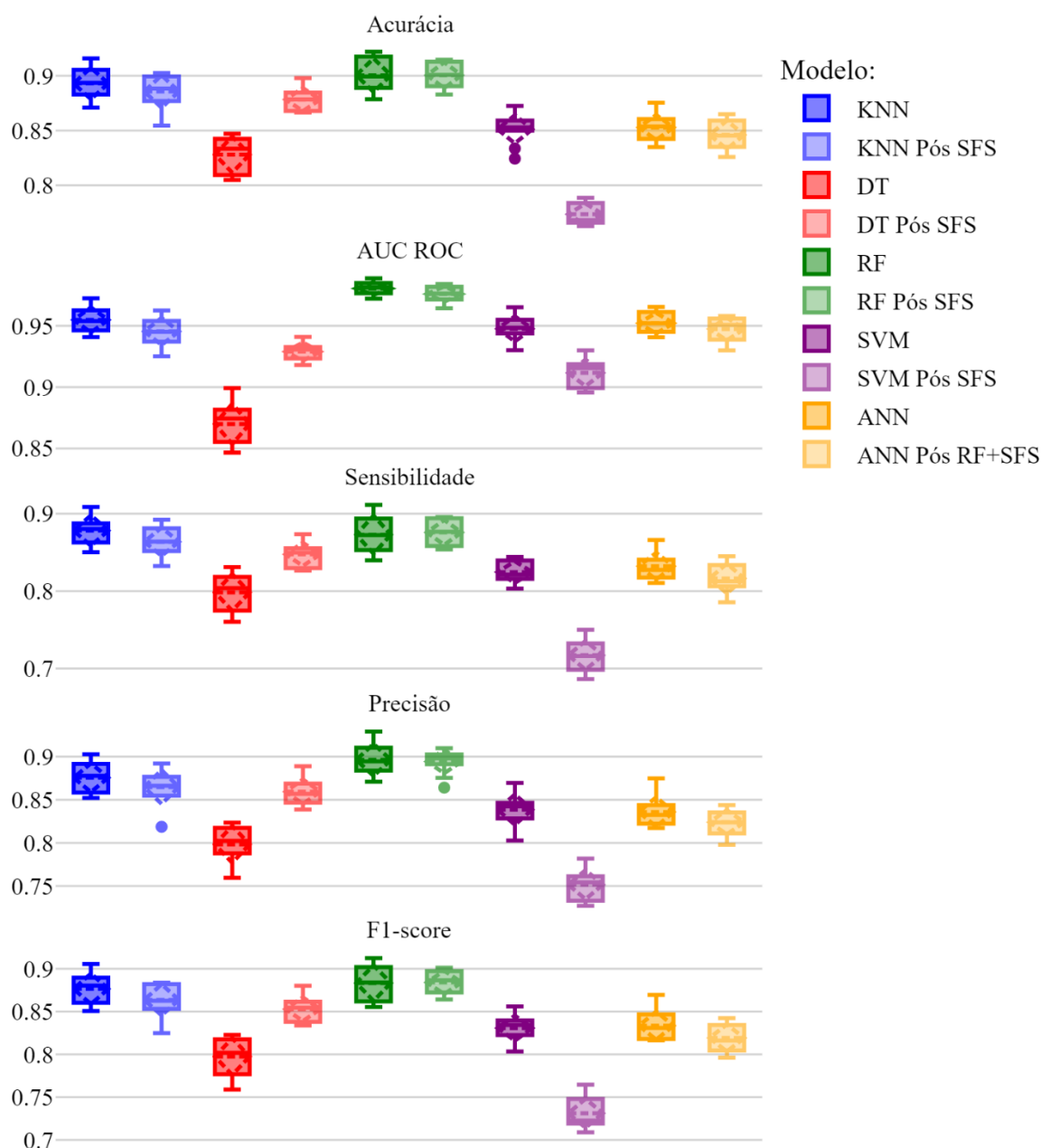
As acurácias após a redução das variáveis de entrada, conforme apresentado na Tabela 4.17, foram relativamente as mesmas para todos os modelos considerando as diferentes classes da função estratégica. Todos tiveram maior êxito em reconhecer as classes líder e vice-líder, majoritárias, e menor êxito em reconhecer as classes premium e primeiro-preço, minoritárias. Ainda, os modelos RF e KNN tiveram acertos relativos e capacidades de discriminação muito próximas, especialmente com relação às classes majoritárias.

As classes líder e vice-líder foram as menos confundidas, tendo a líder tido acertos relativos de 93,75%, 92,41%, 91,32% pelos RF, DT e KNN, nessa ordem, e 89,13% e 86,19% pelo ANN e SVM. Já, a vice-líder, de 89,20%, 89,04% e 86,87% pelos RF, KNN e DT, e 89,20% e 74,61% também pelo ANN e SVM. Por outro lado, as classes minoritárias primeiro-preço e premium tiveram acertos relativos próximos a 80% pelos modelos RF, KNN e DT, a 75% pelo ANN e a 60% pelo SVM. Essas classes primeiro-preço e premium, assim como a

líder, quando confundidas, na maioria das vezes foram com a classe líder, e está última, com a vice-líder.

A distribuição dos resultados da avaliação dos modelos, antes e após a redução das variáveis pela técnica de seleção adotada, está ilustrada na Figura 4.7.

Figura 4.7 – *Boxplots* do Desempenho dos modelos Pré e Pós Seleção das Variáveis



Fonte: A autora (2022)

Os modelos RF e KNN, conforme ilustrado na Figura 4.7, foram os que apresentaram, na média, os melhores desempenhos, respectivamente. Em seguida, antes da redução das variáveis de entrada, os modelos ANN e SVM e, após a redução das variáveis de entrada, os

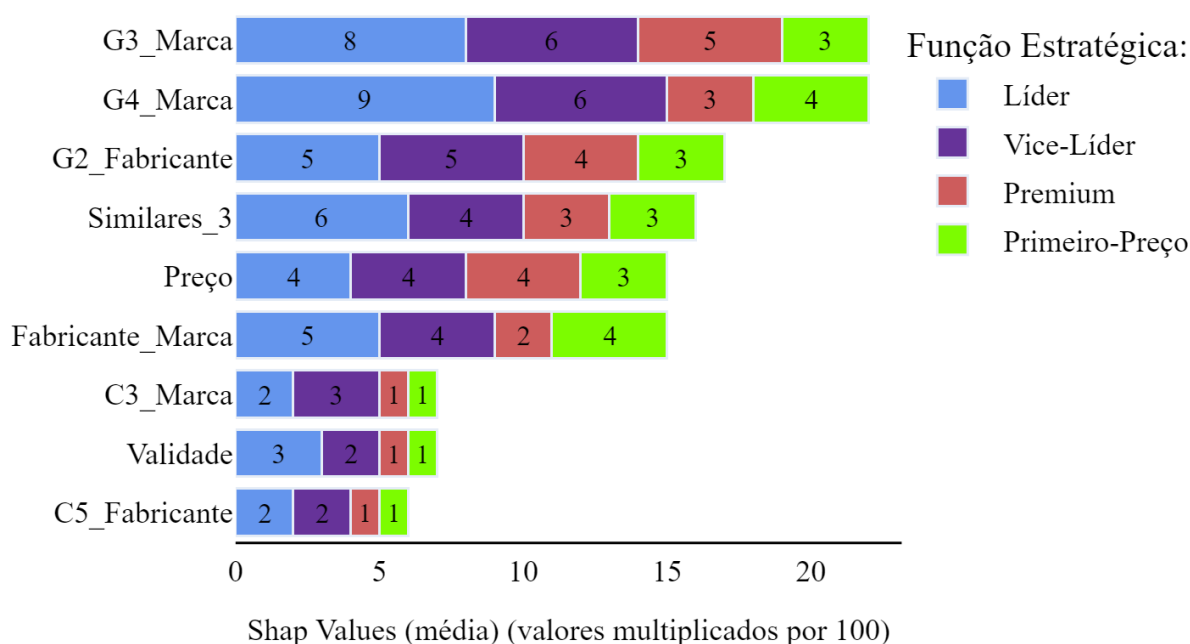
modelos DT e ANN. O modelo com o pior desempenho foi, antes da redução das variáveis, o DT, e, após a redução das variáveis, o SVM. Ainda, esses dois modelos DT e SVM foram os que apresentaram, após a redução das variáveis de entrada, respectivamente, maior melhoria, média geral de 6% para o DT, e maior pioria, média geral de 8% para o SVM, enquanto o modelo RF se manteve praticamente igual, média geral de 0%, e os modelos KNN e ANN apresentaram pouca diferença, mas para pior, média geral de 1% e 1,5%, respectivamente.

Já, com relação à variabilidade que os resultados desses modelos apresentaram, os KNN, ANN e SVM não apresentaram nenhuma diferença significativa antes e após a redução das variáveis de entrada, mantendo-se com uma variabilidade média na ordem de $1,6 \cdot 10^{-2}$, $1,4 \cdot 10^{-2}$ e $1,4 \cdot 10^{-2}$, respectivamente. Por outro lado, os modelos DT e RF, exceto com relação à métrica AUC ROC média, relativa à qual o modelo RF apresentou variabilidade pequena desde antes da redução das variáveis de entrada, de $5 \cdot 10^{-2}$, apresentaram, no geral, redução de variabilidade mais significativa, de $0,8 \cdot 10^{-2}$ e $0,5 \cdot 10^{-2}$, respectivamente.

4.4. Interpretação do melhor Modelo Implementado

A importância das variáveis de entrada que foram selecionadas via técnica SFS para o modelo RF segundo o método SHAP está ilustrada na Figura 4.8.

Figura 4.8 – Importância das Variáveis de entrada segundo a técnica SHAP



Fonte: A autora (2022)

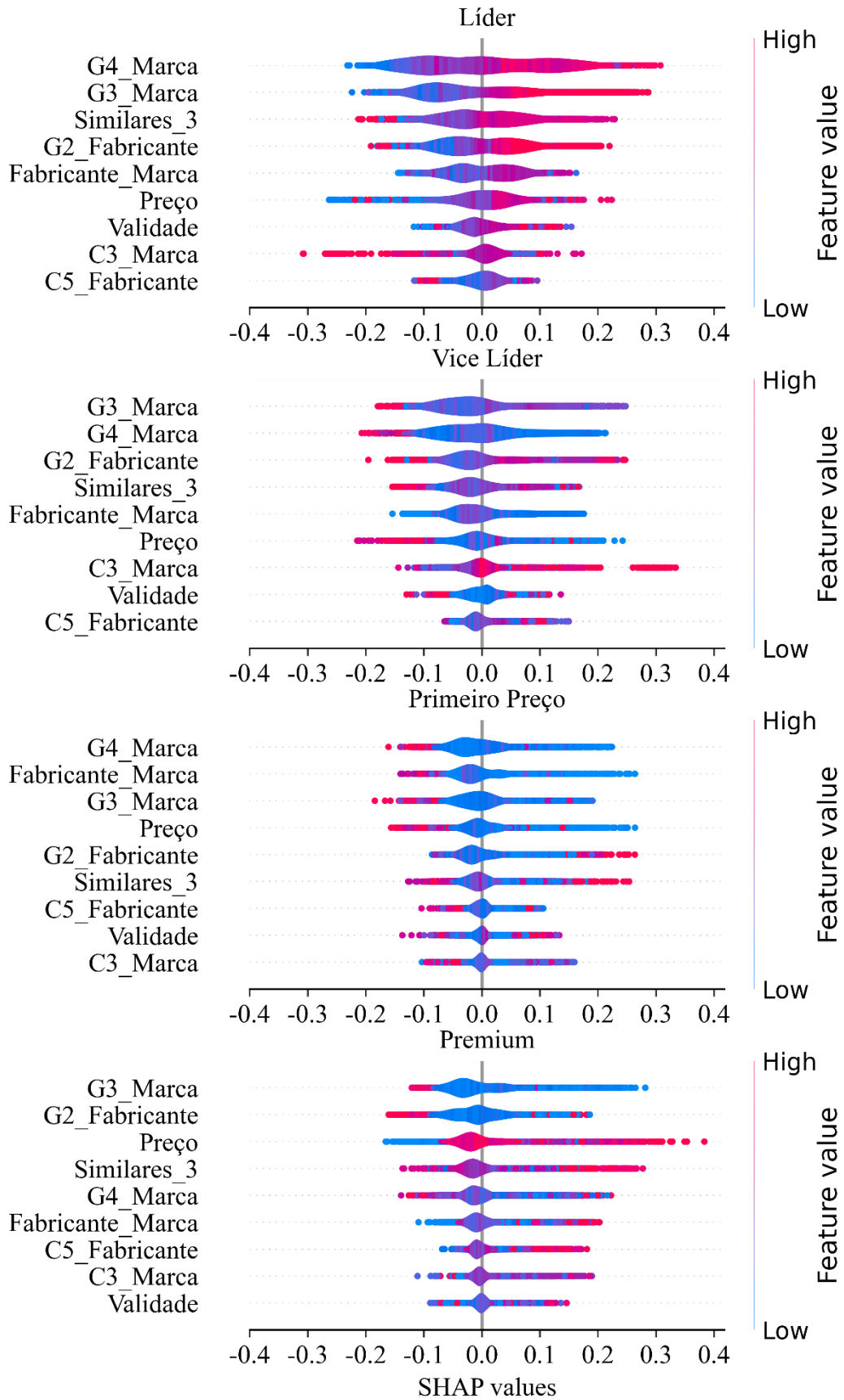
Com relação à classificação geral das diferentes classes de função estratégica, conforme ilustrado na Figura 4.8, as duas variáveis que mais se destacaram foram as correspondentes às participações da marca em termos de itens nos grupos 3 e 4, seguidas do fabricante no grupo 2, da quantidade de similares no grupo 3 e posteriormente do preço e da quantidade de marcas que o fabricante possui. Por fim, as três menos importantes foram a participação da marca em termos de categorias no grupo 3, o prazo de validade do item e a participação do fabricante também em termos de categorias no grupo 5.

Com relação à classificação individual de cada classe, para a líder e vice-líder as duas variáveis mais importantes foram as mesmas, assim como as quatro intermediárias e as três últimas. Já, para a classificação como primeiro-preço e premium essa correspondência ocorreu apenas no caso de duas intermediárias, preço e quantidade de similares no grupo 3, além das três menos importantes, que foram as mesmas também para a classificação como líder e vice-líder, sendo essas a participação da marca e fabricante em termos de categorias no grupo 3 e 5, respectivamente, e o prazo de validade do item.

As duas variáveis mais importantes para a classificação como primeiro-preço foram a participação da marca em termos de itens no grupo 4 e a quantidade de marcas que o fabricante possui, enquanto que para a classificação como premium foram a participação da marca e do fabricante em termos de itens nos grupos 3 e 2, respectivamente. Ainda, o preço teve importância relativa maior para as classes premium e primeiro-preço, ocupando a terceira e quarta posição, respectivamente, e menor para as classes líder e vice-líder, ocupando a sexta posição para ambas.

A relação entre o valor de cada variável e o quanto cada uma incorporou à probabilidade atribuída pelo modelo de cada item pertencer a determinada classe está ilustrada na Figura 4.9

Figura 4.9 - Valores Shap para cada Função Estratégica individualmente



Fonte: A autora (2022)

Quanto maior a participação da marca em termos de itens nos grupos 4 e 3, conforme ilustrado na Figura 4.9, maior foi a probabilidade do mesmo ser classificado como líder, e menor como uma das demais classes, assim como quanto maior a participação do fabricante em termos de itens no grupo 2, menor foi a chance do mesmo ser classificado como premium. Também, quanto maior foi o preço do item, maior foi a sua probabilidade de ser classificado como líder ou premium, e, quanto menor, maior foi a chance de ser classificado como vice-líder ou primeiro-preço. Por fim, quanto menor a quantidade de marcas que o fabricante do item possuía, maior foi a chance do item ser classificado como primeiro-preço.

O resultado da classificação realizada para 4 exemplos de itens, assim como suas probabilidades base e probabilidades atribuídas pelo modelo, estão apresentados na Tabela 4.18.

Tabela 4.18 – Exemplos de itens e suas respectivas predições e probabilidades

Predição/ Classificação	*Probab. Base	Probabilidades de predição atribuídas pelo modelo			
		Líder	Vice-Líder	Primeiro- Preço	Premium
Líder	0,48	0,54	0,21	0,08	0,17
Vice-Líder	0,28	0,00	1,00	0,00	0,00
Primeiro-Preço	0,12	0,12	0,23	0,60	0,06
Premium	0,12	0,06	0,06	0,02	0,87

Obs. 1: *Probabilidade Base corresponde à proporção existente de cada classe na base de dados objeto desta pesquisa

Obs. 2: Os maiores valores das Probabilidades de Predição estão salientados em negrito

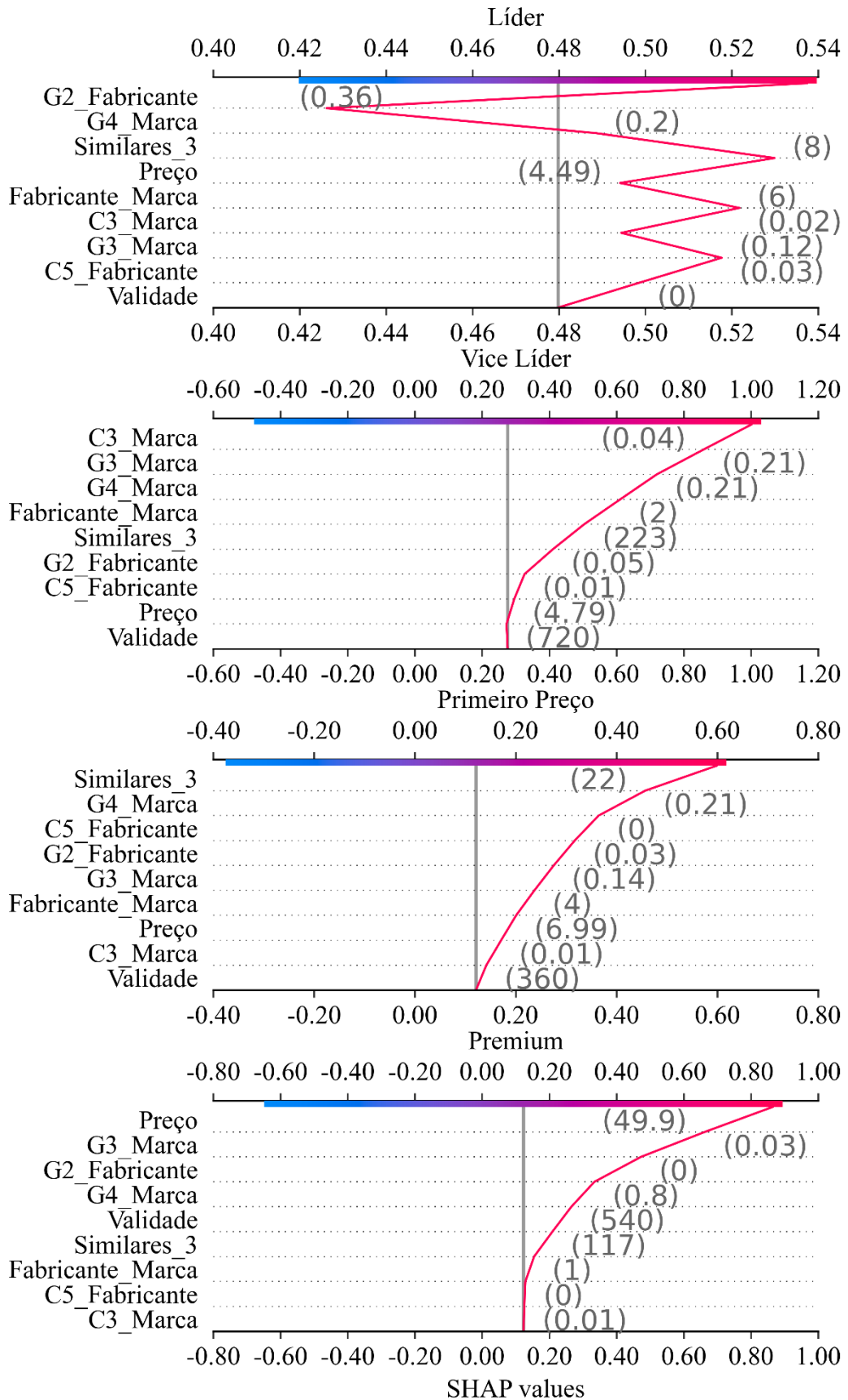
Fonte: A autora (2022)

Para o item classificado como líder, conforme apresentado na Tabela 4.18, a probabilidade base do item ser dessa classe era de 0,48, e a probabilidade final atribuída pelo modelo foi de 0,54. As probabilidades do modelo designar as outras classes foram de 0,21, 0,08 e 0,17. Da mesma forma, para o item classificado como vice-líder, a probabilidade base do item ser dessa classe era de 0,28, enquanto a probabilidade final atribuída pelo modelo foi de 1,0 e as probabilidades do modelo designar as demais classes foram de 0,0.

Para o item classificado como primeiro-preço, a probabilidade base do item pertencer a essa classe era de 0,12, enquanto a probabilidade final foi de 0,60 e as probabilidades do modelo atribuir as demais classes ao item foram de 0,12, 0,23 e 0,12. Por fim, para o item classificado como premium, a probabilidade base do item pertencer a essa classe era de 0,12, a probabilidade final foi de 0,87 e as probabilidades do modelo atribuir as outras classes ao item foram de 0,06, 0,06 e 0,02, respectivamente.

As probabilidades atribuídas por cada variável e seus efeitos na probabilidade final atribuída pelo modelo para os mesmos itens já exemplificados, estão ilustradas na Figura 4.10.

Figura 4.10 - Valores Shap Cumulativos - Exemplo de um item de cada Função Estratégica



Fonte: A autora (2022)

Para o item classificado como líder, conforme ilustrado na Figura 4.10, as variáveis que fizeram o modelo aumentar sua probabilidade base de 48% de pertencer a essa classe foram o prazo de validade do item, a participação do fabricante em termos de categorias no grupo 5, a participação da marca no grupo 3, bem como seu preço e participação do fabricante em termos de itens no grupo 2, respectivamente. Os valores desses aumentos foram de 2%, 2%, 3%, 4% e 11%, nessa ordem.

Por outro lado, ainda para esse item, as variáveis que diminuíram a probabilidade do modelo realizar essa classificação foram a participação da marca em termos de itens no grupo 3, a quantidade de marcas que o fabricante possui, a quantidade de similares no grupo 3 e a participação da marca também em termos de itens no grupo 4. Os valores dessas diminuições foram de, respectivamente, 2%, 3%, 4% e 6%, e assim como os valores de aumento, se somaram à probabilidade base do item, resultando na probabilidade final atribuída pelo modelo de 54%.

Por fim, para o item classificado como vice-líder, todas as variáveis de entrada aumentaram em algum valor a probabilidade do item pertencer a essa classe, e nenhuma diminuiu. No total acumulado à probabilidade base do item ser classificado como essa classe, a probabilidade final atribuída pelo modelo foi de 100%. Comportamento semelhante ocorreu para os itens classificados como primeiro-preço e premium, casos em que todas as variáveis contribuíram de forma positiva para a classificação final realizada pelo modelo.

4.5. Implicações Práticas para a Rede de Supermercados

As implicações práticas que os resultados desta pesquisa apresentam com relação à rede de supermercados a que dizem respeito abrangem dois aspectos diferentes. Um deles refere-se aos modelos de aprendizagem de máquina que foram construídos para reproduzir as classes dos itens de portfólio da organização quanto à função estratégica que desempenham. O outro, por sua vez, refere-se à base de dados que foi construída para esse fim.

Com relação à base de dados construída, a organização passou a ter acesso a informações que antes desconhecia a respeito da quantidade de similares que cada um de seus itens possui, da dimensão dos fabricantes em termos do quantitativo de marcas que engloba, assim como da participação relativa de seus fornecedores, seja no papel de fabricante ou de marca, em seu portfólio de suprimentos. Tais informações dão subsídios para que a organização entenda a representatividade de seus fornecedores e a dependência que possui destes.

Uma vez que a organização passe a ter noção dessa representatividade e dependência, possui condições de melhor delinear as práticas que permeiam seus relacionamentos. Essas práticas podem ser a busca por novos fornecedores, a disponibilidade de recursos para alavancagem ou manutenção dos relacionamentos, a implementação de critérios de avaliação, a tolerância a eventuais problemas como a frequência de indisponibilidade do item, além do atendimento a exigências recebidas e a pressão realizada para redução de preço ou obtenção de diferentes condições de pagamento.

Com relação aos modelos construídos, estes dão possibilidade da organização automatizar a classificação da função estratégica dos itens atualmente realizada de forma subjetiva, o que daria replicabilidade, escalabilidade e dinamicidade à classificação. Também daria condições, com um possível modelo DT simplificado/aprimorado, da organização passar a conhecer todas as regras envolvidas à classificação a ponto de conseguir entendê-las e sistematizá-las como a maneira oficial de classificar os itens, gerando uma maior confiança com relação às classes atribuídas.

Uma vez que a organização passe a ter mais confiança da assertividade da classe da função estratégica de seus itens, pode começar a utilizá-las para apoiar outras práticas que não apenas as relacionadas à manutenção do seu portfólio, como por exemplo sua política de preços, focando na oferta de valores mais baixos para os itens primeiro-preço ou em maior prazo de pagamento para os itens premium, por exemplo, sua política de promoções, focando na concessão de descontos para os itens vice-líder, por exemplo, assim com seu planejamento de exposição dos itens nas gôndolas, focando nos itens premium para serem dispostos à altura dos olhos dos clientes ou nos primeiro-preço para serem dispostos no começo dos corredores.

5. CONSIDERAÇÕES FINAIS

Este capítulo apresenta as constatações e conclusão central da pesquisa (5.1), suas contribuições acadêmicas e práticas (5.2), limitações (5.3) e oportunidades de pesquisas futuras decorrentes de sua realização (5.4).

5.1. Constatações e Conclusão da Pesquisa

A aprendizagem de máquina por meio de sua tarefa de classificação possui como pontos positivos o fato de ser capaz de reproduzir conhecimento tácito, de ser escalável e passível de adaptação à medida da transformação dos registros e respectivos atributos que processa. Por outro lado, possui como ponto negativo, a depender do tipo e grau de complexidade do modelo e da base de dados sobre o qual é implementado, a baixa interpretabilidade dos resultados que produz. Tais questões podem ser contornadas, respectivamente, pela aplicação de técnicas de seleção de variáveis de entrada e de métodos de interpretação dos modelos que possibilitem o entendimento dos resultados gerados.

Para o problema desta pesquisa, que consistiu na implementação/avaliação de modelos de aprendizagem de máquina para a classificação de itens do portfólio de uma rede varejista de supermercados quanto às funções estratégicas que desempenham, além da implementação e avaliação dos modelos de aprendizagem de máquina em si, houve também a aplicação e avaliação de técnicas de seleção de variáveis de entrada e a aplicação de um método para interpretação do modelo aplicado sobre o que teve o melhor desempenho quando implementado sobre a respectiva base de dados reduzida segundo a melhor técnica de seleção de atributos.

Com relação aos modelos em si, os resultados revelaram um melhor desempenho dos menos complexos, especificamente o modelo *ensemble* baseado em árvore de decisão, o RF, e o modelo baseado em instância, o KNN, nessa ordem. O modelo de árvore de decisão, por sua vez, apresentou desempenho intermediário, mas ainda com acurácia geral próxima a 80%, ao passo que, de forma contrária, os modelos mais complexos, SVM e ANN, respectivamente,

apresentaram os piores desempenhos. Esses resultados apontam os modelos menos complexos como um potencial meio de classificação dos itens, alternativo ao método subjetivo adotado pela rede de supermercados a que dizem respeito.

Quanto às técnicas de seleção de variáveis de entrada, os resultados mostraram que as técnicas baseadas em *Feature Importance* e *Permutation Importance* tiveram um desempenho parecido entre si, assim como a *F Classification* e *Mutual Info Classification*, com essas últimas tendo desempenhos superiores às primeiras para o modelo KNN, praticamente iguais para o RF, embora a *Feature Importance* ainda tenha se sobressaído, e inferiores para os DT, SVM e ANN. Também mostraram que as técnicas que tiveram um melhor desempenho foram a *wrapper* SFS e, especificamente para o modelo KNN, mas quase sem diferença, a técnica que consistiu na aplicação do resultado da combinação do modelo RF e da própria técnica SFS.

No que se refere ao método de interpretação de modelo, o único aplicado, o SHAP, por ser suficiente para prover a interpretação tanto local quanto globalmente, cumpriu a sua proposta de explicitar o quanto cada variável de entrada contribuiu para a probabilidade final atribuída pelo modelo do item pertencer a determinada classe. Esse método, entretanto, não foi utilizado para explicitar a forma global da relação entre cada variável de entrada e a probabilidade atribuída pela mesma, por requerer, para isso, a não existência de correlação entre a variável de entrada em questão e as demais.

Por fim, a conclusão da pesquisa foi a aplicabilidade, no contexto da gestão de suprimentos, da aprendizagem de máquina para a reprodução de classes de itens de portfólio sob sua tarefa de classificação quando implementada via modelos de floresta aleatória (RF), k-vizinhos mais próximos (KNN) e de árvore de decisão (DT), considerados menos complexos do que os também implementados rede neural artificial sob arquitetura do perceptron multicamadas (ANN) e máquinas de vetores de suporte (SVM). Também, a viabilidade da aplicação da técnica de seleção de variáveis de entrada, com destaque para a *wrapper* SFS, e do método SHAP para interpretabilidade do modelo.

5.2. Contribuições da Pesquisa

A contribuição acadêmica da pesquisa foi ter colaborado para o preenchimento da lacuna científica referente à escassa literatura sobre gestão de suprimentos apoiada por aprendizagem de máquina ao constatar empiricamente a aplicabilidade da classificação de unidades de manutenção de estoque (SKUs) no contexto da gestão de portfólio sendo apoiada pela tarefa de classificação principalmente por meio do modelo baseado em instância de k-

vizinhos mais próximos (KNN) e dos modelos baseados em árvore de decisão de árvore única (DT) e também o *ensemble* do tipo *bagging* floresta aleatória (RF).

Já, a contribuição prática foi ter desenvolvido um mecanismo embrionário de classificação de SKUs do portfólio de uma rede varejista de supermercados via aprendizagem de máquina capaz de dar dinamicidade à classificação à medida da mudança dos atributos dos itens, além de poder explicitar o conhecimento subjetivo que deu origem às suas classes. Ainda, essa classificação pode, aliada à extração de outros conhecimentos provenientes da base de dados, ser estendida aos fornecedores e relacionamentos e constituir um passo inicial para ser desdobrada em estratégias e abordagens de suprimentos e de relacionamento.

5.3. Limitações da Pesquisa

Constituíram limitações da pesquisa o fato das bases de dados bases de dados avulsas que deram origem à base objeto do estudo não terem sido desenvolvidas para apoiar qualquer prática de aprendizagem de máquina; dos modelos terem sido implementados somente sem o emprego prévio de quaisquer técnicas para lidar dados desbalanceados; dos valores de correlação entre os atributos da base não terem sido utilizados como técnica de seleção de variáveis de entrada; da busca pela melhor configuração dos parâmetros dos modelos ter sido feita por um único método, o *GridSearchCV*, sem a comparação com outros possíveis.

Também, o fato dos modelos terem sido implementados apenas sem estarem envoltos a alguma estratégia para lidar com multiclasse, *ovr* (*one versus rest*) ou *ovo* (*one versus one*), à exceção do modelo SVM para o qual a definição dessa estratégia constitui um parâmetro; somente considerando o cenário onde cada classe da função estratégica era única, sem contemplar a possibilidade das líder e vice-líder serem unificadas e comparadas às demais, premium e primeiro-preço, e estas, entre si; do modelo KNN ter sido implementado considerando unicamente os atributos de forma não ponderada; e, no geral, de terem sido dispensados os atributos correspondentes ao grupo 1 bem como a forma de aquisição do item.

Por fim, o fato da interpretação do melhor modelo ter sido realizada para cada registro somente em termos da apresentação da relação quantitativa entre o valor de cada um de seus atributos e o quanto cada um incorporou da probabilidade final atribuída pelo modelo do item pertencer a determinada classe, sem contemplar a apresentação da forma dessa relação, devido à técnica para isso, o gráfico de dependência parcial (*Partial Dependence Plot*, PDP), ter como premissa a não existência de correlação entre as variáveis de entrada, o que não ocorre para o problema da pesquisa.

5.4. Oportunidades de Pesquisas Futuras

Com relação à implementação da aprendizagem de máquina em si, as oportunidades de pesquisas futuras englobam a combinação de diferentes técnicas de seleção de variáveis; a implementação do modelo de árvore de decisão sem as limitações a que esteve sujeito neste trabalho de modo a tentar chegar a um desempenho equiparável ao do RF, mas com uma árvore aprimorada/simplificada com profundidade que possibilite a visualização e a consequente replicação de suas regras; a implementação/avaliação de outros modelos, como o *Naive Bayes* (NB), de Regras de Decisão e o *XGBoost*; e a implementação da tarefa de clusterização para posterior análise conjunta dos resultados ao desta pesquisa.

Já, com relação às possibilidades de exploração da base de dados que foi objeto desta pesquisa, as oportunidades futuras englobam a construção de novos atributos a partir dos disponíveis, como por exemplo, a quantidade de similares que são de marca ou fabricante concorrentes ao item em questão, e a extração de outros conhecimentos em termos de constructos que permeiam o relacionamento entre empresa e fornecedor, como por exemplo vulnerabilidade e força, de modo que a classificação possa ser estendida aos fornecedores e, consequentemente, vir a constituir um passo inicial para ser desdobrada em abordagens de suprimentos, políticas de compra e estratégias de relacionamento.

Por fim, com relação ao tema desta pesquisa, as oportunidades futuras englobam a realização de uma revisão sistemática de literatura sobre classificação/segmentação no contexto da gestão de suprimentos que venha a abarcar trabalhos atuais, levantando de modo sistematizado os objetos da classificação bem como os constructos que permeiam o relacionamento entre empresa e fornecedor e seus elementos constituintes. Também, uma revisão sistemática sobre gestão de suprimentos apoiada pela aprendizagem de máquina que venha a contemplar a escassa literatura existente sobre o tema, levantando os modelos implementados e seus objetos de estudo.

REFERÊNCIAS

- AGGARWAL, Charu C. **Data Mining**. Cham: Springer International Publishing, 2015. *E-book*. Disponível em: <http://link.springer.com/10.1007/978-3-319-14142-8>.
- BAUM, Eric B. On the capabilities of multilayer perceptrons. **Journal of Complexity**, [s. l.], v. 4, n. 3, p. 193–215, 1988. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/0885064X88900209>.
- BENSAOU, M. Portfolios of Buyer-Supplier Relationships. **Sloan Management Review**, [s. l.], v. 40, n. 4, p. 33–44, 1999.
- BREIMAN, Leo. **Bagging Predictors**. [S. l.: s. n.], 1994.
- BREIMAN, Leo *et al.* **Classification And Regression Trees**. New York: Taylor & Francis Group, 1984. *E-book*. Disponível em: <https://www.taylorfrancis.com/books/9781351460491>.
- BREIMAN, Leo. Random Forests. **Machine Learning**, [s. l.], v. 45, n. 1, p. 5–32, 2001. Disponível em: <https://doi.org/10.1023/A:1010933404324>.
- CANIÈLS, Marjolein C.J.; GELDERMAN, Cees J. Power and interdependence in buyer supplier relationships: A purchasing portfolio approach. **Industrial Marketing Management**, [s. l.], v. 36, n. 2, p. 219–229, 2007. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S001985010500132X>.
- CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, [s. l.], v. 20, n. 3, p. 273–297, 1995. Disponível em: <http://link.springer.com/10.1007/BF00994018>.
- COVER, T. Estimation by the nearest neighbor rule. **IEEE Transactions on Information Theory**, [s. l.], v. 14, n. 1, p. 50–55, 1968. Disponível em: <http://ieeexplore.ieee.org/document/1054098/>.
- COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, [s. l.], v. 13, n. 1, p. 21–27, 1967. Disponível em: <http://ieeexplore.ieee.org/document/1053964/>.
- DAY, Marc; MAGNAN, Gregory M.; MOELLER, Morten Munkgaard. Evaluating the bases of supplier segmentation: A review and taxonomy. **Industrial Marketing Management**, [s. l.], v. 39, n. 4, p. 625–639, 2010. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0019850109000819>.
- DOUMPOS, Michael; ZOPOUNIDIS, Constantin. A multicriteria classification approach based on pairwise comparisons. **European Journal of Operational Research**, [s. l.], v. 158, n. 2, p. 378–389, 2004. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0377221703004909>.
- DYER, H. J.; CHO, D. S.; CHU, W. Strategic supplier segmentation: The next “Best practice” in supply chain management. **California Management Review**, [s. l.], v. 40, n. 2, p. 57–77,

1998.

ELLRAM, Lisa M. *et al.* Purchasing and supply management's identity: Crisis? What crisis? **Journal of Purchasing and Supply Management**, [s. l.], v. 26, n. 1, p. 100583, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1478409218301468>.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, [s. l.], v. 17, n. 3, 1996. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1230/1131>.

FIX, E; HODGES, J L. Discriminatory analysis, nonparametric discrimination: Consistency properties. **Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties**, [s. l.], 1951.

FIX, E; HODGES, J. Discriminatory analysis: Small sample performance. **Discriminatory Analysis: Small Sample Performance**, [s. l.], 1952.

FLORES, Benito E.; OLSON, David L.; DORAI, V.K. Management of multicriteria inventory classification. **Mathematical and Computer Modelling**, [s. l.], v. 16, n. 12, p. 71–82, 1992. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/089571779290021C>.

FÜRNKRANZ, Johannes. Decision Tree. *In*: SAMMUT, Claude; WEBB, Geoffrey I. (org.). **Encyclopedia of Machine Learning**. Boston, MA: Springer US, 2010. p. 263–267. *E-book*. Disponível em: https://doi.org/10.1007/978-0-387-30164-8_204.

GELFAND, S.B.; RAVISHANKAR, C.S.; DELP, E.J. An iterative growing and pruning algorithm for classification tree design. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [s. l.], v. 13, n. 2, p. 163–174, 1991. Disponível em: <http://ieeexplore.ieee.org/document/67645/>.

GERÓN, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. 2. ed. Sebastopol, Canada: O'Reilly Media, Inc, 2019.

GOLDSTEIN, M. k-Nearest Neighbor Classification. **IEEE Transactions on Information Theory**, [s. l.], v. 18, n. 5, p. 627–630, 1972. Disponível em: <http://ieeexplore.ieee.org/document/1054888/>.

GUPTA, Sushil; VERMA, Rohit; VICTORINO, Liana. Empirical Research Published in Production and Operations Management (1992-2005): Trends and Future Research Directions. **Production and Operations Management**, [s. l.], v. 15, n. 3, p. 432–448, 2006. Disponível em: <http://doi.wiley.com/10.1111/j.1937-5956.2006.tb00256.x>.

HALLIKAS, Jukka *et al.* Risk-based classification of supplier relationships. **Journal of Purchasing and Supply Management**, [s. l.], v. 11, n. 2–3, p. 72–82, 2005. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1478409205000713>.

HUNT, E. B.; MARIN, J.; STONE, P. J. **Experiments in Induction**. 1. ed. New York: [s. n.], 1966.

IBGE. **Pesquisa Anual de Comércio - 2019**. [S. l.], 2019. Disponível em: <https://sidra.ibge.gov.br/tabela/1418#resultado>. Acesso em: 22 jun. 2022.

IBGE. **Pesquisa Nacional por Amostra de Domicílios Contínua trimestral - 2022**. [S. l.], 2022. Disponível em: <https://sidra.ibge.gov.br/tabela/5434#resultado>. Acesso em: 22 jun. 2022.

IVAKNENKO, A. G.; LAPA, V. G. **Cybernetic predicting devices**. New York: CCM Information Corp., 1965.

KAMATH, Rajan R.; LIKER, Jeffrey. **A Second Look at Japanese Product Development**.

Harvard Business Review, [s. l.], v. 72, n. 6, p. 154–170, 1994.

KARTAL, Hasan *et al.* An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. **Computers and Industrial Engineering**, [s. l.], v. 101, p. 599–613, 2016.

KAUFMAN, Allen; WOOD, Craig H.; THEYEL, Gregory. Collaboration and technology linkages: a strategic supplier typology. **Strategic Management Journal**, [s. l.], v. 21, n. 6, p. 649–663, 2000. Disponível em: [https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0266\(200006\)21:6%3C649::AID-SMJ108%3E3.0.CO;2-U](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0266(200006)21:6%3C649::AID-SMJ108%3E3.0.CO;2-U).

KEARNS, Michael; VALIANT, L.G. Cryptographic limitations on learning Boolean formulae and finite automata. *In:* , 1989. **Proceedings of the twenty-first annual ACM symposium on Theory of computing**. [S. l.: s. n.], 1989. p. 433–444.

KEARNS, Michael; VALIANT, L.G. **Learning Boolean formulae or finite automata is as hard as factoring** Technical Report TR-14-88 Harvard University Aikem Computation Laboratory. [S. l.: s. n.], 1988.

KOHAVI, Ron; JOHN, George H. Wrappers for feature subset selection. **Artificial Intelligence**, [s. l.], v. 97, n. 1–2, p. 273–324, 1997. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S000437029700043X>.

KRALJIC, Peter. Purchasing Must Become Supply Management. **Harvard Business Review**, [s. l.], v. 61, n. September-October, p. 109–117, 1983.

KUBAT, Miroslav. **An Introduction to Machine Learning**. 2. ed. Cham: Springer International Publishing, 2017. *E-book*. Disponível em: <http://link.springer.com/10.1007/978-3-319-63913-0>.

LAJILI, Imen; LADHARI, Talel; BABAI, Zied. Adaptive machine learning classifiers for the class imbalance problem in ABC inventory classification. *In:* , 2016. **ILS 2016 - 6th International Conference on Information Systems, Logistics and Supply Chain**. [S. l.]: International Conference on Information Systems, Logistics and Supply Chain, 2016.

LAJIMI, F. H.; MAJIDI, S. Supplier segmentation: A systematic literature review. **Journal of Supply Chain Management Science**, [s. l.], v. 2, n. 3–4, p. 138–158, 2021. Disponível em: <https://journals.open.tudelft.nl/jscms/article/view/6151>.

LOLLI, F. *et al.* Machine learning for multi-criteria inventory classification applied to intermittent demand. **Production Planning and Control**, [s. l.], v. 30, n. 1, p. 76–89, 2019.

LUNDBERG, Scott M; LEE, Su-In. A unified approach to interpreting model predictions. *In:* , 2017. **Advances in Neural Information Processing Systems**. [S. l.: s. n.], 2017. p. 4765–4774.

MISHRA, Ayaskanta; MOHAPATRO, Manaswini. Real-time RFID-based item tracking using IoT efficient inventory management using Machine Learning. *In:* , 2020. **4th IEEE Conference on Information and Communication Technology, CICT 2020**. [S. l.]: Institute of Electrical and Electronics Engineers Inc., 2020.

MOLNAR, Christoph. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**. 2. ed. [S. l.: s. n.], 2022. *E-book*. Disponível em: <https://christophm.github.io/interpretable-ml-book>.

NELLORE, Rajesh; SÖDERQUIST, Klas. Portfolio approaches to procurement: Analysing the missing link to specifications. **Long Range Planning**, [s. l.], v. 33, n. 2, p. 245–267, 2000. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0024630100000273>.

- NEUMANN, John Von; MORGENSTERN, Oskar. **Theory of Games and Economic Behavior**. [S. l.]: Princeton, 1944. *E-book*. Disponível em: <https://jmvidal.cse.sc.edu/library/neumann44a.pdf>.
- NTAKOLIA, C. *et al.* An explainable machine learning model for material backorder prediction in inventory management. **Sensors**, [s. l.], v. 21, n. 23, 2021.
- OLSEN, Rasmus Friis; ELLRAM, Lisa M. A portfolio approach to supplier relationships. **Industrial Marketing Management**, [s. l.], v. 26, n. 2, p. 101–113, 1997. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0019850196000892>.
- PONTIL, Massimiliano; VERRI, Alessandro. Properties of Support Vector Machines. **Neural Computation**, [s. l.], v. 10, n. 4, p. 955–974, 1998. Disponível em: <https://direct.mit.edu/neco/article/10/4/955-974/6169>.
- R.A. REID. The ABC method in hospital inventory management: A practical approach. **Production and Inventory Management Journal**, [s. l.], v. 28, n. 4, p. 67–70, 1987.
- RAHIM, Robbi *et al.* C4.5 classification data mining for inventory control. **International Journal of Engineering and Technology(UAE)**, [s. l.], v. 7, p. 68–72, 2018.
- RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTRIN, Carlos. “Why Should I Trust You?” *In:* , 2016, New York, NY, USA. **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2016. p. 1135–1144. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939778>.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, [s. l.], v. 65, n. 6, p. 386–408, 1958. Disponível em: <http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042519>.
- SAFAVIAN, S.R.; LANDGREBE, D. A survey of decision tree classifier methodology. **IEEE Transactions on Systems, Man, and Cybernetics**, [s. l.], v. 21, n. 3, p. 660–674, 1991. Disponível em: <http://ieeexplore.ieee.org/document/97458/>.
- SAMMUT, Claude; WEBB, Geoffrey I. (org.). **Encyclopedia of Machine Learning**. Boston, MA: Springer US, 2010. *E-book*. Disponível em: <http://link.springer.com/10.1007/978-0-387-30164-8>.
- SANTIS, R.B. de; AGUIAR, E.P. de; GOLIATT, L. Predicting material backorders in inventory management using machine learning. *In:* , 2018. **2017 IEEE Latin American Conference on Computational Intelligence, LA-CCI 2017 - Proceedings**. [S. l.: s. n.], 2018. p. 1–6.
- SCHAPIRE, Robert E. The strength of weak learnability. **Machine Learning**, [s. l.], v. 5, n. 2, p. 197–227, 1990. Disponível em: <http://link.springer.com/10.1007/BF00116037>.
- SCHMIDHUBER, Jürgen. Deep learning in neural networks: An overview. **Neural Networks**, [s. l.], v. 61, p. 85–117, 2015. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0893608014002135>.
- SCIKIT-LEARN. **1.10. Decision Trees: 1.10.6. tree algorithms: id3, c4.5, c5.0 and cart**. [S. l.], 2022. Disponível em: <https://scikit-learn.org/stable/modules/tree.html>. Acesso em: 17 jun. 2022.
- SHAPLEY, L. S. A Value for n-Person Games. *In:* CONTRIBUTIONS TO THE THEORY OF GAMES (AM-28), VOLUME II. [S. l.]: Princeton University Press, 1953. p. 307–318. *E-book*. Disponível em: <https://www.degruyter.com/document/doi/10.1515/9781400881970-018/html>.
- SVENSSON, Göran. Interactive vulnerability in buyer-seller relationships: a dyadic approach.

International Journal of Physical Distribution & Logistics Management, [s. l.], v. 34, n. 8, p. 662–682, 2004. Disponível em:

<https://www.emerald.com/insight/content/doi/10.1108/09600030410557785/full/html>.

SVOBODA, J.; MINNER, S. Tailoring inventory classification to industry applications: the benefits of understandable machine learning. **International Journal of Production Research**, [s. l.], 2021.

TANG, Christopher S. Supplier Relationship Map. **International Journal of Logistics Research and Applications**, [s. l.], v. 2, n. 1, p. 39–56, 1999. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/13675569908901571>.

TING, Kai Ming. Precision. In: SAMMUT, Claude; WEBB, Geoffrey I (org.). **Encyclopedia of Machine Learning**. Boston, MA: Springer US, 2010. p. 780. *E-book*. Disponível em:

https://doi.org/10.1007/978-0-387-30164-8_651.

WALTERS-WILLIAMS, Janett; LI, Yan. Comparative Study of Distance Functions for Nearest Neighbors. In: **ADVANCED TECHNIQUES IN COMPUTING SCIENCES AND SOFTWARE ENGINEERING**. Dordrecht: Springer Netherlands, 2010. p. 79–84. *E-book*. Disponível em: http://link.springer.com/10.1007/978-90-481-3660-5_14.

WERBOS, Paul J. Applications of advances in nonlinear sensitivity analysis. In: **SYSTEM MODELING AND OPTIMIZATION**. Berlin/Heidelberg: Springer-Verlag, 1981. p. 762–770. *E-book*. Disponível em: <http://link.springer.com/10.1007/BFb0006203>.

WOLPERT, David H. Stacked generalization. **Neural Networks**, [s. l.], v. 5, n. 2, p. 241–259, 1992. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0893608005800231>.

YANG, Qinghong *et al.* Machine Learning Based Prediction and Prevention of Malicious Inventory Occupied Orders. **International Journal of Mobile Computing and Multimedia Communications**, [s. l.], v. 6, n. 4, p. 56–72, 2014.