

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS
CURSO ENGENHARIA DE PRODUÇÃO ELÉTRICA

Vitor Rodrigues

Um Modelo de Previsão de Demanda no Varejo do Setor de Saúde e Bem-Estar

Florianópolis

2021

Vitor Rodrigues

Um Modelo de Previsão de Demanda no Varejo do Setor de Saúde e Bem-Estar

Trabalho Conclusão do Curso de Graduação em Engenharia de Produção Elétrica do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito para a obtenção do título em Engenharia, área Eletricidade, habilitação Engenharia de Produção Elétrica

Orientador: Prof. Carlos Ernani Fries, Dr.

Florianópolis

2021

Ficha de identificação da obra

Rodrigues, Vitor
Um Modelo de Previsão de Demanda no Varejo do Setor de
Saúde e Bem Estar / Vitor Rodrigues ; orientador, Carlos
Ernani Fries, 2021.
90 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia de Produção Elétrica, Florianópolis,
2021.

Inclui referências.

1. Engenharia de Produção Elétrica. 2. Previsão de
demanda. 3. Clusterização de séries temporais. 4. Varejo.
I. Fries, Carlos Ernani. II. Universidade Federal de Santa
Catarina. Graduação em Engenharia de Produção Elétrica. III.
Título.

Vitor Rodrigues

Um Modelo de Previsão de Demanda no Varejo do Setor de Saúde e Bem-Estar

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título em Engenharia, área Eletricidade, habilitação Engenharia de Produção Elétrica e aprovado em sua forma final pelo Curso de Graduação em Engenharia de Produção Elétrica

Florianópolis, 10 de maio de 2021.



Documento assinado digitalmente
Monica Maria Mendes Luna
Data: 19/05/2021 16:00:22-0300
CPF: 314.104.993-91
Verifique as assinaturas em <https://v.ufsc.br>

Profa. Mônica Maria Mendes Luna, Dra.
Coordenadora do Curso

Banca Examinadora:



Documento assinado digitalmente
Carlos Ernani Fries
Data: 19/05/2021 14:49:25-0300
CPF: 304.616.699-00
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Carlos Ernani Fries, Dr.
Orientador

Universidade Federal de Santa Catarina



Documento assinado digitalmente
Daniel Christian Henrique
Data: 19/05/2021 14:55:59-0300
CPF: 288.178.928-58
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Daniel Christian Henrique, Dr.
Avaliador

Universidade Federal de Santa Catarina



Documento assinado digitalmente
Diego de Castro Fettermann
Data: 19/05/2021 15:00:49-0300
CPF: 931.923.710-72
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Diego de Castro Fettermann, Dr.
Avaliador

Universidade Federal de Santa Catarina

Este trabalho é dedicado à minha família, por sempre incentivarem meus estudos, à minha namorada, pelo companheirismo nessa jornada, e ao meu orientador Carlos, por toda dedicação e apoio no desenvolvimento desta pesquisa.

AGRADECIMENTOS

Agradeço ao meu pai, Paulo Sérgio Rodrigues, por me incentivar, me ensinar lições da vida e ser meu exemplo desde criança. Agradeço à minha mãe, Sandra Helena Rodrigues, por todo o amor, apoio e ajuda que recebi. Agradeço à minha irmã, Maitê Rodrigues, que sempre me apoiou e cuidou de mim. Agradeço ao meu cunhado, Nicolas Keunecke, que foi meu colega e me ajudou em diversos momentos.

Agradeço à minha namorada, Gabriela Soares, por sempre estar ao meu lado, me apoiando e me incentivando, e por ser minha companheira desde o início da graduação. Agradeço aos meus sogros, Regina Célia e José Messias, por sempre serem receptivos e solícitos.

Ao meu orientador, Carlos Ernani Fries, por me guiar e participar ativamente de toda a pesquisa. Agradeço também a todos os professores que me ensinaram durante toda a graduação, em especial os professores Carlos Manuel Taboada Rodriguez e Sérgio Fernando Mayerle.

Agradeço ao Grupo de Estudos Logísticos da UFSC (GELOG), por me propiciar um enorme desenvolvimento pessoal e profissional, além de me permitir conhecer pessoas incríveis que se tornaram amigos que quero levar para a vida.

Agradeço a todos que trabalharam comigo nas empresas em que fiz estágio, por toda o suporte e conhecimento transmitido. Agradeço em especial aos meus colegas de profissão, André Cotta de Azevedo e Álvaro Dresch Zomkowski, pelo companheirismo nessa jornada.

Agradeço à Universidade Federal de Santa Catarina por tudo o que vivi e aprendi nesse período de graduação, além de todas as oportunidades oferecidas. Agradeço aos meus amigos, em especial o Eugenio Rovaris, pela camaradagem e pelos bons momentos vividos juntos.

Por fim, agradeço a todos que de alguma forma contribuíram direta ou indiretamente para a minha pesquisa e para a minha formação. Muito obrigado!

“O aumento do conhecimento é como uma esfera dilatando-se no espaço: quanto maior a nossa compreensão, maior o nosso contato com o desconhecido.” (Blaise Pascal)

RESUMO

No setor de varejo, a previsão de demanda é uma informação crítica, podendo afetar diretamente a eficiência operacional das empresas e o nível de serviço prestado ao cliente. Entretanto, prever não é uma tarefa trivial. A crescente variedade e complexidade dos problemas de previsão resultou na necessidade de modelos preditivos cada vez mais complexos e de difícil parametrização. De forma geral, modelos mais complexos requerem mais dados e ajustes para serem treinados e, conseqüentemente, exigem um maior custo computacional. Quando se há um grande número de itens a serem previstos, o processo de previsão demanda pode demorar várias horas, ou até mesmo dias, o que pode ser prejudicial para a operação. Dessa forma, o presente estudo visa propor um modelo de previsão de demanda no varejo do setor de saúde e bem-estar que melhore a acurácia de predição e seja equilibrado em termos de custo computacional e desempenho. Para isto, o modelo desenvolvido incorpora classes de métodos preditivos como Suavização Exponencial, modelos ARIMA, SARIMA e Redes Neurais Recorrentes. As abordagens de seleção individual, na qual determina-se e aplica-se o melhor modelo preditivo em cada série individual, e seleção agregada, onde o modelo com melhor performance para a população como um todo é determinado e aplicado, foram testadas e comparadas. Técnicas de clusterização de séries temporais foram empregadas com o intuito de aprimorar o método da seleção agregada. Dessa forma, buscou-se o método preditivo a ser aplicado em cada *cluster*, ao invés de buscar um único método para toda a população. As séries temporais dos centroides obtidos foram utilizadas para eleger o método preditivo a ser utilizado em cada *cluster* e o resultado dessa abordagem foi comparado com uma árvore de regressão. O custo computacional na seleção agregada foi 96,7% menor comparado com a seleção individual. Em contrapartida, houve um aumento de 9,5% no erro médio de previsão. Em vista dos resultados, empregou-se a abordagem seleção agregada. Em comparação com o modelo corrente na empresa objeto de estudo, os resultados do modelo proposto demonstraram reduções consideráveis no erro de previsão a um custo médio de processamento do modelo de apenas 2,09 segundos por SKU. Ademais, a seleção do método preditivo para cada *cluster* através dos centroides demonstrou ser uma estimativa com grande potencial de aplicação.

Palavras-chave: Previsão de demanda. Clusterização de séries temporais. Varejo.

ABSTRACT

In the retail sector, the demand forecasting is a critical information, which can directly affect companies' operational efficiency and the provided customer level of service. However, predicting is not a trivial task. The increasing variety and complexity of the forecasting problems resulted in the need for increasingly complex predictive models with difficult parameterization. In general, more complex models need more data and adjustments to be trained and, consequently, a higher computational cost. When there are a large number of items to be foreseen, the forecasting process can take several hours, or even days, which can be detrimental to an operation. Thus, the present study aims to propose a model for forecasting retail demand in the health and wellness sector that improves the accuracy of prediction and is balanced in terms of computational cost and performance. For this, the developed model incorporates classes of predictive methods such as Exponential Smoothing, ARIMA, SARIMA models and Recurrent Neural Networks. Both individual selection approach, in which the best predictive model is determined and applied for each individual series, and aggregated selection approach, where the model with the best performance for the population as a whole is determined and applied, were tested and compared. Time series clustering techniques were employed in order to improve the aggregate selection method. Thus, the predictive method to be applied in a cluster was sought, instead of seeking a single method for the entire population. Thus, it sought the predictive method to be applied in each cluster, instead of seeking a single method for the whole population. The time series of the obtained centroids were used to choose the predictive method to be used in each cluster and the result of this approach was compared with a regression tree. The computational cost in the aggregate selection was 96.7% lower compared to the individual selection. In contrast, there was a 9.5% increase in the average prediction error. In view of the results, it was decided to use the aggregate selection approach. In comparison with the current model applied in the company under study, the results of the proposed model showed considerable reductions in the forecast error. Furthermore, the selection of the predictive method for each cluster through the centroids proved to be an estimate with great application potential.

Keywords: Demand forecast. Time series clustering. Retail.

LISTA DE FIGURAS

Figura 1 – Métodos quantitativos mais utilizados nas organizações.....	22
Figura 2 – Perfis de previsão de suavização exponencial.	28
Figura 3 –Modelo básico de um neurônio artificial.	31
Figura 4 – Camadas básicas de uma rede neural <i>feed-forward</i>	32
Figura 5 – Exemplos de ABC por diferentes regras.....	34
Figura 6 – Resultado de um agrupamento <i>k-means</i>	35
Figura 7 – Número ótimo de clusters pelo método de <i>Elbow</i>	36
Figura 8 – O resultado patológico correspondente obtido por DTW.	37
Figura 9 – Exemplo de <i>overfitting</i> e <i>underfitting</i>	40
Figura 10 – Caracterização da pesquisa.	45
Figura 11 – Etapas básicas de uma tarefa de previsão.	46
Figura 12 – Etapas da pesquisa.....	49
Figura 13 – Série histórica da quantidade de itens vendidos pela empresa por mês entre Jan/16 e Fev/20.	54
Figura 14 – Total faturado vs. quantidade vendida por cada família de produtos.	55
Figura 15 – Resultado da classificação ABC dos produtos.....	56
Figura 16 – Série temporal da quantidade mensal vendida de um produto antes e depois da normalização <i>Min-Max</i>	57
Figura 17 – Curva do somatório do erro quadrático dentro dos <i>clusters</i> em função do número de <i>clusters</i> formados.....	58
Figura 18 – Contagem de produtos agrupados em cada <i>cluster</i> para K igual a 10.	58
Figura 19 – Série temporal normalizada dos centroides obtidos de cada <i>cluster</i>	60
Figura 20 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 1.....	61
Figura 21 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 2.....	62
Figura 22 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 3.....	63
Figura 23 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 4.....	64
Figura 24 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 5.....	65
Figura 25 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 6.....	66
Figura 26 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 7.....	67
Figura 27 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 8.....	68

Figura 28 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 9.	69
Figura 29 – Centroide e amostra aleatória de 5 produtos do <i>cluster</i> 10.	70
Figura 30 – Previsões obtidas pelo programa computacional para o <i>SKU</i> 813500.	74
Figura 31 – Erros obtidos no conjunto de treino e teste para o <i>SKU</i> 813500.	75
Figura 32 – Histograma dos NRMSE obtidos na seleção individual.	78
Figura 33 – Contagem de <i>SKUs</i> em que cada família de métodos preditivos foi eleita como melhor modelo na seleção individual.	78
Figura 34 – Árvore de regressão do NRMSE em função das famílias de métodos preditivos aplicados e do cluster de cada respectivo <i>SKU</i> na seleção individual.	83
Figura 35 – Comparativo dos histogramas dos NRMSE obtidos na seleção vs. na seleção agregada.	86
Figura 36 – Fluxograma do modelo de previsão proposto.	87
Figura 37 – Comparativo do NRMSE médio obtido em cada cluster no modelo proposto vs. método corrente.	90

LISTA DE QUADROS

Quadro 1 – Padrões de séries temporais.....	24
Quadro 2 – Lista de métricas de agregação de erro comumente usadas.	42
Quadro 3 – Significado das colunas dos dados extraídos.....	53
Quadro 4 – Abordagem de escolha dos hiperparâmetros das Redes Neurais.	73
Quadro 5 – Hardware utilizado.	73
Quadro 6 – Hiperparâmetros escolhidos pelo programa para o <i>SKU 813500</i>	74
Quadro 7 – Hiperparâmetros escolhidos para a família de métodos com menor NRMSE em cada centroide.	81

LISTA DE TABELAS

Tabela 1 – Custo computacional por etapa para o <i>SKU 813500</i>	76
Tabela 2 – Resultados obtidos da aplicação da seleção individual nas 1193 séries temporais priorizadas.	77
Tabela 3 – Primeira e segunda família de métodos preditivos com menor NRMSE em cada centroide.	80
Tabela 4 – Custo computacional médio por série por etapa na seleção individual.	85
Tabela 5 – Resultados obtidos da aplicação da seleção agregada nos 1193 produtos priorizados.	86
Tabela 6 – Comparativo das estatísticas descritivas dos NRMSE obtidos no modelo proposto vs. método corrente nos produtos classificados como “A”.	88

LISTA DE ABREVIATURAS E SIGLAS

ACF – Função de Autocorrelação de Amostra
AIC – Critério de Informação de Akaike
ARIMA – Modelo de Média Móvel Integrada Autorregressiva
ARMA – Modelo de Média Móvel Autorregressiva
BIC – Critério de Informação Bayesiana
BU – *Bottom-Up*
CHAID – *CHi-squared Automatic Interaction Detector*
DTW – Sincronização Dinâmica de Tempo
LSTM – *Long Short-Term Memory*
MAE – Erro Médio Absoluto
MAPE – Erro Percentual Médio Absoluto
MLE – *Maximum Likelihood Estimation*
MSE – Erro Quadrático Médio
NRMSE – Raiz do Erro Quadrático Médio Normalizado
PACF – Função de Autocorrelação Parcial
PIB – Produto Interno Bruto
RMSE – Raiz do Erro Quadrático Médio
RNA – Redes Neurais Artificiais
RNR – Rede Neural Recorrente
SARIMA – Modelo de Média Móvel Integrada Autorregressiva Sazonal
SKU – Unidades de Manutenção de Estoque
SSE – Soma dos Erros Quadráticos
TD – *Top-Down*

SUMÁRIO

1	INTRODUÇÃO	16
1.1	CONTEXTUALIZAÇÃO	16
1.2	APRESENTAÇÃO DA EMPRESA OBJETO DE ESTUDO.....	17
1.3	OBJETIVO GERAL	18
1.4	OBJETIVOS ESPECÍFICOS	18
1.5	JUSTIFICATIVA	19
1.6	DELIMITAÇÃO DO TRABALHO	19
2	FUNDAMENTAÇÃO TEÓRICA.....	21
2.1	PREVISÃO DE DEMANDA	21
2.2	MÉTODOS DE SÉRIES TEMPORAIS.....	23
2.3	SUAVIZAÇÃO EXPONENCIAL	24
2.3.1	Suavização exponencial simples	25
2.3.2	Suavização exponencial dupla	26
2.3.3	Suavização exponencial tripla	26
2.4	MODELOS ARIMA.....	28
2.5	REDES NEURAIS ARTIFICIAIS	30
2.5.1	Redes neurais recorrentes	32
2.6	CLASSIFICAÇÃO ABC.....	33
2.7	CLUSTERIZAÇÃO	34
2.7.1	Clusterização de séries temporais	36
2.8	NORMALIZAÇÃO	38
2.9	VALIDAÇÃO CRUZADA	39
2.10	MÉTRICAS DE AVALIAÇÃO DE ERRO	40
2.11	ÁRVORE DE DECISÃO	43
3	PROCEDIMENTOS METODOLÓGICOS.....	45
3.1	CARACTERIZAÇÃO DA PESQUISA	45

3.2	ETAPAS BÁSICAS DE UMA TAREFA DE PREVISÃO	46
3.3	ETAPAS DA PESQUISA	48
4	DESENVOLVIMENTO DO MODELO.....	51
4.1	REQUISITOS DO MODELO	51
4.2	COLETA E TRATAMENTO DOS DADOS.....	52
4.3	CLASSIFICAÇÃO ABC.....	55
4.4	CLUSTERIZAÇÃO DE SÉRIES TEMPORAIS	56
4.5	ESCOLHA E IMPLEMENTAÇÃO DOS MÉTODOS PREDITIVOS	70
4.5.1	Suavização Exponencial	71
4.5.2	Modelos ARIMA e SARIMA.....	71
4.5.3	Redes Neurais Recorrentes	71
4.6	CONCEPÇÃO DAS PREVISÕES E DEFINIÇÃO DO MÉTODO DE SELEÇÃO	73
4.7	COMPARAÇÃO COM O MÉTODO ATUAL DA EMPRESA	88
5	CONCLUSÃO.....	92
5.1	ATENDIMENTO DOS OBJETIVOS DA PESQUISA.....	92
5.2	LIMITAÇÕES DO TRABALHO E OPORTUNIDADES E SUGESTÕES PARA FUTURAS PESQUISAS.....	93
6	REFERÊNCIAS.....	95

1 INTRODUÇÃO

Este capítulo tem como objetivo contextualizar o tema abordado no trabalho, justificar a pesquisa, definir seus objetivos e apresentar a delimitação do mesmo.

1.1 CONTEXTUALIZAÇÃO

A história do varejo no Brasil tem início a partir das primeiras formações populacionais nas vilas litorâneas. Inicialmente limitado ao oferecimento de itens básicos de subsistência, o varejo estabelece-se como um setor de grande importância a datar do século XIX, motivado pela expansão dos mercados produtor e consumidor (VAROTTO, 2006).

Após o final da Segunda Guerra, com o advento da globalização e o impacto das tecnologias de informação e comunicação, as transformações no varejo brasileiro se intensificaram e deram origem a um mercado cada vez mais dinâmico e competitivo (TAMASHIRO; MERLO; DA SILVEIRA, 2011). Nesse contexto, as empresas varejistas precisam buscar sempre estratégias e investimentos para alcançar novos clientes, fidelizar clientes antigos, aumentar vendas e reduzir custos (BORGES, 2001).

Em qualquer ambiente altamente competitivo, os níveis de estoque e serviço são sempre uma preocupação e um fator competitivo importante. No setor de varejo, a disponibilidade do produto é geralmente considerada um importante indicador de qualidade (SALAM; PANAHI FAR; BYRNE, 2016). Níveis de serviço baixos podem resultar em perda de clientes e vendas, enquanto, por outro lado, estoques excessivos resultam em custos desnecessários devido à realização de grandes estoques (HBNER; KUHN; STERNBECK, 2013; SALAM; PANAHI FAR; BYRNE, 2016).

O principal desafio no gerenciamento de estoques no varejo é equilibrar a reposição e a demanda, ou seja, fornecer itens na prateleira justificados por uma demanda futura do comprador (EHRENTHAL; HONHON; VAN WOENSEL, 2014). Por conseguinte, a previsão de demanda é uma parte integral dos sistemas de gerenciamento de estoque. As previsões atualizadas periodicamente são utilizadas para determinar os níveis da meta de estoque e os tamanhos dos pedidos necessários para atingir esses níveis (DZIUBA; KADŁUBEK; INGALDI, 2017; LEE; PALANIAPPAN, 2014).

Prever envolve fazer estimativas sobre uma questão ou problema desconhecido (SCHOEMAKER, 2004). As previsões possuem valor apenas se forem precisas e em tempo oportuno (SHAHABUDDIN, 1987). No entanto, prever com precisão é uma tarefa desafiadora, posto que a mudança da demanda com o tempo pode ser afetada por um número elevado de fatores de complexa natureza e modelagem (ISLAM et al., 2020).

Essa crescente variedade e complexidade dos problemas de previsão resultou na necessidade de modelos preditivos cada vez mais complexos e de difícil parametrização (BATTY; TORRENS, 2001). De forma geral, modelos mais complexos requerem mais dados e ajustes para serem treinados. À vista disso, há um correspondente aumento do custo computacional exigido (JUSTUS et al., 2019). O custo computacional de um algoritmo é a quantidade de recursos necessários para executá-lo, que se reflete fundamentalmente nos requisitos de tempo e memória (CAO et al., 2020).

A maioria das organizações envolvidas na gestão de estoques, especialmente no setor varejista, se depara com a tomada de decisões para grandes números (100's, 1.000's, 10.000's ou 1000.000's) de itens individuais (SILVER, 1981). Como resultado, o processo completo de previsão demanda em muitas empresas pode levar várias horas, ou até mesmo dias, o que pode ser prejudicial para a operação. Entretanto, apenas empregar modelos mais simples poderia resultar em uma perda significativa na acurácia de previsão. Dessa forma, um dos desafios na previsão de vendas no varejo é estimar efetivamente a demanda futura sem impor cargas computacionais ou inferenciais muito pesadas.

1.2 APRESENTAÇÃO DA EMPRESA OBJETO DE ESTUDO

A empresa alvo de estudo desta monografia está presente há mais de 50 anos no varejo catarinense, sendo especializada no comércio de materiais médicos, ortopédicos, para fitness e mobilidade. Após passar por um período de grande expansão na última década, no qual seu faturamento anual aumentou em mais de 1100%, a outrora empresa familiar detém atualmente 8 lojas físicas, um centro de distribuição e um escritório. Um dos fatores responsáveis por esse crescimento foi a intensificação das suas operações no *e-commerce*, que representaram em 2019 cerca de 20% das suas vendas.

Um rápido crescimento organizacional, no entanto, não costuma ocorrer sem o surgimento de problemas. A falta de utilização de ferramentas de previsão de demanda

adequadas, para embasar as compras de suprimentos e os pedidos de reposição nas lojas, acarretou à empresa uma série de adversidades. Dentre os problemas relatados, pode-se citar: o excesso e a falta de produtos nas lojas, o atraso e cancelamento de entregas no *e-commerce* e gastos elevados com recursos humanos na gestão e controle de estoques.

Um dos principais fatores que dificulta o planejamento da demanda na empresa é o grande número de itens individuais comercializados. Conforme dados da varejista, foram comercializados 8.844 produtos distintos em 2019. Para embasar as decisões é utilizada atualmente a técnica de média móvel com janela de tempo fixa de 6 meses. Uma das vantagens dessa técnica é que ela é simples e de fácil aplicação, porém é adequada apenas para dados univariados, sem tendência ou sazonalidade. Logo, a fim de melhorar a acurácia de predição, é primordial empregar técnicas mais sofisticadas.

1.3 OBJETIVO GERAL

O objetivo deste trabalho consiste em propor um modelo de previsão de demanda no varejo do setor de saúde e bem-estar que melhore a acurácia de predição e seja equilibrado em termos de custo computacional e desempenho.

1.4 OBJETIVOS ESPECÍFICOS

Os objetivos específicos que precisam ser atendidos para que se alcance o objetivo geral são:

- I. Revisar a literatura, com o propósito de identificar as “melhores práticas” em previsão de demanda;
- II. Coletar e tratar os dados necessários para a concepção das previsões;
- III. Classificar os produtos, visando priorizar aqueles com maior participação no faturamento da empresa;
- IV. Definir o conjunto de métodos preditivos a ser aplicado e implementá-lo em um programa computacional;
- V. Selecionar o método preditivo a ser utilizado em cada produto e comparar as suas previsões com o modelo atual da empresa.

1.5 JUSTIFICATIVA

Os sistemas de informação estão em evolução contínua e os seus impactos na sociedade são cada vez mais expressivos. Na última década, a utilização de tecnologias envolvendo o aprendizado de máquina e a inteligência artificial se popularizaram. Como consequência, a disponibilidade de informações de alta qualidade proporcionou às organizações melhores decisões de negócios e cursos de ação mais inteligentes (AL-MAMARY; SHAMSUDDIN; ABDUL HAMID, 2013).

No entanto, sistemas informacionais mais sofisticados são uma realidade observada majoritariamente nas grandes empresas (PADILHA; MARINS, 2005). Conforme Stachewski (2020), existem cerca de 1,3 milhão de micro, pequenas e médias empresas do varejo no Brasil. Ainda conforme a autora, essas empresas representam mais de 96% dos negócios do setor e empregam quase 5 milhões de pessoas. Assim sendo, uma parcela considerável das empresas do varejo brasileiro ainda não usufrui de tecnologias adequadas para uma previsão de demanda eficaz.

Dessa forma, o modelo proposto neste trabalho justifica-se pela sua contribuição sobre a aplicação de conceitos da Engenharia de Produção ao setor de varejo, cuja grande parte das empresas ainda não emprega técnicas mais sofisticadas para prever a demanda.

1.6 DELIMITAÇÃO DO TRABALHO

O presente trabalho propõe-se a desenvolver um modelo de gestão das práticas de previsão da demanda, envolvendo a aplicação de técnicas de previsão, tendo como objetivo principal aumentar a acurácia de predição da empresa objeto de estudo desta monografia. A primeira limitação a ser citada é o número restrito de empresas e setores em que o modelo foi desenvolvido e aplicado.

Outra limitação está relacionada ao número limitado de métodos de previsão empregados. Para facilitar a generalização e aplicação do modelo proposto, optou-se por não englobar métodos causais de previsão de demanda. Há também limitações relacionadas à estratégia de escolha dos hiperparâmetros, ao desenvolvimento do programa computacional, como também limitações associadas com a abordagem de clusterização de séries temporais adotada.

O modelo foi concebido para ser aplicado em empresas de médio e pequeno porte que possuem registrados os dados de histórico de vendas de seus produtos e possuem sistemas de informações básicos. Para empresas de grande porte, com sistemas de informações avançados e mais automatizados, outras soluções de gestão das práticas de previsão da demanda seriam recomendadas. Por fim, ressalta-se que não faz parte do escopo o estudo da integração do modelo proposto com sistemas informacionais existentes.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo visa trazer uma fundamentação teórica sobre previsão de demanda, métodos de séries temporais, suavização exponencial, modelos ARIMA, SARIMA, redes neurais artificiais, classificação ABC, clusterização de séries temporais, normalização, validação cruzada e métricas de avaliação de erros. Dessa forma, objetiva-se expor a teoria relacionada à pesquisa e contextualizá-la em meio à literatura.

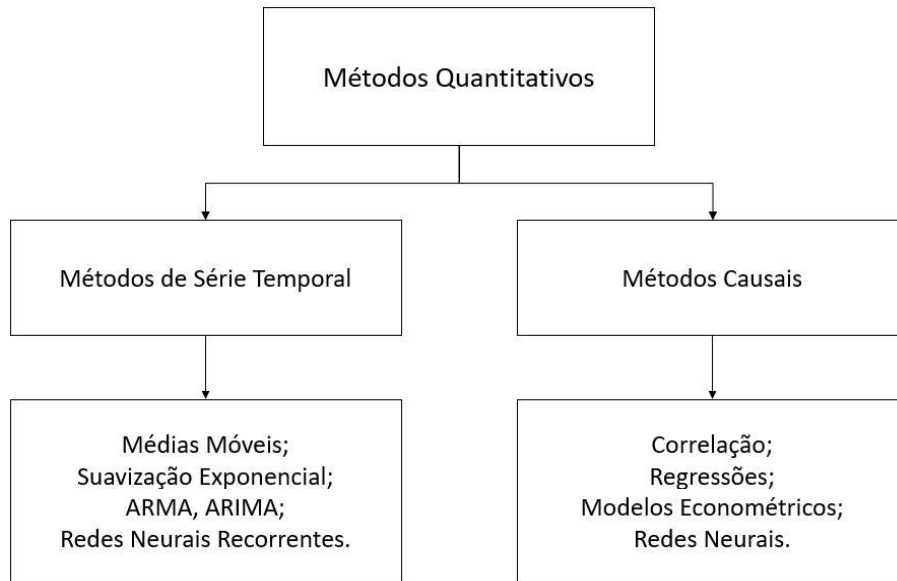
2.1 PREVISÃO DE DEMANDA

Segundo Moran *et al.* (2016), previsão é a capacidade de estimar o que acontecerá no futuro com base na análise de dados passados e atuais. Já de acordo com Kaur *et al.* (2014), previsão é o processo de fazer declarações sobre eventos cujo resultado real ainda não foi observado. Em complemento ao tema, o mesmo autor comenta que risco e incerteza são centrais para qualquer previsão; sendo considerado uma boa prática indicar o grau de incerteza associado às previsões.

A literatura científica sobre previsão de demanda tem tradicionalmente classificado os vários métodos e ferramentas em duas famílias: abordagens quantitativas e qualitativas (CANIATO; KALCHSCHMIDT; RONCHI, 2011). Os métodos quantitativos de previsão são baseados em modelos matemáticos e utilizam-se de dados para realizar as projeções. Em contrapartida, os métodos qualitativos são baseados no julgamento e intuição de especialistas (MILESKI JUNIOR, 2007). Em termos gerais, os métodos quantitativos tendem a ter um desempenho melhor do que a previsão qualitativa (CHU, 1998).

Na abordagem quantitativa, uma vez que os dados foram capturados para a série temporal a ser prevista, a próxima etapa do analista é selecionar um modelo para previsão (KALEKAR, 2004). Os métodos quantitativos de previsão podem ser subdivididos em dois grupos: métodos causais e métodos de série temporal (VERRUCK; BAMPI; MILAN, 2009). Os métodos de séries temporais usam dados históricos como base para estimar resultados futuros. Por outro lado, os métodos causais buscam descrever a demanda como função de outras variáveis no ambiente, tendo como exemplo o PIB, inflação, clima, entre outros (MILESKI JUNIOR, 2007). A Figura 1 apresenta alguns dos métodos quantitativos mais utilizados.

Figura 1 – Métodos quantitativos mais utilizados nas organizações.



Fonte: Adaptado de Mileski Junior (2007).

A seleção do método de previsão não é uma tarefa trivial. De acordo com Browne (2000), cada modelo que provavelmente será considerado estará errado até certo ponto, o que significa que buscar o modelo correto é buscar uma impossibilidade. Ainda segundo o autor, em vez de buscar um modelo correto inexistente, deve-se buscar um modelo que seja uma aproximação plausível da realidade e transmita uma visão adequada para o tomador de decisão. Em complemento ao tema, Yang e Chang (2020) afirmam que embora novos métodos venham sendo desenvolvidos, cada um com suas próprias vantagens e limitações, inexistente um método que seja o "melhor" em todas as situações.

Segundo Fildes (1989), ao fazerem previsões para várias séries temporais, os analistas geralmente tentam melhorar a precisão da previsão selecionando o modelo mais apropriado dentro de um conjunto de alternativas. A principal vantagem dessa abordagem é a preocupação com a escolha de um modelo a partir de um conjunto de modelos concorrentes, em vez de decidir se um único modelo é bom ou não. Ainda conforme o autor, uma organização pode basear empiricamente sua seleção em qualquer uma das duas abordagens distintas:

- I. Seleção individual: onde cada série temporal é analisada, o modelo com melhor performance para aquela série é determinado e as previsões são feitas utilizando os modelos selecionados para cada série individual.

- II. Seleção agregada: onde a população inteira (ou uma amostra aleatória) é analisada, o modelo com melhor performance para a população (ou amostra) como um todo é determinado e as previsões são feitas utilizando o modelo selecionado para aquela população.

Para Fildes e Petropoulos (2015), a primeira abordagem é intuitivamente atraente e pode resultar em ganhos substanciais. Entretanto, a seleção individual torna-se um problema complicado e os previsores necessitam equilibrar os ganhos potenciais na precisão das previsões e da complexidade adicional introduzida (KANG et al., 2019). A este respeito, alguns autores propõem a segmentação da população em clusters, visando identificar e agrupar séries temporais com padrões semelhantes e, assim, obter melhores resultados na seleção agregada (MAHARAJ, 2000; VAN WIJK; VAN SELOW, 1999). Essa estratégia pressupõe que o melhor modelo para uma dada série temporal possui uma alta probabilidade de ser um bom modelo para as demais séries similares, uma vez que possuem perfis de demanda semelhantes.

Muitas organizações julgam ser necessário prever itens individuais que constituem uma classificação de família ou grupo (SCHWARZKOPF; TERSINE; MORRIS, 1988). Segundo Widiarta *et al.* (2009), duas abordagens gerais foram sugeridas para desenvolver previsões para modelos ou itens individuais. Elas podem ser rotuladas como estratégias *top-down* (TD) e *bottom-up* (BU). A estratégia BU prevê itens individuais separadamente, sem colocar nenhuma restrição geral no agregado. Por sua vez, a estratégia TD prevê um total agregado e o distribui para itens individuais ou famílias proporcionalmente. De acordo com Dangerfield e Morris (1992), a estratégia BU produz previsões mais precisas na maioria das situações.

2.2 MÉTODOS DE SÉRIES TEMPORAIS

Conforme descrito por Falk *et al.* (2006), séries temporais podem ser definidas como uma sequência de observações que são organizadas de acordo com a ordem cronológica de seu desfecho. Para Cordeiro e Neves (2012), uma série temporal surge sempre que algo é observado ao longo do tempo. Esta observação pode ser feita quer em tempo contínuo, quer discreto.

De acordo com Kalekar (2004), os métodos de séries temporais pressupõem que uma série temporal é uma combinação de um padrão e algum erro aleatório. Ainda de acordo com o autor, o objetivo é isolar os padrões subjacentes e quantificar até que ponto cada componente influencia a forma dos dados observados. Se um ou mais componentes individuais de uma série

temporal forem isolados e identificados, uma previsão pode projetar o padrão subjacente no futuro (BEE DAGUM; BIANCONCINI, 2016). Tradicionalmente cinco tipos de padrões podem ser distinguidos em uma série temporal, conforme apresenta o Quadro 1.

Quadro 1 – Padrões de séries temporais.

Padrão	Descrição
Horizontal	Existe quando os valores dos dados flutuam horizontalmente em torno de uma média constante. Um produto cujas vendas não aumentam ou diminuem com o tempo seria desse tipo.
Tendência	Existe quando há um aumento ou diminuição de longo prazo nos dados. Tal comportamento não precisa ser linear
Sazonal	Ocorre quando uma série temporal é afetada por fatores sazonais (por exemplo, o trimestre do ano, o mês ou o dia da semana). A sazonalidade é sempre de uma frequência fixa e conhecida.
Cíclico	Existe quando os dados exibem aumentos e quedas cíclicas que não são de um período fixo. A duração dessas flutuações é geralmente mais longa do que a sazonalidade.
Aleatoriedade	Ocorre quando a série temporal apresenta flutuações de curto prazo que não são sistemáticas nem previsíveis.

Fonte: Adaptado de Makridakis *et al.* (1998).

A identificação desses padrões em séries temporais é muito importante sempre que se deseja prever valores futuros. Um modelo de série temporal eficaz deve ser suficientemente flexível para capturar essas características principais sem impor cargas computacionais ou inferenciais muito pesadas (CORDEIRO; NEVES, 2012). Diversos métodos de previsão de séries temporais estão disponíveis, como o método de médias móveis, regressão linear com o tempo, suavização exponencial, modelos ARIMA, entre outros (KALEKAR, 2004). Este trabalho concentra-se nas técnicas de suavização exponencial, modelos ARIMA, SARIMA e redes neurais recorrentes.

2.3 SUAVIZAÇÃO EXPONENCIAL

Os métodos de suavização exponencial são uma família de modelos de previsão que usam médias ponderadas de observações anteriores para prever novos valores

(OSTERTAGOVÁ; OSTERTAG, 2013). Segundo Valles (2018), a suavização exponencial atribui pesos exponencialmente decrescentes conforme a observação envelhece. Em outras palavras, as observações recentes recebem um peso relativamente maior na previsão do que as observações mais antigas.

Os métodos de suavização exponencial são um dos métodos de previsão mais amplamente utilizados (OSTERTAGOVÁ; OSTERTAG, 2013). Possivelmente a razão mais importante para a popularidade da suavização exponencial seja a surpreendente precisão que pode ser obtida com mínimo esforço na identificação do modelo (GARDNER, 1985). Conforme Kalekar (2004), existem três tipos principais de métodos de suavização exponencial para previsão de séries temporais: a suavização exponencial simples, a suavização exponencial dupla e a suavização exponencial tripla.

2.3.1 Suavização exponencial simples

A técnica mais simples deste tipo, a suavização exponencial simples (SES), baseia-se na premissa de que os dados flutuam em torno de uma média razoavelmente estável, sem tendência clara ou padrão sazonal (HYNDMAN; ATHANASOPOULOS, 2018). Formalmente, a equação de suavização exponencial simples assume a forma de:

$$y_{t+h} = l_t \tag{1}$$

$$l_t = \alpha y_t + (1 - \alpha)l_{t-1} \tag{2}$$

Onde y_t é o valor real da série conhecido no momento t ; y_{t+h} é o valor de previsão da variável Y no momento $t + h$; l_t é o nível suavizado da série no tempo t ; l_{t-1} é o valor suavizado anterior; α é a constante de suavização para o nível da série (BROWN; MEYER, 1961). Quando aplicado recursivamente a cada observação sucessiva na série, cada novo valor suavizado (previsão) é calculado como a média ponderada da observação atual e da observação suavizada anterior; a observação suavizada anterior foi calculada por sua vez a partir do valor observado anterior e do valor suavizado antes da observação anterior, e assim por diante (KALEKAR, 2004).

Para iniciar o algoritmo, precisa-se de um valor suavizado inicial, um valor real e uma constante de suavização. Visto que l_1 não é conhecido, podemos defini-lo como y_1 . Outra possibilidade seria utilizar a média das primeiras quatro ou cinco observações (OSTERTAGOVÁ; OSTERTAG, 2013). α é geralmente definido com um valor entre 0 e 1. Valores grandes significam que o modelo se atenta principalmente às observações mais recentes, enquanto valores menores significam que o modelo considera mais o histórico das observações ao fazer uma previsão (HYNDMAN; ATHANASOPOULOS, 2018).

2.3.2 Suavização exponencial dupla

A suavização exponencial dupla, também conhecida como modelo de previsão linear de Holt, é uma extensão da suavização exponencial simples que adiciona explicitamente suporte para tendências na série temporal (HANSUN, 2016). Segundo Kalekar (2004), este método envolve uma equação de previsão e duas equações de suavização, uma para o nível e outra para a tendência. O nível é uma estimativa suavizada do valor dos dados no final de cada período. Já a tendência é uma estimativa suavizada do crescimento médio no final de cada período. A fórmula específica para suavização exponencial dupla é:

$$y_{t+h} = l_t + h \cdot b_t \quad (3)$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (4)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (5)$$

Na qual b_t é a tendência aditiva suavizada no final do período t ; β é a constante de suavização para a tendência, $0 < \beta < 1$; b_{t-1} é a tendência suavizada anterior (HOLT, 2004). Existem vários métodos para a escolha dos valores iniciais para l_t e b_t . Geralmente l_1 é definido como y_1 e b_1 costuma ser definido como $y_2 - y_1$ (KALEKAR, 2004).

2.3.3 Suavização exponencial tripla

A suavização exponencial tripla é uma extensão da suavização exponencial dupla que adiciona explicitamente suporte para sazonalidade à série temporal univariada. Esse método

ocasionalmente é chamado de método de previsão de Holt-Winters, em homenagem a dois colaboradores do método: Charles Holt e Peter Winters (HYNDMAN; ATHANASOPOULOS, 2018). Existem duas variações desse método que diferem na natureza do componente sazonal. O método aditivo é preferido quando as variações sazonais são aproximadamente constantes ao longo da série, enquanto o método multiplicativo é preferido quando as variações sazonais mudam proporcionalmente ao nível da série (KALEKAR, 2004). As fórmulas do método aditivo de Holt-Winters são dadas por:

$$y_{t+h} = l_t + h \cdot b_t + s_{t+h-p} \quad (6)$$

$$l_t = \alpha(y_t - s_{t-p}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (7)$$

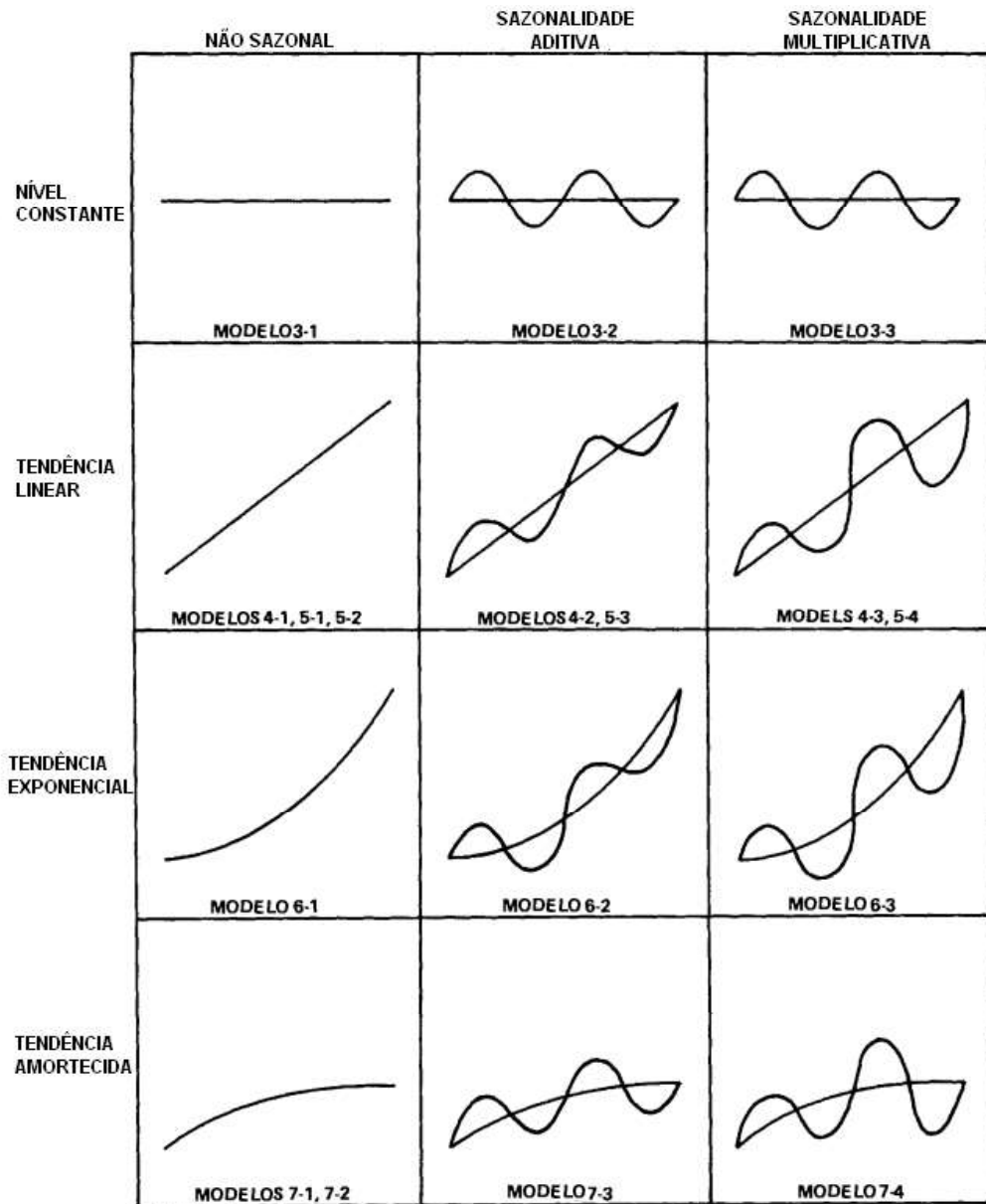
$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (8)$$

$$s_t = \gamma(y_t - l_t) + (1 - \gamma)s_{t-p} \quad (9)$$

Onde s_t é o fator sazonal aditivo suavizado no final do período t ; p é o número de períodos no ciclo sazonal; γ é a constante de suavização para os índices sazonais, $0 < \gamma < 1$; s_{t-p} é o fator sazonal aditivo suavizado no final do período $t - p$ (GARDNER, 1985). Como regra geral, um mínimo de duas temporadas completas (ou períodos de $2p$) de dados históricos é necessário para inicializar um conjunto de fatores sazonais (KALEKAR, 2004). Em Makridakid *et al.* (1998) diversas alternativas para inicialização do método Holt-Winters são apresentadas.

Uma grande série de modelos subjacentes à suavização exponencial foram propostos desde 1960, onde diferentes abordagens para modelar a tendência e a sazonalidade foram criadas. Gardner (2006) em seu trabalho apresenta 12 diferentes tipos de perfis de previsão de suavização exponencial e suas respectivas equações, onde cada perfil é definido por uma combinação de um padrão de tendência (linear, exponencial, amortecida ou sem tendência) e um padrão de sazonalidade (aditiva, multiplicativa ou não sazonal). A Figura 2 ilustra os diferentes perfis obtidos.

Figura 2 – Perfis de previsão de suavização exponencial.



Fonte: Adaptado de Gardner (1985).

2.4 MODELOS ARIMA

Introduzido em 1970 por Box e Jenkins, o modelo de média móvel integrada autorregressiva (ARIMA) é uma generalização do modelo de média móvel autorregressiva (ARMA) (LI; LI, 2017). De acordo com Al-Shiab (2006), ao contrário dos modelos de regressão, em que y_t é explicado por k variáveis explicativas, a metodologia de Box-Jenkins permite que y_t seja explicado por valores passados ou defasados do próprio y e termos de erro estocástico. Conforme Hyndman e Athanasopoulos (2018), os modelos ARIMA e a suavização

exponencial são as duas abordagens mais amplamente utilizadas para a previsão de séries temporais.

A parte autorregressiva do ARIMA indica que a variável em evolução de interesse é regredida linearmente em seus próprios valores defasados, ou seja, usa suas próprias defasagens como preditores (YOOSEFDOOST *et al.*, 2017). A parte de média móvel indica que o erro de regressão é na verdade uma combinação linear de termos de erro cujos valores ocorreram simultaneamente e em vários momentos no passado (BAKAR; ROSBI, 2017). Finalmente, caso a série não seja estacionária, a parte integrada do ARIMA indica que os valores dos dados foram substituídos pela diferença entre seus valores e os valores anteriores, sendo que este processo diferenciador pode ter sido realizado mais de uma vez (YANG *et al.*, 2020).

De acordo com Wu *et al.* (2019), um processo estacionário é um processo estocástico cuja distribuição de probabilidade conjunta incondicional não muda quando deslocada no tempo. Consequentemente, parâmetros estatísticos como média e variância também não mudam com o tempo. O conceito de estacionariedade é fundamental para a aplicação de um modelo ARIMA porque o uso de valores passados não estacionários da variável independente pode causar problemas de multicolinearidade (MCGOUGH; TSOLACOS, 1995; WHITE, 1987).

Os modelos ARIMA não sazonais são geralmente denotados $ARIMA(p, d, q)$, onde os parâmetros p , d e q são inteiros não negativos, p é a ordem (número de defasagens de tempo) do modelo autorregressivo, d é o grau de diferenciação (o número de vezes que os dados tiveram valores passados subtraídos), e q é a ordem do modelo de média móvel (BAKAR; ROSBI, 2017). O modelo completo pode ser escrito como:

$$y'_t - \phi_1 y'_{t-1} - \dots - \phi_p y'_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (10)$$

Onde y'_t é a série diferenciada, ϕ_p são os parâmetros da parte autorregressiva do modelo, θ_q são os parâmetros da parte média móvel e ε_t são os termos de erro (HYNDMAN; ATHANASOPOULOS, 2018). Diversas técnicas podem ser empregadas para determinar os parâmetros p , d e q , como os gráficos da função de autocorrelação de amostra (ACF) e função de autocorrelação parcial (PACF), Grid Search ou métodos alternativos como o Critério de informação de Akaike (AIC), Critério de informação bayesiana (BIC), etc. (HYNDMAN; ATHANASOPOULOS, 2018).

Na prática, os dados de série temporal geralmente exibem padrões periódicos e não estacionários, como sazonalidade. Dessa forma, os modelos ARIMA precisam ser generalizados para lidar com os diferentes perfis de demanda (CHEN; CHANG; CHANG, 2009).

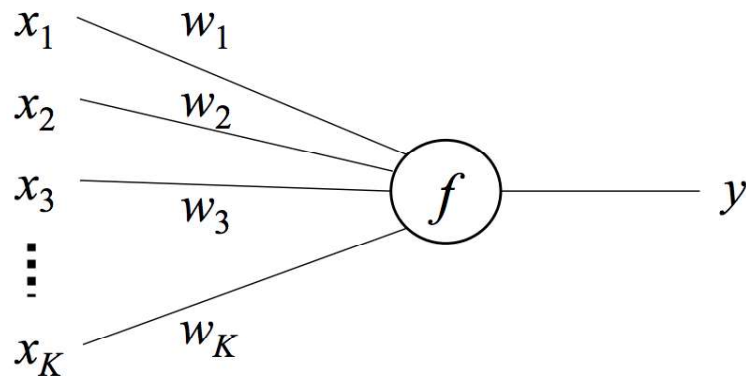
O modelo ARIMA usado para acomodar a sazonalidade é chamado ARIMA Sazonal (SARIMA). No modelo SARIMA, uma diferenciação adicional, chamada de diferenciação sazonal, é necessária a fim de tratar o efeito sazonal (SHAADAN et al., 2019). De acordo com Permanasari *et. al* (2013), o modelo SARIMA é denotado como $SARIMA(p, d, q)(P, D, Q)_m$, onde p , d e q denotam as ordens não sazonais e P, D e Q indicam as ordens sazonais do processo. Todos esses parâmetros são inteiros não negativos e a letra subscrita ' m ' representa a duração do período sazonal. A teoria do modelo SARIMA e suas equações é melhor abordada em Brockwell e Davis (1991).

2.5 REDES NEURAIS ARTIFICIAIS

As redes neurais artificiais (RNA), ou simplesmente redes neurais, são uma família de métodos de aprendizado de máquina desenvolvidas a partir da ideia de simular o cérebro humano (ZOU; HAN; SO, 2008). As RNAs têm sido usadas para criar relações gerais entre variáveis em um determinado problema (ANASARI, 2019). Devido à sua capacidade de reproduzir e modelar processos não lineares, as RNAs encontraram aplicações em muitas disciplinas (MITRAN; MEENA; CHAKRABORTY, 2021). Conforme Oancea e Ciucu (2014), as redes neurais artificiais têm sido usadas com sucesso para previsões.

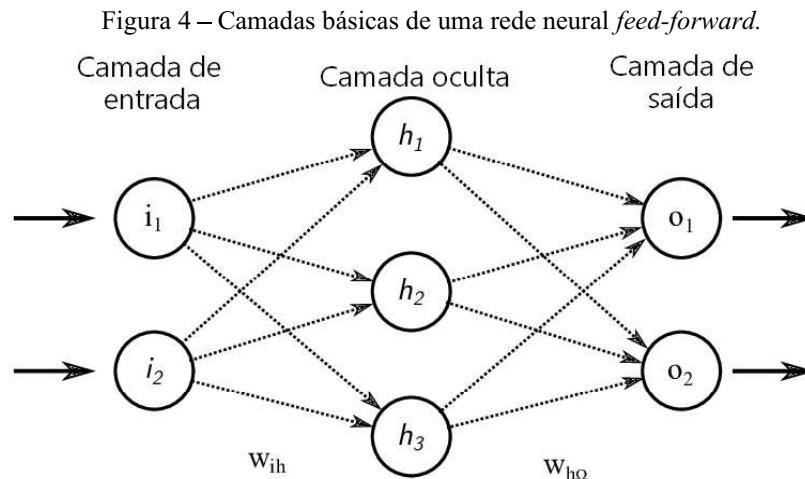
Uma RNA é baseada em uma coleção de unidades conectadas chamadas neurônios artificiais (PAUL *et al.*, 2017). O modelo básico de um neurônio artificial é mostrado na Figura 3. Os neurônios artificiais imitam neurônios biológicos, que recebem entrada, combinam a entrada com seu estado interno e produzem saída usando uma função de saída (MITRAN; MEENA; CHAKRABORTY, 2021). Um determinado neurônio pode ter várias conexões de entrada e saída (MONIKA; GUPTA; SUSHIL KUMAR, 2020). Normalmente, cada entrada recebe um peso, que representa sua importância relativa, e a soma é passada por uma função não linear conhecida como função de ativação ou função de transferência (CARVALHO, 2010). As funções de transferência geralmente têm uma forma de sigmoide, mas também podem assumir a forma de outras funções não lineares, funções lineares por partes ou funções de *step* (ARZAMASTSEV; TROITZSCH, 2011).

Figura 3 –Modelo básico de um neurônio artificial.



Fonte: Adaptado de Zou *et al.* (2008).

Os neurônios costumam ser organizados em camadas. Com base no padrão de conexão entre os neurônios (arquitetura), as RNAs podem ser agrupadas em duas categorias: redes *feed-forward* e redes recorrentes (JAIN; MAO; MOHIUDDIN, 1996). Em uma rede *feed-forward*, os neurônios de uma camada conectam-se apenas aos neurônios das camadas imediatamente anteriores e posteriores e o fluxo de informações ocorre apenas na direção para frente (RACCA, 2019). Logo, não há ciclos ou *loops* na rede. A Figura 4 exemplifica a arquitetura geral de uma rede neural *feed-forward*. A camada de neurônios que recebe dados externos é chamada de camada de entrada, enquanto a camada que produz a saída é chamada de camada de saída. Entre a camada de entrada e saída, pode haver camadas ocultas (ZOU; HAN; SO, 2008).



Fonte: Jahr *et al.* (2015).

Conforme Jahr *et al.* (2015), as RNAs se adaptam a diferentes aplicações por meio do aprendizado. O progresso na aprendizagem é alcançado ajustando-se os pesos das sinapses até que um conjunto de dados de entrada resulte nos valores de saída desejados. Ainda segundo os autores, as RNAs podem ser personalizadas para vários objetivos usando diferentes topologias, funções de neurônios e estratégias de aprendizagem.

2.5.1 Redes neurais recorrentes

A rede neural recorrente (RNR) é semelhante à rede *feed-forward*, tendo como principal diferença a ocorrência de *loops* devido às conexões de *feedback* (GENÇAY; LIU, 1997). Pode-se também defini-la como uma rede onde a conexão entre os nós formam um grafo direcionado (LIU; LIU, 2020). Isso permite que ela exiba um comportamento dinâmico temporal. De acordo com Jain *et al.* (1996), as redes *feed-forward* são sem memória no sentido de que sua resposta a uma entrada é independente do estado da rede anterior. As redes recorrentes, por outro lado, podem responder ao mesmo padrão de entrada de forma diferente em momentos diferentes, dependendo dos padrões que foram apresentados anteriormente como entradas (DEMATOS *et al.*, 1996). Assim, a sequência dos padrões é tão importante quanto o próprio padrão de entrada. Conforme Kim (2020), a rede neural recorrente é conhecida como um tipo de RNA naturalmente adequada para dados de séries temporais, uma vez que estes dados podem ter autocorrelação ou dependência do tempo.

Segundo Lyu *et al.* (2015), em princípio, uma rede neural recorrente grande o suficiente deveria ser capaz de modelar sequências de complexidade arbitrária. Na prática,

entretanto, os mesmos autores afirmam que é difícil para a RNR armazenar informações históricas extensas sobre uma sequência. As redes de *Long Short-Term Memory* (LSTM), ou redes de memória de curto e longo prazo, são um tipo especial de rede neural recorrente projetadas especificamente para ajudar a memorizar e recuperar informações em sequência melhores do que a RNC padrão. Elas foram introduzidas por Hochreiter e Schmidhuber em 1997, e foram refinadas e popularizadas por muitas pessoas em trabalhos seguintes (LI; SUN; PANG, 2018).

De acordo com Graves (2012), a arquitetura da LSTM consiste em um conjunto de sub-redes conectadas de forma recorrente, conhecidas como blocos de memória. Ainda segundo o autor, cada bloco contém uma ou várias células de memória e três portas: porta de entrada, porta de saída e porta de esquecimento. Essas portas podem regular o estado da célula adicionando ou removendo as informações (DU et al., 2017). O portal de esquecimento tende a decidir que tipo de informação jogaremos fora do estado da célula. A porta de entrada decide os valores que precisamos atualizar e, após atualizar o estado da célula antiga para um novo, a porta de saída decide quais informações iremos produzir (LI; SUN; PANG, 2018).

Estudos empíricos conduzidos e relatados na literatura sugerem que redes LSTM podem alcançar melhores resultados para séries temporais, quando comparadas com algoritmos tradicionais (ABDOLI; MEHRARA; EBRAHIM ARDALANI, 2020; JOZEFOWICZ; ZAREMBA; SUTSKEVER, 2015; SIAMI-NAMINI; TAVAKOLI; SIAMI NAMIN, 2019).

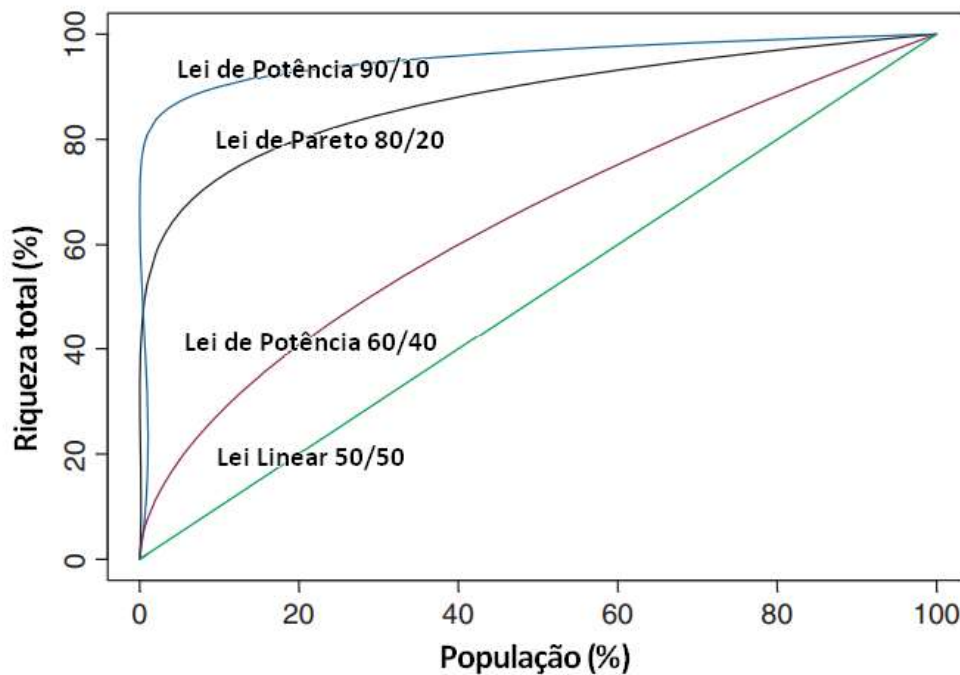
2.6 CLASSIFICAÇÃO ABC

Uma questão operacional importante envolvida na previsão de demanda é a de categorizar as unidades de manutenção de estoque (SKUs) relevantes, a fim de facilitar a tomada de decisão, ou seja, selecionar métodos apropriados de previsão e permitir que os gerentes concentrem sua atenção nas SKUs mais importantes (SYNTETOS; KEYES; BABAI, 2009).

Um grande número de SKUs torna inviável o gerenciamento de itens individualmente. Portanto, eles são comumente agrupados e políticas genéricas de previsão de demanda e controle de estoque são aplicadas a cada grupo (SILVER, 2008). O método mais comum para classificar itens de estoque é o método de classificação ABC, que assume, de acordo com o princípio de Pareto, que um pequeno número de itens é responsável por uma grande parte do

volume de faturamento, uma categoria intermediária de itens de volume de faturamento moderado e um grande número de itens de baixo faturamento ou uso (LIIV, 2006). A Figura 5 ilustra como uma análise ABC poderia parecer, onde é apresentado o percentual acumulado do faturamento total pela quantidade de produtos. De acordo com o princípio de Pareto, tipicamente 80% do faturamento é concentrado em 20% dos produtos (LIPOVETSKY, 2009).

Figura 5 – Exemplos de ABC por diferentes regras.



Fonte: Adaptado de Lipovetsky (2009)

Obviamente, a classificação ABC foi originalmente nomeada e projetada para três classes: A, B e C; com os itens A sendo considerados os mais críticos e exigindo os mais altos níveis de serviço (COX; BLACKSTONE, 2002). Os tamanhos de cada classe são determinados usando a regra comumente aplicada de que as classes A, B e C contêm, respectivamente, cerca de 20%, 30% e 50% de todas as SKUs (JACOBS; CHASE, 2014).

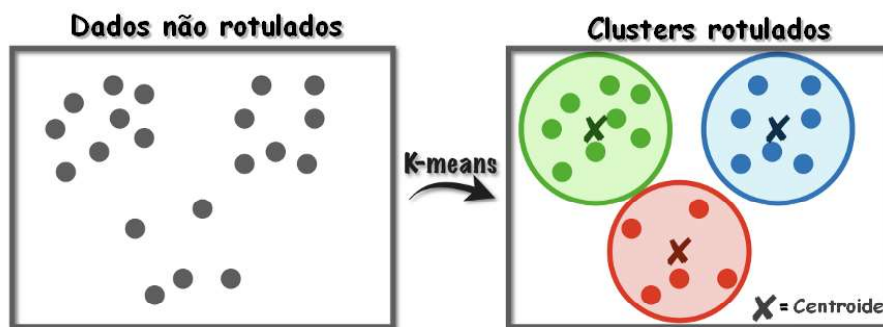
2.7 CLUSTERIZAÇÃO

Conforme Rai e Singh (2010), *clustering*, ou clusterização, é uma técnica de mineração de dados onde dados semelhantes são colocados em grupos relacionados ou homogêneos sem conhecimento avançado das definições dos grupos. Cada grupo, chamado *cluster*, consiste em

objetos que são semelhantes entre si e diferentes de objetos de outros grupos. É uma abordagem útil para análise exploratória de dados, pois identifica estruturas em um conjunto de dados não rotulado (AGHABOZORGI; SEYED SHIRKHORSHIDI; YING WAH, 2015).

De acordo com Kanungo *et al.* (2002), entre as técnicas de agrupamento que se baseiam na minimização de uma função objetivo formal, talvez a mais amplamente utilizada e estudada seja o agrupamento *k-means*. Ainda segundo o autor, dado um conjunto de n pontos de dados no espaço d -dimensional real, \mathbf{R}^d , e um inteiro k , o problema é determinar um conjunto de k pontos em \mathbf{R}^d , chamados centros, de modo a minimizar a distância quadrada média de cada ponto de dados para seu centro mais próximo. Em outras palavras, o algoritmo *k-means* identifica o número k de centróides e, em seguida, aloca cada ponto de dados para o *cluster* mais próximo, enquanto mantém os centróides o mais longe possível (GARBADE, 2018). A Figura 6 ilustra o resultado de um agrupamento *k-means*, onde $k = 3$.

Figura 6 – Resultado de um agrupamento *k-means*.

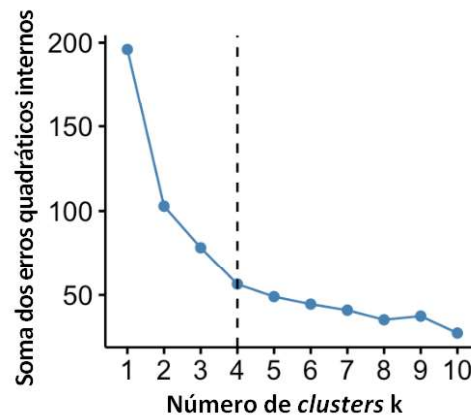


Fonte: Adaptado de Jeffares (2019).

Várias abordagens têm sido propostas para determinar o número ideal k de *clusters* para o algoritmo de agrupamento *k-means* (KODINARIYA; MAKWANA, 2013). Um dos métodos mais antigos e possivelmente o mais popular é deslealmente chamado de método de *Elbow*, ou método do cotovelo. De acordo com Yuan e Yang (2019), a ideia básica por trás desse método é que conforme o valor de k aumenta, haverá menos elementos no *cluster* e, portanto, a soma total das distâncias euclidianas de cada ponto de dados para seu centro mais próximo diminuirá. Ainda segundo o autor, essa grandeza é comumente chamada de soma dos erros quadráticos (SSE). Logo, deve-se escolher um número k de *clusters* de modo que a adição de outro cluster não forneça uma SSE muito menor (PURNIMA; ARVIND, 2014). O valor de

k pode ser melhor determinado traçando a curva K-SSE e encontrando o ponto de inflexão para baixo. A Figura 7 ilustra a escolha do número de *clusters* através do método do cotovelo.

Figura 7 – Número ótimo de clusters pelo método de *Elbow*.



Fonte: Adaptado de Kassambara (2018).

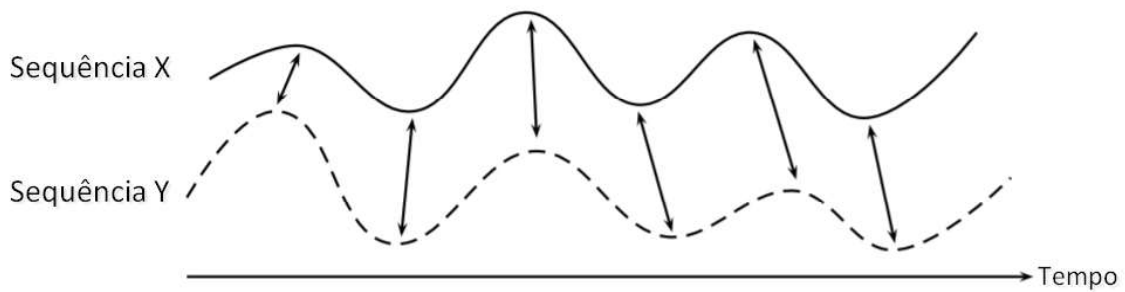
2.7.1 Clusterização de séries temporais

Um tipo especial de clusterização é a clusterização de séries temporais. Esse tipo de procedimento tem o intuito de fracionar dados de séries temporais em grupos com base na similaridade ou distância, de forma que as séries temporais no mesmo cluster sejam semelhantes (LI et al., 2020). Segundo Maharaj (2000), um dos benefícios de agrupar séries temporais é que em vez de prever cada uma das séries temporais fornecidas, a previsão pode ser realizada em um representante de cada grupo. Essa estratégia é especialmente útil se for necessário prever um grande número de séries temporais.

Agrupar séries temporais diferentes em grupos semelhantes, no entanto, é uma tarefa desafiadora porque cada ponto de dados é uma sequência ordenada (AGHABOZORGI; SEYED SHIRKHORSHIDI; YING WAH, 2015). No algoritmo *k-means*, as tradicionais métricas de distância utilizadas não são adequadas para séries temporais. Entretanto, motivados por essa observação, diversos métodos avançados têm sido propostos para medir as distâncias entre diferentes séries (ATEV; MILLER; PAPANIKOLOPOULOS, 2010). Estudos anteriores propuseram métodos de medição de distâncias populares, como a distância euclidiana, a distância de Hausdorff, o modelo de Markov oculto, a sincronização dinâmica no tempo e a subsequência comum mais longa (LI et al., 2020; LOH; MANE; SRIVASTAVA, 2011).

A sincronização dinâmica de tempo (DTW) é a técnica mais conhecida para avaliar similaridade/dissimilaridade de séries temporais em relação às suas informações de perfil (IZAKIAN; PEDRYCZ; JAMAL, 2015). Segundo Müller (2007), usando alongamento e compactação de séries temporais, o DTW determina uma combinação ótima entre duas séries temporais nos cálculos de suas diferenças. Ainda conforme o autor, nessa técnica cada ponto proveniente da primeira série temporal é comparado com qualquer ponto arbitrário da segunda série temporal. Como resultado, séries temporais com padrões semelhantes ocorridos em diferentes períodos de tempo são consideradas semelhantes. A Figura 8 ilustra o resultado patológico correspondente obtido por DTW.

Figura 8 – O resultado patológico correspondente obtido por DTW.



Fonte: Adaptado de Müller (2007).

De acordo com Sakoe e Chiba (1978), dada a série $X = (x_0, \dots, x_n)$ e a série $Y = (y_0, \dots, y_m)$, a distância DTW de X a Y é formulada como o seguinte problema de otimização:

$$DTW(x, y) = \min_{\pi} \sqrt{\sum_{(i,j) \in \pi} d(x_i, y_j)^2} \quad (11)$$

Onde $d(x_i, y_j)$ é uma medida da diferença entre x_i e y_j (ex.: $d(x_i, y_j) = |x_i - y_j|$) e $\pi = [\pi_0, \dots, \pi_K]$ é o caminho que satisfaz as seguintes propriedades:

1. É uma lista de índices pares $\pi_k = (i_k, j_k)$ com $0 \leq i_k < n$ e $0 \leq j_k < m$;
2. $\pi_0 = (0, 0)$ e $\pi_K = (n - 1, m - 1)$;
3. Para todo $k > 0$, $\pi_k = (i_k, j_k)$ é relacionado com $\pi_{k-1} = (i_{k-1}, j_{k-1})$ da seguinte forma:

- a. $i_{k-1} \leq i_k \leq i_{k-1} + 1$;

$$b. j_{k-1} \leq j_k \leq j_{k-1} + 1.$$

Para resumir a equação: DTW é calculado como a raiz quadrada da soma das distâncias quadradas entre cada elemento em X e seu ponto mais próximo em Y. É importante destacar que $DTW(X, Y) \neq DTW(Y, X)$.

Conforme Jang *et al.* (2011), o algoritmo de agrupamento *k-means* pode ser aplicado a séries temporais utilizando a distância DTW. Ainda segundo os autores, como resultado, os centroides têm uma forma média que imita a forma dos membros do cluster, independentemente de onde ocorrem mudanças temporais entre os membros.

2.8 NORMALIZAÇÃO

A normalização é uma técnica de escalonamento ou um estágio de pré-processamento onde pode-se encontrar um novo intervalo a partir de um intervalo existente (PATRO; SAHU, 2015). Segundo Shalabi *et al.* (2006), um atributo é normalizado dimensionando seus valores para que fiquem dentro de um pequeno intervalo especificado, como 0,0 a 1,0. Ainda de acordo com os autores, a normalização pode melhorar a precisão e eficiência de algoritmos de classificação que envolvem redes neurais ou medições de distância, como a classificação de vizinho mais próximo e *clustering*.

Vários tipos de técnicas de normalização estão disponíveis na literatura, tais como normalização *Min-Max*, normalização *Z-Score* e normalização de escala decimal (SARANYA; MANIKANDAN, 2013). A normalização *Min-Max*, também conhecida como padronização *MinMaxScaler*, é uma das abordagens mais amplamente reconhecidas para normalizar informações (RAJU et al., 2020). Suponha que min_a e max_a sejam os valores mínimo e máximo para o atributo A. A normalização *Min-Max* mapeia um valor v de A no intervalo $[min_a, max_a]$ para um novo intervalo $[novo_min_a, novo_max_a]$ calculando:

$$v' = \left(\frac{v - min_a}{max_a - min_a} \right) * (novo_max_a - novo_min_a) + novo_min_a \quad (12)$$

Onde v' é o novo valor no intervalo necessário (SHALABI; SHAABAN; KASASBEH, 2006). A normalização *Min-Max* executa uma transformação linear nos dados originais. Por padrão, $novo_min_a$ é escolhido como 0 e $novo_max_a$ como 1 (RAJU et al., 2020). Conforme Ogasawara *et al.* (2010), depois que os processos classificatórios ou as

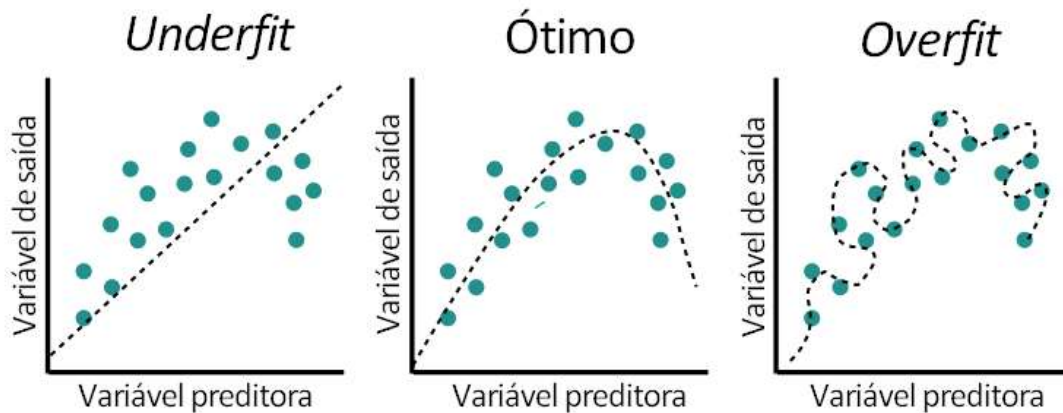
previsões são feitas, é de praxe mapear os valores de saída para os valores originais na série temporal através de um processo de desnormalização.

2.9 VALIDAÇÃO CRUZADA

De acordo com Dufour *et al.* (2014), a validação cruzada, às vezes chamada de estimativa de rotação ou teste fora da amostra, é uma técnica de validação de modelo para avaliar como os resultados de uma análise estatística serão generalizados para um conjunto de dados independente. Ainda conforme o autor, é usada principalmente em ambientes onde o objetivo é previsão, e se quer estimar quão precisamente um modelo preditivo funcionará na prática.

Na validação cruzada, é prática comum separar os dados disponíveis em duas partes, dados de treinamento e teste, onde os dados de treinamento são usados para estimar quaisquer parâmetros de um método de previsão e os dados de teste são usados para avaliar sua precisão (HYNDMAN; ATHANASOPOULOS, 2018). O objetivo é testar a capacidade do modelo de prever novos dados que não foram utilizados na estimativa do mesmo, a fim de sinalizar problemas como *overfitting* ou viés de seleção (CAWLEY; TALBOT, 2010; ORRÙ *et al.*, 2020).

Em estatística, *overfitting* é a produção de uma análise que corresponde muito próxima ou exatamente a um determinado conjunto de dados e pode, portanto, não ajustar dados adicionais ou prever observações futuras de forma confiável (ARAVANIS; ARAVANIS; PAPADOPOULOS, 2019). Por outro lado, o *underfitting* ocorre quando o modelo não consegue capturar adequadamente a estrutura subjacente dos dados (CHAOJI; RASTOGI; ROY, 2015). A Figura 9 ilustra um exemplo de *overfitting*, *underfitting* e o que seria considerado um “modelo ideal” para o conjunto de dados em questão.

Figura 9 – Exemplo de *overfitting* e *underfitting*.

Fonte: Adaptado de “Overfitting and underfitting”, [s.d.].

Segundo Kasturi (2019), um dos principais desafios com *overfitting*, e com a aprendizagem de máquina em geral, é que não se sabe de antemão quão bem um modelo irá desempenhar com novos conjuntos de dados até que seja realmente testado. Ainda conforme o autor, a maneira mais simples de determinar o *overfitting* é se o modelo desempenhar muito bem em dados de treinamento, mas muito mal em dados de teste. Isso é um sinal direto de que se está, em grande parte, com *overfitting* no modelo de previsão. Por outro lado, se o modelo tiver um desempenho ruim tanto nos dados de treinamento quanto nos dados de teste, isso pode ser por causa de *underfitting*.

Quando se lida com séries temporais, é uma prática comum reservar uma parte do final de cada série temporal para teste e usar o restante da série para treinamento (BERGMEIR; BENÍTEZ, 2012). De acordo com Hyndman e Athanasopoulos (2018), o tamanho do conjunto de teste é normalmente cerca de 20% da amostra total, embora esse valor dependa de quanto tempo é a amostra e quão adiantado deseja-se prever. Ainda conforme os autores, o conjunto de teste deve ser idealmente pelo menos tão grande quanto o horizonte de previsão máximo necessário.

2.10 MÉTRICAS DE AVALIAÇÃO DE ERRO

De acordo com Fildes e Kingsman (2011), erro de previsão é a diferença entre o valor observado e sua previsão. Ainda segundo os autores, esse erro, no entanto, combina a

aleatoriedade no processo que gera as demandas e os erros decorrentes do não uso da previsão ótima. Para Hyndman e Athanasopoulos (2018), o erro de previsão pode ser escrito como:

$$e_t = y_t - \hat{y}_t \quad (13)$$

Onde e_t é o erro de previsão no período t , y_t é a demanda real nesse período e \hat{y}_t é a sua previsão. Essa métrica pode ser usada para calcular o erro de cada combinação individual, porém quando há a necessidade de calcular a métrica em um nível agregado, os erros negativos e positivos se cancelam e uma imagem muito mais otimista do que a realidade é obtida (SINGH, 2015).

Para medir o erro agregado, pode-se agregar os erros de previsão de diferentes maneiras. O Quadro 2 apresenta uma coleção de medidas de precisão comumente utilizadas e suas equações (BOTCHKAREV, 2018; DIEBOLD; LOPEZ, 1996; FLORES, 1986). A seleção de uma medida de erro tem efeito importante nas conclusões sobre qual de um conjunto de métodos de previsão é o mais preciso (KALEKAR, 2004).

Quadro 2 – Lista de métricas de agregação de erro comumente usadas.

Nome da métrica	Fórmula
Erro médio absoluto (MAE)	$MAE = \frac{1}{T} \sum_{t=1}^T e_t $
Erro percentual médio absoluto (MAPE)	$MAPE = \frac{1}{T} \sum_{t=1}^T \left \frac{e_t}{y_t} \right * 100$
Erro quadrático médio (MSE)	$MSE = \frac{1}{T} \sum_{t=1}^T e_t^2$
Raiz do erro quadrático médio (RMSE):	$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T e_t^2}$
Raiz do erro quadrático médio normalizado (NRMSE):	$NRMSE = \frac{\sqrt{T}}{\sum_{t=1}^T y_t} \sqrt{\sum_{t=1}^T e_t^2}$

Fonte: Elaborado pelo autor (2021).

A função de erro apropriada depende da situação em questão (DIEBOLD; LOPEZ, 1996). As medidas relativas, como o MAPE, possuem a vantagem de não possuírem unidades e, portanto, são frequentemente usadas para comparar desempenhos de previsão entre conjuntos de dados diferentes (HYNDMAN; ATHANASOPOULOS, 2018). Uma das advertências no uso de medidas relativas é quando os valores reais são zero ou próximos de zero (FLORES, 1986).

Quando se comparam os métodos de previsão aplicados a uma única série temporal ou a várias séries temporais com as mesmas unidades, o MAE apresenta-se como uma medida popular porque é de fácil cálculo e entendimento (SNAPP, 2020). O MSE tende a mostrar qual o melhor método para evitar grandes erros, uma vez que atribui maior peso a esses erros devido à parte quadrada da função (THOMPSON, 1990). Esta pode ser uma consideração importante para a escolha de uma métrica. No entanto, o MSE é difícil de ser interpretado devido à arbitrariedade de sua escala (MESAROS et al., 2019). Conforme Chai e Oceanic (2015), a raiz quadrada dos erros quadrados médios produz algumas implicações interessantes para o RMSE. O RMSE representa o desvio padrão dos resíduos (erros de previsão) e pode ser interpretado como o erro médio de predição dentro da mesma escala (unidade). Para evitar a dependência

de escala no RSME, a raiz do erro quadrático médio normalizado (NRMSE) foi proposta (SHCHERBAKOV et al., 2013). A normalização do RMSE facilita a comparação entre conjuntos de dados ou modelos com escalas diferentes.

2.11 ÁRVORE DE DECISÃO

Conforme Goyal e Vohra (2012), a árvore de decisão é uma ferramenta de análise que usa um gráfico semelhante a uma árvore ou modelo de decisão e suas possíveis consequências, incluindo resultados de eventos fortuitos, custos de recursos e utilidade. Ainda segundo os autores, outro uso de árvores de decisão é como um meio descritivo para calcular probabilidades condicionais.

Os modelos de árvore em que a variável de destino pode assumir um conjunto discreto de valores são chamados de árvores de classificação; nessas estruturas de árvore, as folhas representam rótulos de classe e os ramos representam conjunções de recursos que levam a esses rótulos de classe. Árvores de decisão onde a variável de destino pode assumir valores contínuos (normalmente números reais) são chamadas de árvores de regressão (KANUNGO et al., 2002).

O algoritmo *CHi-squared Automatic Interaction Detector* (CHAID), proposto pelo estatístico Kass no final dos anos 1970, é um dos métodos mais populares de aprendizagem supervisionada com base estatística para o desenvolvimento de árvores de decisão (MILANOVIĆ; STAMENKOVIĆ, 2016). O método CHAID emprega uma abordagem semelhante à regressão *stepwise* para seleção dividida (LOH, 2014). Ele foi originalmente projetado para classificação e posteriormente estendido para regressão.

O acrônimo CHAID denota o procedimento automático e iterativo de desenvolvimento da árvore com base na estatística Qui-quadrado de Pearson e o correspondente *p-value* (MILANOVIĆ; STAMENKOVIĆ, 2016). Em outras palavras, o nó com *p-value* estatisticamente significativo é dividido, caso contrário, ele não é dividido e, portanto, considerado um nó terminal.

De forma similar a outras árvores de decisão, as vantagens do CHAID são que seu resultado é altamente visual e fácil de interpretar. Como o método usa divisões de múltiplas vias por padrão, é necessário tamanhos de amostra grandes para que o método funcione de maneira eficaz, já que com tamanhos de amostra pequenos, os grupos de respondentes podem rapidamente se tornar muito pequenos para uma análise confiável (SATYANARAYANA;

RAMALINGASWAMY; RAMADEVI, 2014). Uma vantagem importante do CHAID sobre alternativas, como regressão múltipla, é que ele é não-paramétrico.

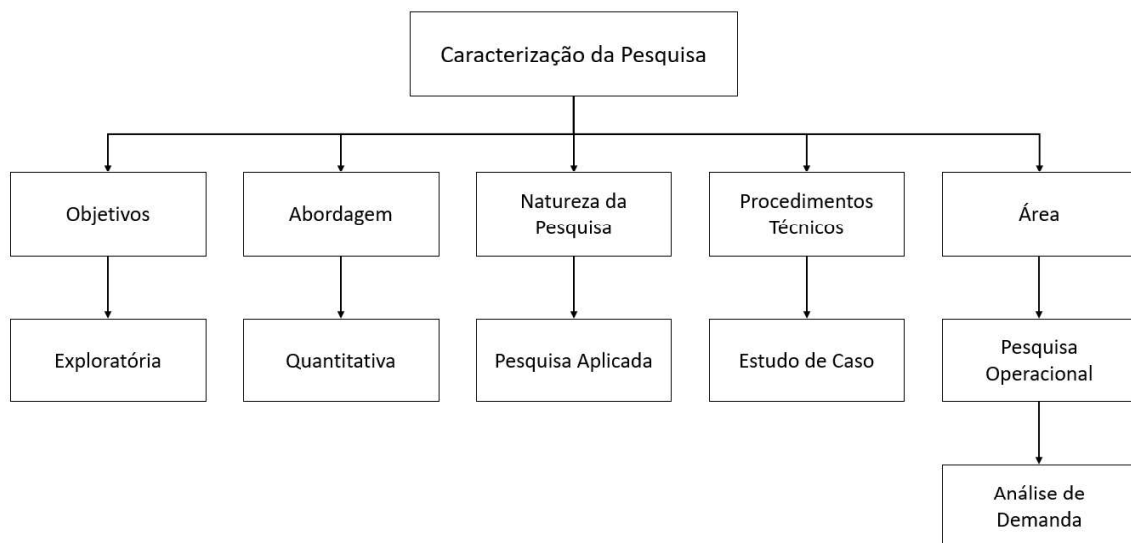
3 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta os métodos deste trabalho, englobando a caracterização e as etapas da pesquisa.

3.1 CARACTERIZAÇÃO DA PESQUISA

O presente trabalho se enquadra na área de pesquisa operacional da engenharia de produção, na subárea de análise de demanda (ABEPRO, [s.d.]). Conforme Silva e Menezes (2001), a pesquisa científica pode ser classificada do ponto de vista da sua natureza, sua abordagem, seus objetivos e seus procedimentos técnicos. Este estudo apresenta como objetivo ser uma pesquisa exploratória, com abordagem quantitativa e natureza considerada como aplicada. Quanto aos procedimentos técnicos, a pesquisa pode ser caracterizada como um estudo de caso. Na Figura 10 ilustra-se a caracterização metodológica da pesquisa.

Figura 10 – Caracterização da pesquisa.



Fonte: Elaborado pelo autor (2021).

Conforme Gil (2008), os objetivos da pesquisa podem ser classificados como de cunho exploratório, descritivo ou explicativo. Ainda segundo o autor, a pesquisa exploratória busca proporcionar maior familiaridade com o problema de pesquisa, com vistas a torná-lo mais explícito ou a construir hipóteses.

A natureza da pesquisa pode ser classificada como básica ou aplicada. A pesquisa aplicada tem por objetivo a aplicação dos conhecimentos com uma finalidade prática e dirigidos à solução de problemas específicos (SILVA; ESTERA MUSZKAT MENEZES, 2005).

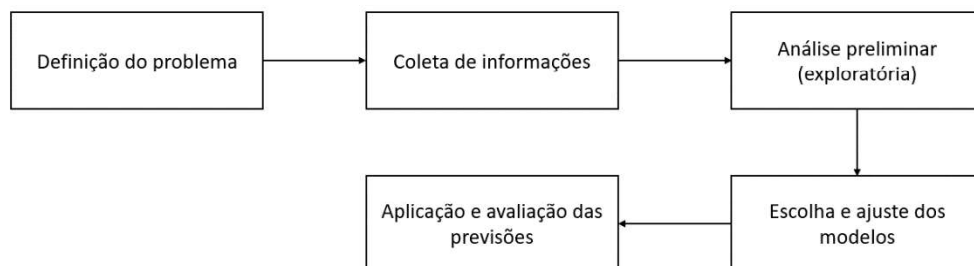
A abordagem da pesquisa, por sua vez, pode ser quantitativa ou qualitativa. A pesquisa quantitativa tem caráter quantificável, requer o uso de recursos e de técnicas estatísticas e possui ênfase na análise de resultados (SILVA; ESTERA MUSZKAT MENEZES, 2005).

Quanto aos procedimentos técnicos, a pesquisa pode ser classificada como: desenvolvimento teórico-conceitual, estudo de caso, levantamentos tipo *survey*, modelamento e simulação, pesquisa bibliográfica, pesquisa experimental ou pesquisa ação (CAUCHICK MIGUEL, 2007). O estudo de caso é marcado pelo estudo profundo e exaustivo de uma entidade bem definida como uma instituição, um programa ou um grupo de indivíduos (CAUCHICK MIGUEL, 2007; GIL, 2008).

3.2 ETAPAS BÁSICAS DE UMA TAREFA DE PREVISÃO

Uma tarefa de previsão geralmente envolve cinco etapas básicas, que podem ser observadas na Figura 11.

Figura 11 – Etapas básicas de uma tarefa de previsão.



Fonte: Elaborado pelo autor (2021).

A definição do problema requer um entendimento de como as previsões serão usadas, quem as requer, a natureza dos itens a serem previstos, o nível de agregação necessário, o horizonte de previsão, dentre outros fatores (HYNDMAN; ATHANASOPOULOS, 2018). Essa etapa pode ser realizada através de pesquisas e entrevistas com as partes interessadas das previsões.

A coleta de informações envolve a identificação de quais dados são necessários e quais estão disponíveis. De acordo com Mentzas *et al.* (1995), nessa fase, o previsor geralmente executa as seguintes tarefas:

- I. Extração dos dados apropriados do sistema de origem;
- II. Modificações, correções apropriadas, etc.
- III. Tratamento de valores ausentes e/ou extremos;
- IV. Carregamento dos dados no destino final.

A análise preliminar, ou análise exploratória, é uma abordagem para analisar conjuntos de dados e resumir suas características principais, geralmente usando gráficos e modelos estatísticos (SAMADI; WILSON; MORADKHANI, 2013). Essa etapa é altamente recomendada na literatura de previsão, uma vez que permite a detecção das características específicas das séries temporais: tendência, sazonalidade, componentes cíclicos, grau de irregularidade e estabilidade da variância (MENTZAS; LINARDOPOULOS; ASSIMAKOPOULOS, 1995).

Na quarta fase, as informações fornecidas pela primeira e terceira etapas são utilizadas para especificar a família dos métodos de previsão. Diversos métodos de previsão de séries temporais estão disponíveis na literatura, como o método de médias móveis, modelos de regressão, suavização exponencial, modelos Box-Jenkins ARIMA, previsão hierárquica, redes neurais, dentre outros. Cada modelo é em si uma construção artificial que se baseia em um conjunto de suposições (explícitas e implícitas) e geralmente envolve um ou mais parâmetros que devem ser estimados (HYNDMAN; ATHANASOPOULOS, 2018).

Na última etapa, uma vez que um ou mais modelos foram especificados e seus parâmetros estimados, é possível realizar previsões para determinada série. A seleção do melhor método preditivo pode ser baseada em qualquer uma das duas abordagens distintas: seleção individual ou seleção agregada (FILDES, 1989). Na seleção individual, todos os métodos preditivos são testados em cada série e o melhor modelo é eleito individualmente. Na seleção agregada, a população inteira (ou uma amostra aleatória) é analisada e o modelo com melhor performance para a população como um todo é determinado. A primeira abordagem é intuitivamente atraente e pode resultar em ganhos substanciais. Entretanto, a seleção individual pode apresentar custo computacional elevado quando há necessidade de se prever um grande número de séries (FILDES; PETROPOULOS, 2015). Complementarmente, alguns autores propõem a segmentação da população em clusters, visando identificar e agrupar séries

temporais com padrões semelhantes e, dessa forma, obter melhores resultados na seleção agregada (MAHARAJ, 2000; VAN WIJK; VAN SELOW, 1999).

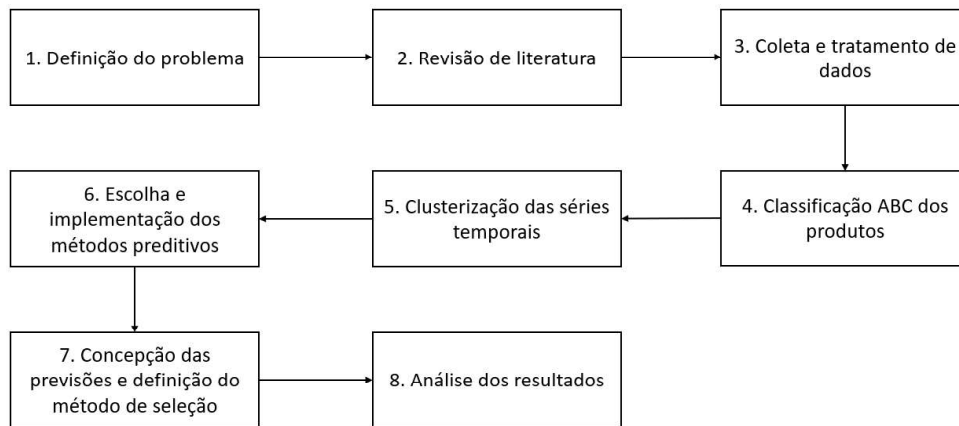
A eleição do melhor método preditivo pode ser baseada no erro da previsão. Para isto, é uma prática comum reservar uma parte do final de cada série para teste e usar o restante da série para treinamento (BERGMEIR; BENÍTEZ, 2012). O erro de previsão pode ser definido como o desvio da demanda real da demanda prevista (FILDES; KINGSMAN, 2011). Diferentes medidas podem ser utilizadas para agregar o erro, como o erro médio absoluto, o erro médio percentual absoluto, o erro quadrático médio.

Por fim, uma etapa adicional comumente aplicada em tarefas de previsão e de gestão de inventário é a classificação ABC dos produtos. Esse método classificatório assume, de acordo com o princípio de Pareto, que um pequeno número de itens é responsável por uma grande parte do volume de faturamento, uma categoria intermediária de itens de volume de faturamento moderado e um grande número de itens de baixo faturamento ou uso (LIIV, 2006). Esse processo classificatório fornece uma estrutura para definir políticas diferentes para os itens de previsão com base em sua importância. Como exemplo, pode-se restringir os métodos preditivos a serem testados nos produtos de menor importância e, assim, reduzir custos computacionais.

3.3 ETAPAS DA PESQUISA

As etapas englobadas no desenvolvimento deste trabalho podem ser vistas na Figura 12.

Figura 12 – Etapas da pesquisa.



Fonte: Elaborado pelo autor (2021).

A primeira etapa deste trabalho foi a definição do problema, a partir da qual foram traçados os objetivos gerais e específicos da pesquisa. Nessa etapa foram realizadas entrevistas com as partes interessadas da empresa a fim de compreender os requisitos do modelo de previsão a ser desenvolvido.

Na segunda etapa, visando aprofundar os conhecimentos na área e embasar o presente trabalho, efetuou-se uma revisão de literatura nas áreas de previsão de demanda, classificação de produtos e avaliação de erro. Também foram pesquisados artigos referentes ao setor varejista para conhecimento do que é publicado sobre esta área.

Após a revisão da literatura, foram coletadas as informações necessárias para o desenvolvimento e testes do modelo a ser proposto. Foram extraídos dados relacionados ao histórico de vendas da empresa e feitas as devidas correções e tratamentos para a execução das etapas subsequentes.

Em seguida, procedeu-se com a classificação ABC dos produtos, através da qual definiu-se o grupo de *SKUs* a serem priorizados nesse estudo. Na etapa de clusterização das séries temporais, foi aplicado o algoritmo *K-Means* com a métrica DTW para formar os agrupamentos sobre os produtos classificados como “A”. Nessa técnica, os centros de cada *cluster* no algoritmo K-Means são calculados como os baricentros em relação ao DTW e, portanto, permitem recuperar uma forma média que imita a forma dos membros do *cluster*, independentemente dos deslocamentos temporais. Tal informação pode ser de grande valia na seleção do método de previsão.

Na etapa seis prosseguiu-se com a escolha e implementação dos métodos preditivos. Por familiaridade, utilizou-se a linguagem de programação Python e pacotes de funções pré-existentes em suas bibliotecas para implementar os métodos escolhidos em um programa computacional.

Posteriormente, o programa computacional desenvolvido foi utilizado para conceber as previsões na etapa sete. Ambas as abordagens, seleção individual e seleção agregada, foram testadas e comparadas no conjunto de séries de interesse. Com base nos resultados obtidos, a abordagem de seleção é definida e o modelo de previsão final é proposto. Por fim, na etapa oito, foram feitas conclusões e comparações do modelo proposto com o método atual da empresa.

4 DESENVOLVIMENTO DO MODELO

O presente capítulo apresenta os resultados encontrados com a aplicação do roteiro metodológico definido no Capítulo 3. Assim, o capítulo compreende o desenvolvimento e aplicação do modelo de previsão, bem como a comparação dos resultados do modelo com aquele comumente utilizado na empresa alvo do estudo.

4.1 REQUISITOS DO MODELO

Uma série de especificações preliminares devem ser consideradas antes de qualquer decisão adicional, com relação à estratégia de previsão a ser seguida (MENTZAS; LINARDOPOULOS; ASSIMAKOPOULOS, 1995). Para esse fim, foi estruturado uma lista com seis perguntas a serem respondidas, de forma a estabelecer uma melhor compreensão do problema e seus requisitos. As perguntas e suas respostas são apresentadas abaixo.

I. Qual é a natureza dos itens a serem previstos?

A empresa objeto de estudo atua no ramo de varejo, sendo especializada no comércio de materiais médicos, ortopédicos, para fitness e mobilidade. Logo, os itens a serem previstos consistem em bens de consumo duráveis. Todos os produtos são adquiridos diretamente com os fabricantes ou distribuidores licenciados, sendo posteriormente revendidos pela varejista em suas lojas. Dessa forma, a empresa não fabrica e nem realiza a montagem de produtos.

II. Como as previsões serão usadas?

As previsões serão utilizadas para auxiliar o processo de compras dos produtos junto aos fornecedores. Atualmente a varejista detém um centro de distribuição, responsável por receber os pedidos, centralizar o estoque da rede e abastecer as lojas conforme demandado. Dessa forma, o estudo será priorizado na previsão de demanda para o abastecimento desse centro de distribuição.

III. Qual é o número de itens a serem previstos?

Conforme dados da varejista, foram comercializados aproximadamente 15 mil produtos distintos entre janeiro de 2016 e fevereiro de 2020. No entanto, a fim de reduzir a complexidade do problema, este trabalho tem como foco apenas os produtos mais importantes, a serem determinados em etapa posterior.

IV. Qual é o nível de agregação?

Neste trabalho, será utilizada a estratégia Bottom-Up, ou seja, as previsões serão realizadas considerando as demandas individuais de cada produto.

V. Qual é o horizonte de previsão?

O horizonte de previsão do modelo será de seis meses de venda, o que vai de acordo com as necessidades da empresa.

VI. Qual é a natureza da informação histórica disponível?

Estão disponíveis dados de histórico de venda desde o ano de 2009, como também dos pedidos de compras, estoque e reposição das lojas. Há também um registro de vendas perdidas, onde são anotados os produtos que foram demandados por clientes e não estavam disponíveis em certa data.

4.2 COLETA E TRATAMENTO DOS DADOS

A varejista catarinense utiliza em todas as suas unidades operacionais um sistema único e integrado de informações, facilitando a coleta de dados para o trabalho. Para o desenvolvimento deste estudo, foi extraído o histórico completo de 50 meses de venda, abrangendo o período de janeiro de 2016 até fevereiro de 2020. Os dados foram compilados e padronizados em uma única tabela, sendo os campos, ou colunas, e seus significados apresentados no Quadro 3. É importante destacar que o registro de vendas perdidas não foi utilizado pois muitas inconsistências em sua base foram encontradas.

Quadro 3 – Significado das colunas dos dados extraídos.

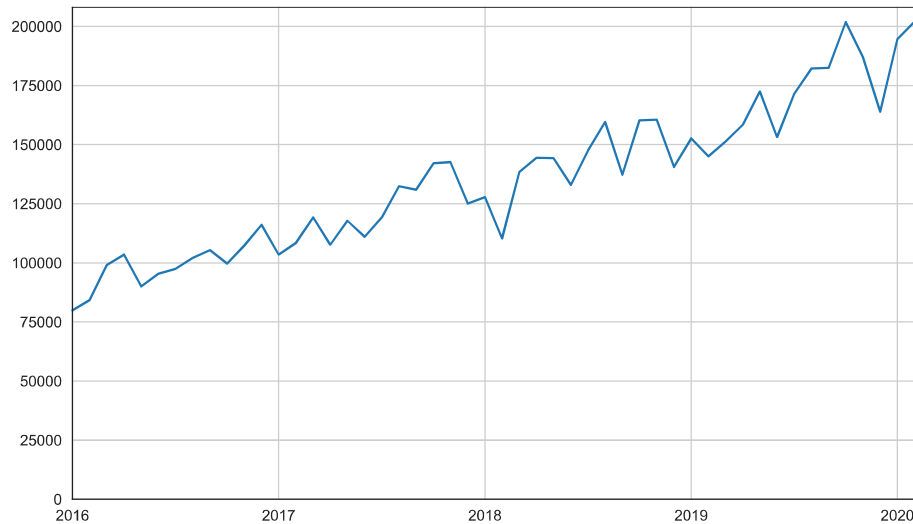
Coluna	Significado
contrato	Código interno que identifica exclusivamente cada registro de venda realizado. Pode abranger um ou mais produtos.
data_venda	Data no formato “dia/mês/ano” em que foi emitida a nota fiscal de determinado contrato.
filial	Código interno que identifica exclusivamente a unidade, ou loja, em que foi efetivado determinado contrato.
codigo_produto	Código interno que identifica o(s) produto(s) comercializado(s) em determinado contrato. É exclusivo para cada <i>SKU</i> .
codigo_marca	Código interno que identifica a marca de determinado <i>codigo_produto</i> .
codigo_grupo	Código interno que identifica a família de produtos em que foi classificado determinado <i>codigo_produto</i> .
quantidade	Quantidade de unidades comercializadas de determinado <i>SKU</i> em determinado contrato. É sempre um número inteiro positivo.
total_liquido	Valor monetário faturado por aquela quantidade comercializada de determinado produto em determinado o contrato.
codigo_vendedor	Código interno que identifica exclusivamente o vendedor, ou atendente, responsável por determinado contrato.
codigo_cliente	Código interno que identifica o cliente, ou comprador, de determinado contrato. É exclusivo para cada CPF ou CNPJ registrado.

Fonte: Elaborado pelo autor (2021).

Através dos dados coletados, pode-se mensurar que, no período de janeiro de 2016 até fevereiro de 2020, foram comercializados 14.816 produtos diferentes em cerca de 1,085 milhão de contratos faturados pela varejista. Todos os produtos estão categorizados em uma das 477 famílias de produtos e 685 marcas registradas no período analisado. A série histórica da quantidade mensal de itens vendidos pela empresa é apresentada na Figura 13, na qual é

possível observar um crescimento constante nas vendas mensais que resulta em um aumento cumulativo de 131% no período de janeiro de 2016 e fevereiro de 2020.

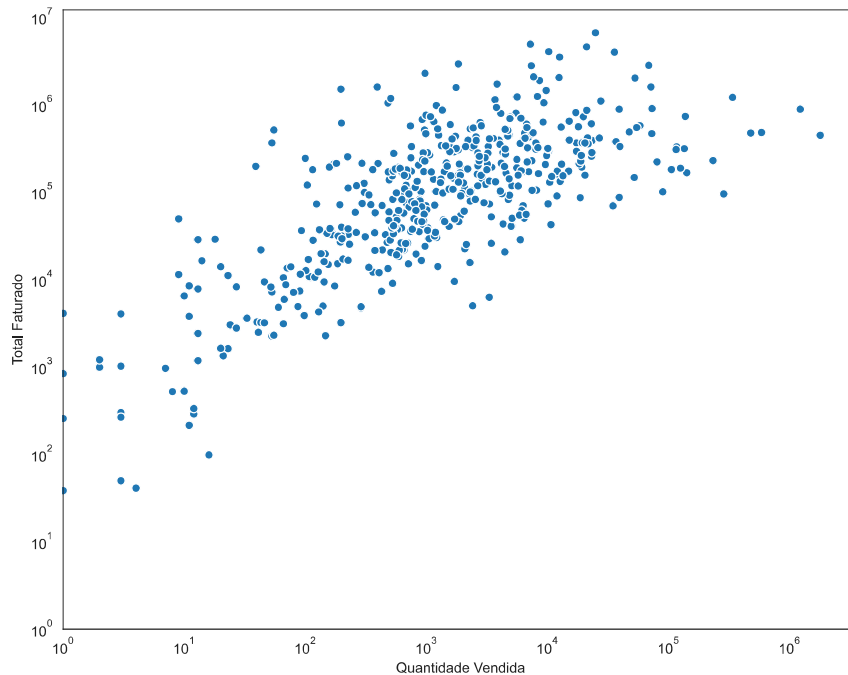
Figura 13 – Série histórica da quantidade de itens vendidos pela empresa por mês entre Jan/16 e Fev/20.



Fonte: Elaborado pelo autor (2021).

A Figura 14 exibe o gráfico de dispersão em escala logarítmica do total faturado vs. a quantidade vendida por cada família de produto no período total extraído.

Figura 14 – Total faturado vs. quantidade vendida por cada família de produtos.



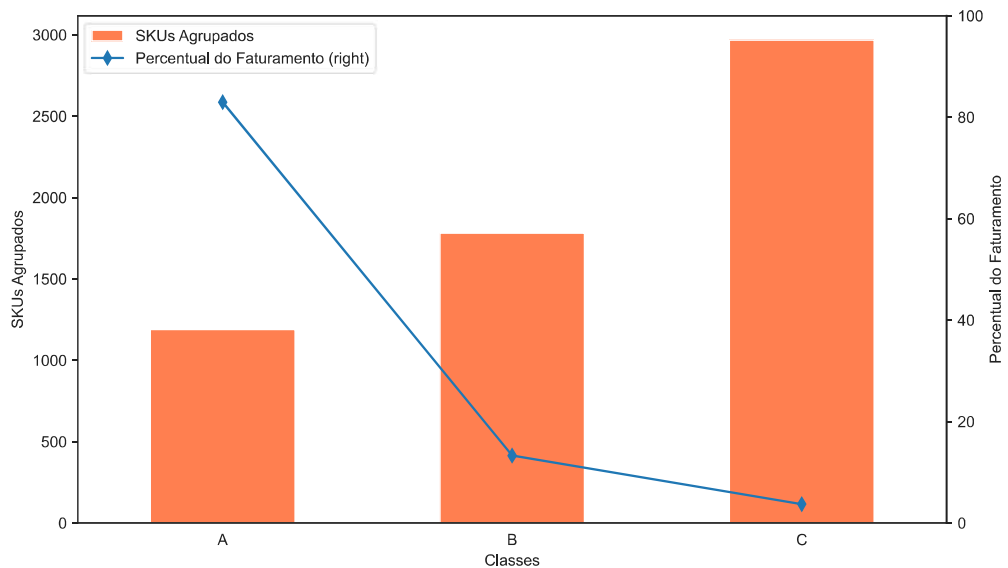
Fonte: Elaborado pelo autor (2021).

Apesar de contabilizar quase 15 mil itens comercializados no período, o volume de produtos de interesse é reduzido para 5.960 ao desconsiderar-se aqueles que estão atualmente fora de linha, assim como os que são adquiridos apenas sob encomenda de clientes.

4.3 CLASSIFICAÇÃO ABC

No que diz respeito ao grande número de SKUs, torna-se necessário classificá-los conforme sua participação no faturamento para, dessa forma, focalizar o estudo nos produtos mais importantes. Assim, realizou-se nesta etapa a classificação ABC dos itens de acordo com a regra comumente aplicada de que as classes A, B e C contêm, respectivamente, cerca de 20%, 30% e 50% de todos os *SKUs* agrupados em ordem decrescente pela participação no faturamento total. Foram excluídos dessa classificação os produtos que estão atualmente fora de linha e aqueles que são adquiridos apenas sob encomenda de clientes. A Figura 15 ilustra o resultado dessa classificação.

Figura 15 – Resultado da classificação ABC dos produtos.



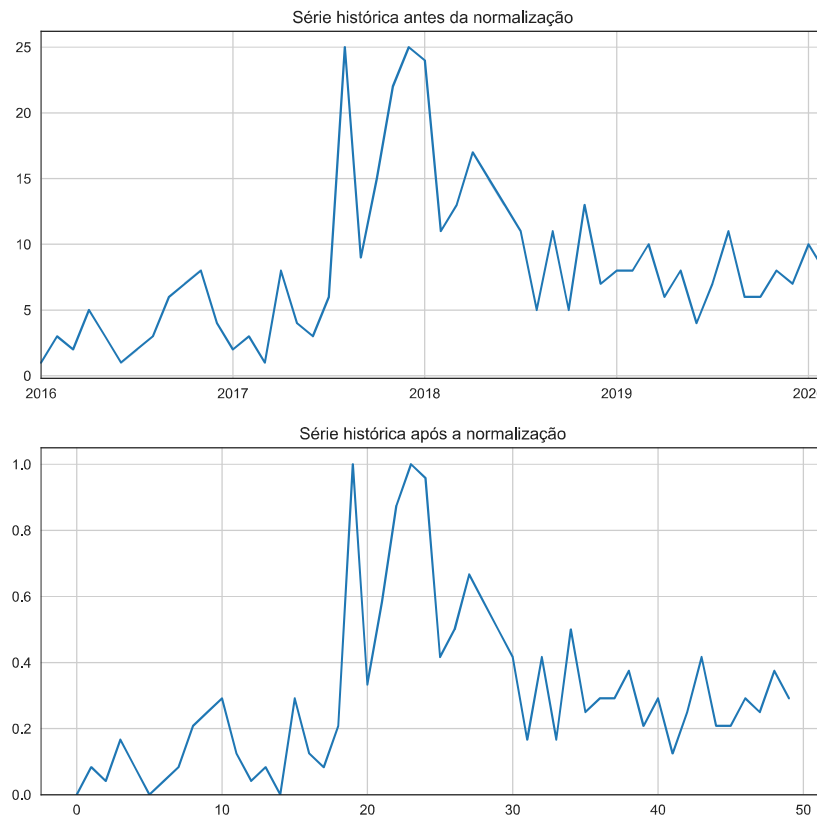
Fonte: Elaborado pelo autor (2021).

Com base nos resultados apresentados, constata-se que os 1.193 produtos pertencentes à classificação A representam juntos 83,05% do faturamento total. Os 1.788 itens pertencentes à classificação B representam cerca de 13,23% da receita. Por sua vez, os 2.979 produtos pertencentes à classificação C juntos somam os 3,72% restantes do faturamento total. Atentando-se para os resultados, este estudo limitar-se-á apenas aos produtos classificados como “A”, posto que são os mais significativos para a receita.

4.4 CLUSTERIZAÇÃO DE SÉRIES TEMPORAIS

A clusterização de séries temporais tem como objetivo auxiliar o método da seleção agregada dos modelos preditivos, dado que a seleção individual pode apresentar alto custo computacional. Para tanto, empregou-se o algoritmo *K-Means* para formar os *clusters* com a métrica DTW para medir a similaridade entre as séries temporais dos produtos classificados como “A”. É prática comum normalizar as séries temporais, mapeando-as dentro da escala [0,1] através do método de normalização *Min-Max*. Essa etapa é sugerida para que a escala de venda dos produtos não interfira no resultado da agrupação. A Figura 16 apresenta o comparativo do gráfico da série temporal de um produto antes e depois da normalização.

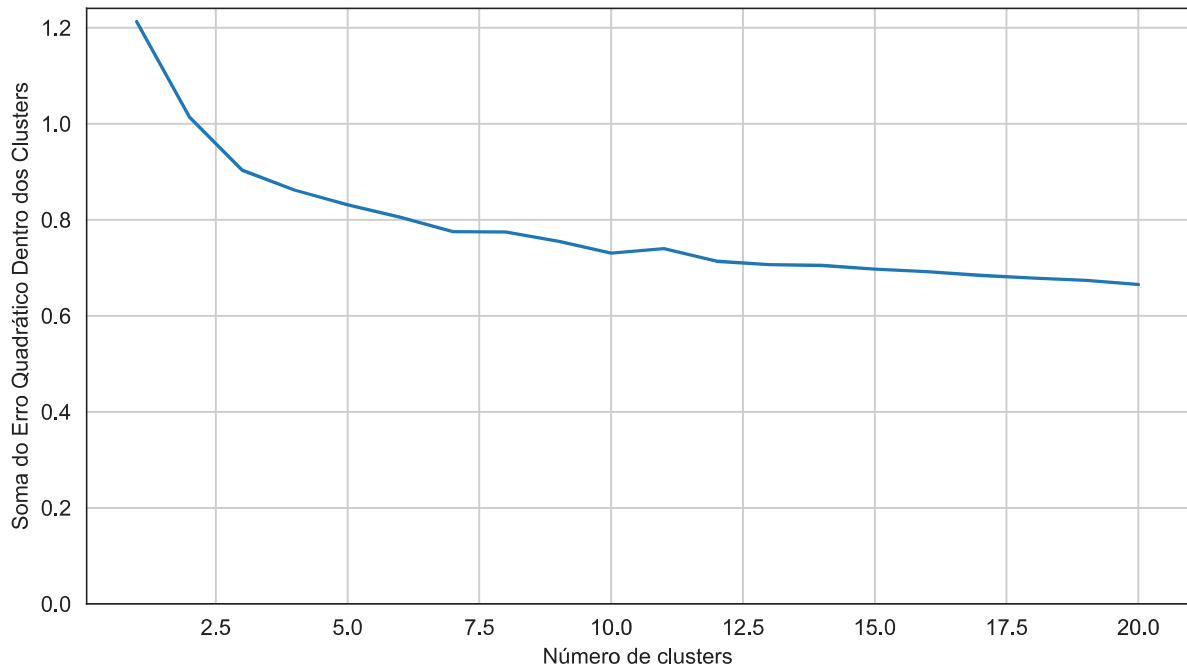
Figura 16 – Série temporal da quantidade mensal vendida de um produto antes e depois da normalização *Min-Max*.



Fonte: Elaborado pelo autor (2021).

Após normalizar as séries temporais, aplicou-se o algoritmo *K-means*. Para determinar o número de *clusters* ideal, pode-se traçar a curva K-SSE e utilizar o método de *Elbow*. A Figura 17 ilustra a curva K-SSE obtida. Por meio da análise visual, decidiu-se adotar K igual a 10. Para esse valor de K, o somatório do erro quadrático dentro dos *clusters* é aproximadamente 40% menor em relação à opção inicial, garantindo assim maior congruência dos grupos formados. Também foi observado que aumentando K de 10 para 20, um aumento de 100% no número de centros, apenas diminuiria o SSE em 8,9%.

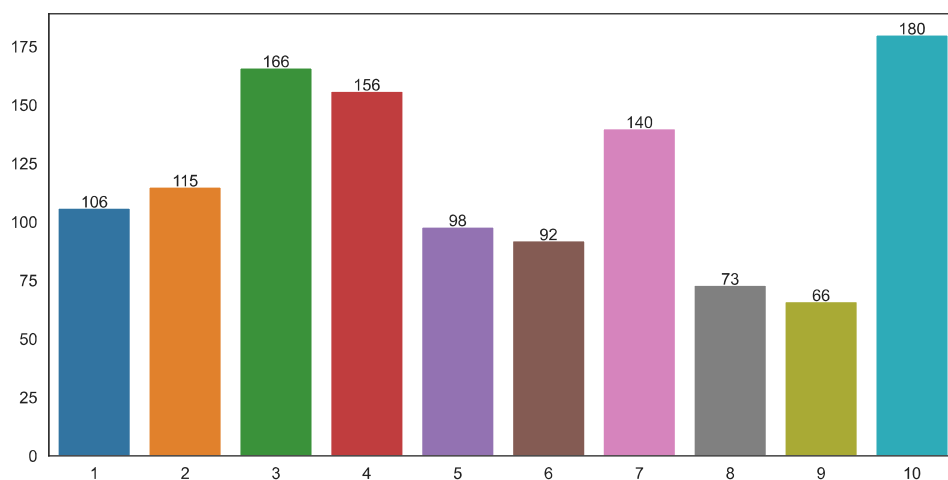
Figura 17 – Curva do somatório do erro quadrático dentro dos *clusters* em função do número de *clusters* formados.



Fonte: Elaborado pelo autor (2021).

Uma vez definido o número K de *clusters*, analisou-se os resultados para esse K . A Figura 18 apresenta a quantidade de produtos agrupados em cada um dos 10 *clusters* formados.

Figura 18 – Contagem de produtos agrupados em cada *cluster* para K igual a 10.



Fonte: Elaborado pelo autor (2021).

Nessa técnica, os centroides de cada *cluster*, ou baricentros, são calculados em relação ao DTW. Um baricentro é a sequência média de um grupo de séries temporais no espaço DTW.

Como resultado, os centroides têm uma forma média que imita a forma dos membros do *cluster*, independentemente de onde ocorrem mudanças temporais. A Figura 19 ilustra as séries temporais dos centroides obtidos.

Figura 19 – Série temporal normalizada dos centroides obtidos de cada *cluster*.

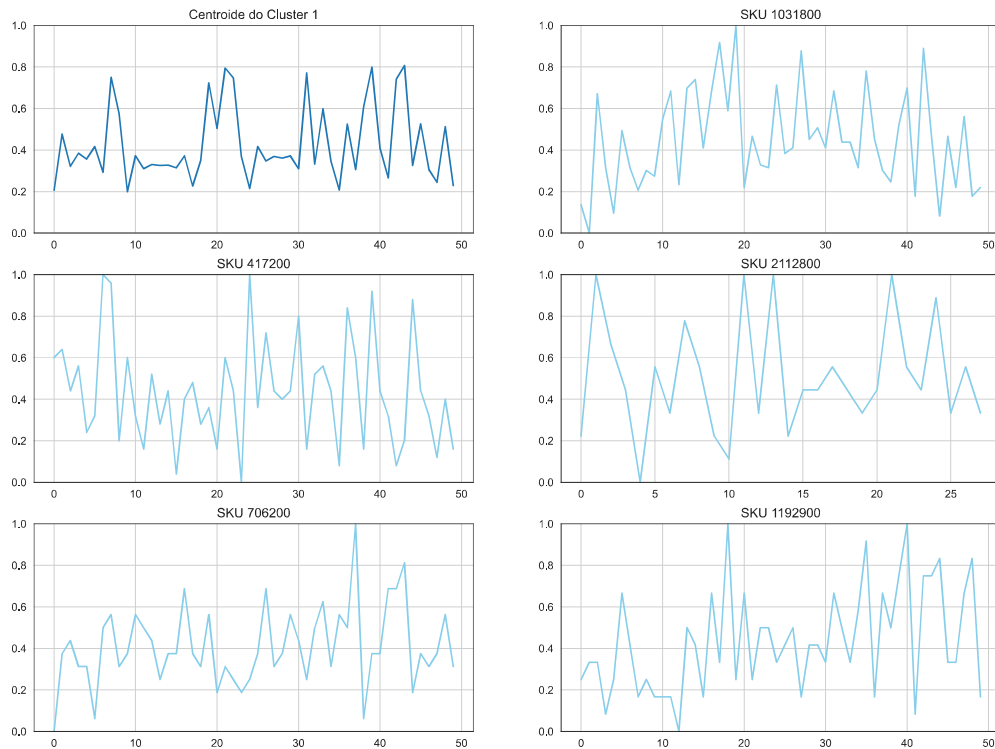


Fonte: Elaborado pelo autor (2021).

Em seguida, realizou-se uma breve comparação visual das séries temporais normalizadas dos centroides com uma amostra aleatória de 5 produtos de cada *cluster*. É desejável que as séries dentro de um *cluster* sejam o mais semelhante possível entre si. Ao mesmo tempo, as séries de dois *clusters* distintos deveriam ser diferentes uma da outra.

O *cluster 1*, com seu centroide e amostra aleatória de cinco produtos expostos na Figura 20, é composto por produtos com histórico de vendas que não apresentam tendência crescente ou decrescente nem sazonalidade bem definida. As séries amostradas exibem característica estacionárias, com valores que variam em torno de uma média constante com flutuações irregulares.

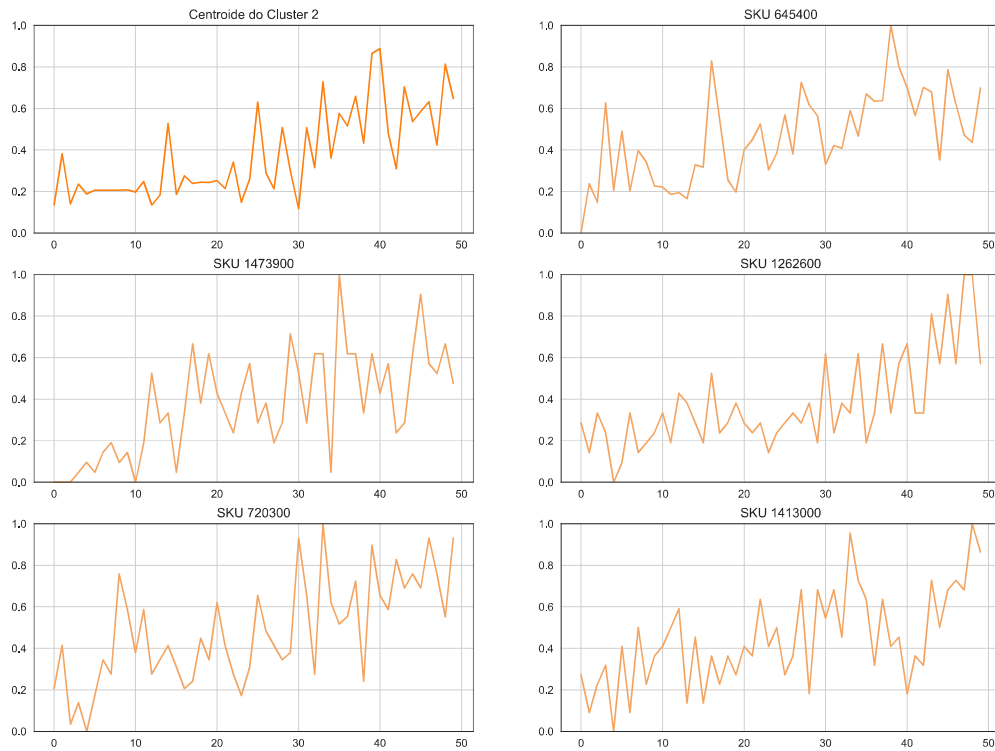
Figura 20 – Centroide e amostra aleatória de 5 produtos do *cluster 1*.



Fonte: Elaborado pelo autor (2021).

Nas séries temporais do *cluster 2*, ilustradas na Figura 21, é possível observar uma tendência crescente desde o início do período analisado. Além disso, nota-se que as flutuações com relação à linha de tendência não são regulares.

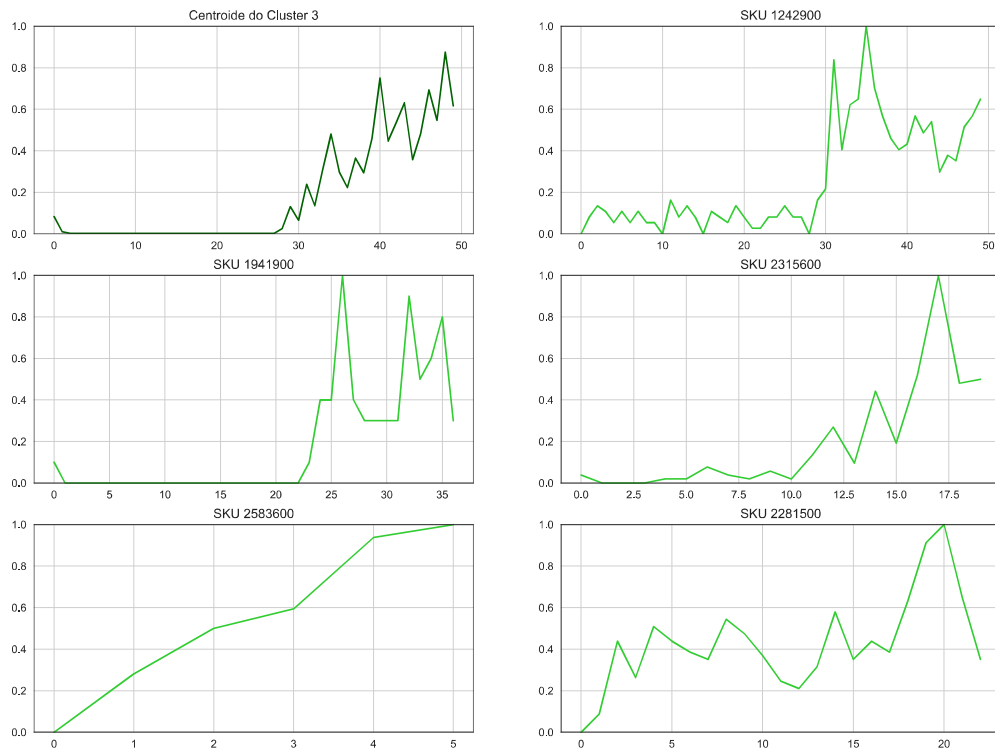
Figura 21 – Centroide e amostra aleatória de 5 produtos do *cluster 2*.



Fonte: Elaborado pelo autor (2021).

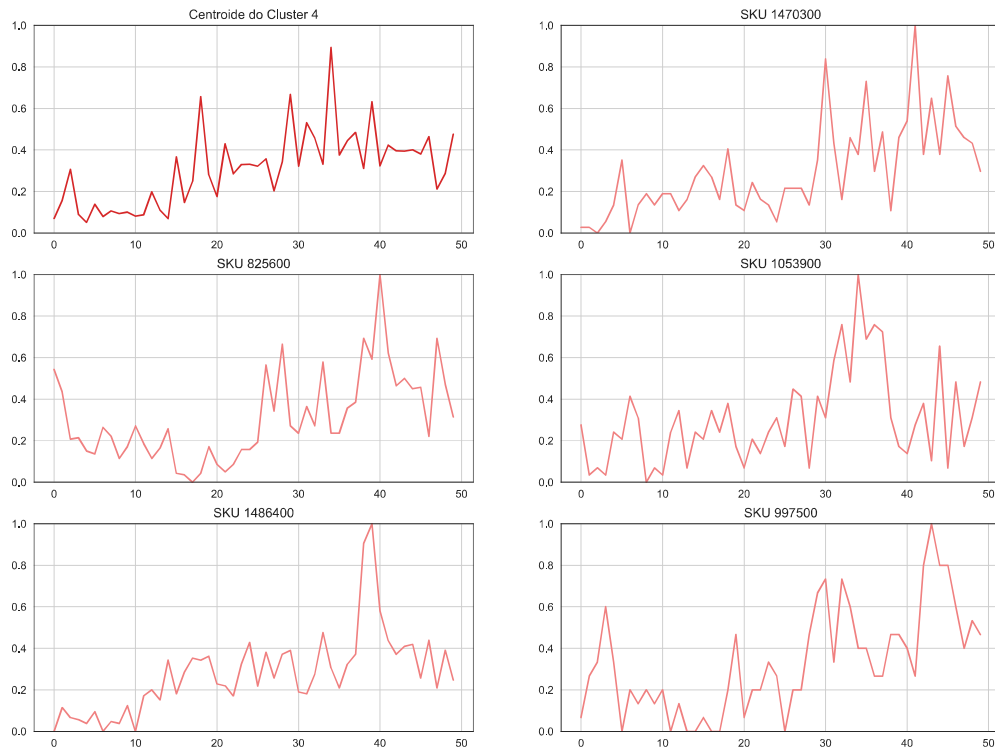
O *cluster 3*, conforme exposto na Figura 22, compreende produtos com grande tendência de crescimento nos últimos períodos e vendas relativamente baixas ou inexistentes nos primeiros meses. Isso indica a possibilidade de se tratar de produtos recém-lançados ou com crescimento acelerado nos últimos períodos.

Figura 22 – Centroide e amostra aleatória de 5 produtos do *cluster* 3.



Fonte: Elaborado pelo autor (2021).

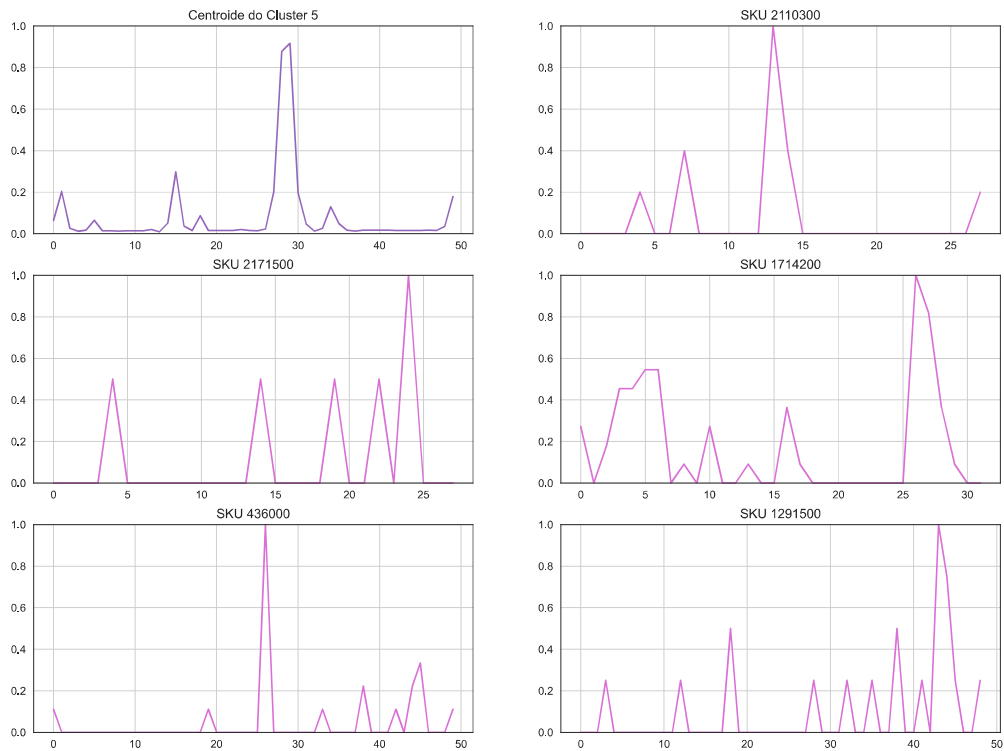
As séries do *cluster* 4, ilustradas na Figura 23, apresentam tendência levemente crescente durante todo o período. Evidencia-se também que as séries amostradas exibem flutuações irregulares com picos grandes de vendas em períodos mais recentes, característica semelhantemente observada na série do centroide.

Figura 23 – Centroide e amostra aleatória de 5 produtos do *cluster 4*.

Fonte: Elaborado pelo autor (2021).

O *cluster 5*, representado na Figura 24, exibe produtos com demanda intermitente, ou seja, com períodos sem vendas e picos com periodicidade alta e irregular entre si.

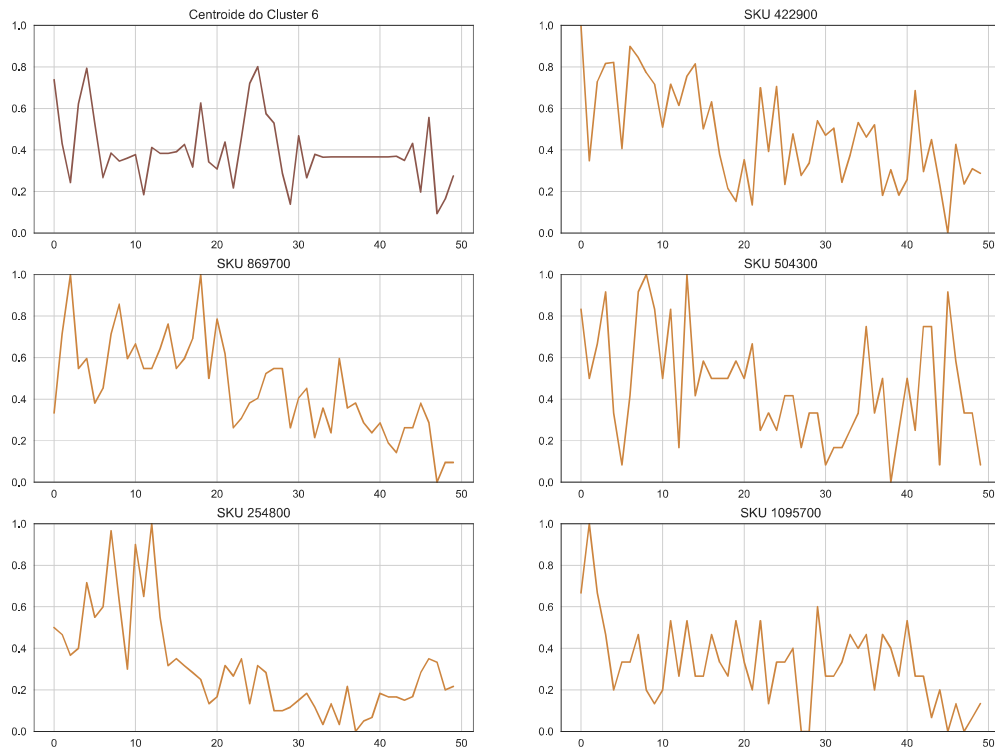
Figura 24 – Centroide e amostra aleatória de 5 produtos do *cluster 5*.



Fonte: Elaborado pelo autor (2021).

Por sua vez, o *cluster 6*, com seu centroide e amostras de produtos mostrados na Figura 25, contempla séries com tendência decrescente. Ademais, as séries apresentam picos e vales nas vendas oscilando de forma irregular em torno da linha de tendência.

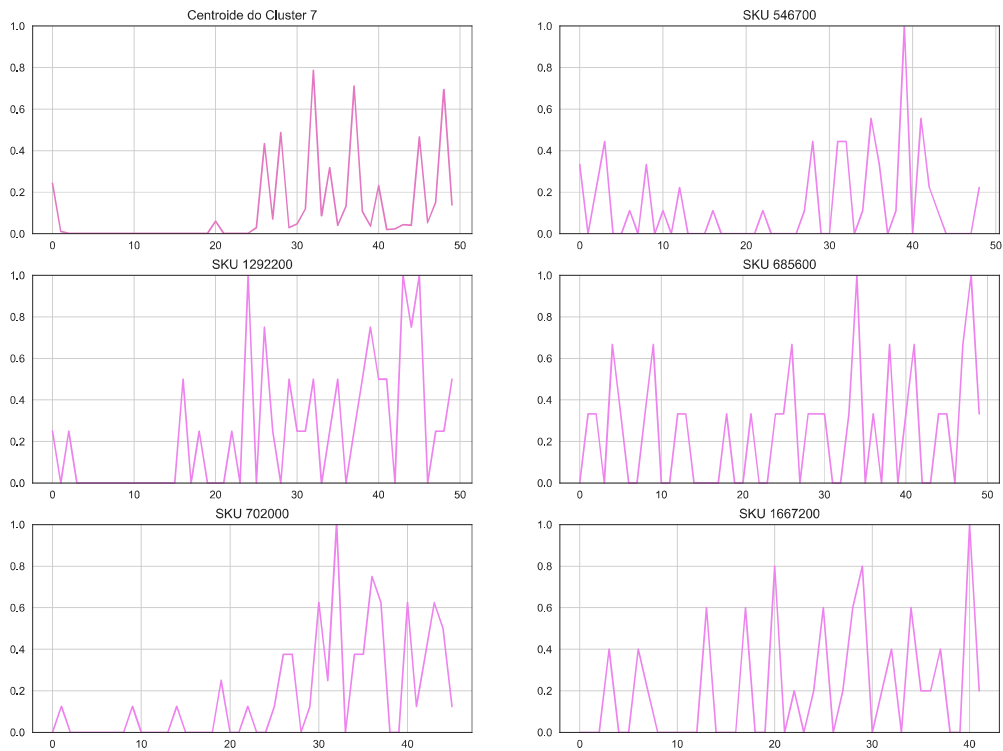
Figura 25 – Centroide e amostra aleatória de 5 produtos do *cluster 6*.



Fonte: Elaborado pelo autor (2021).

O *cluster 7*, conforme ilustra a Figura 26, engloba produtos com demanda intermitente, porém com periodicidade menor entre dois períodos de venda se comparado ao *cluster 5*, embora também apresente picos irregulares ao longo da série.

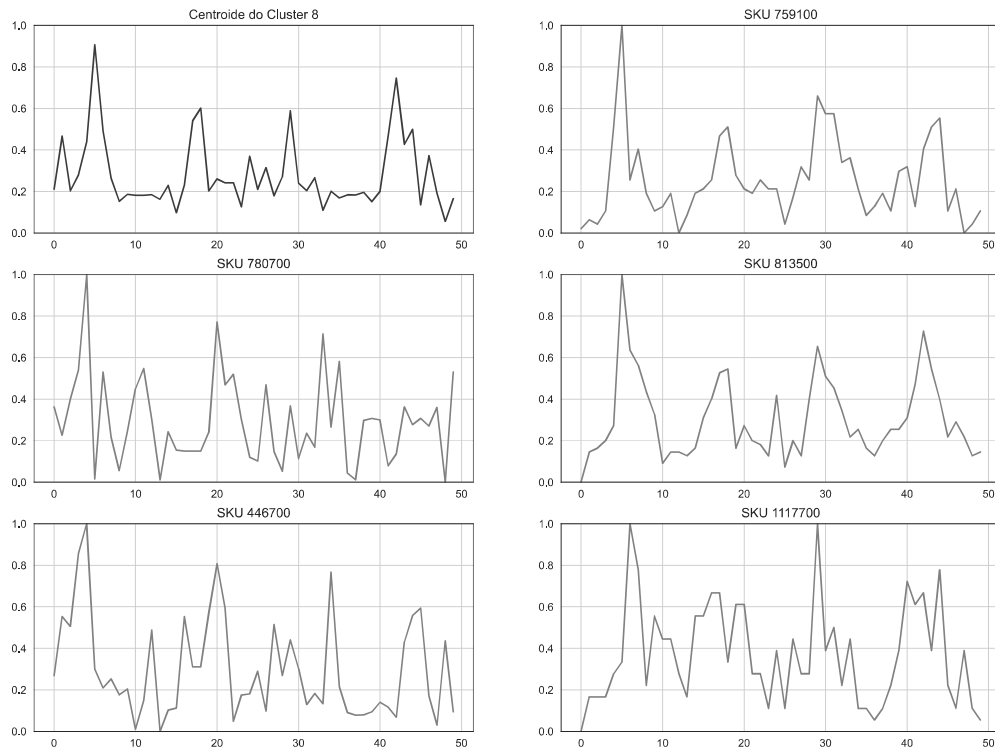
Figura 26 – Centroide e amostra aleatória de 5 produtos do *cluster* 7.



Fonte: Elaborado pelo autor (2021).

A Figura 27 apresenta o centroide e amostra de produtos do *cluster* 8. Esse *cluster* contém produtos com média de vendas estacionária, ou seja, sem tendência crescente ou decrescente, no entanto, é possível observar sazonalidade nos picos e vales das vendas.

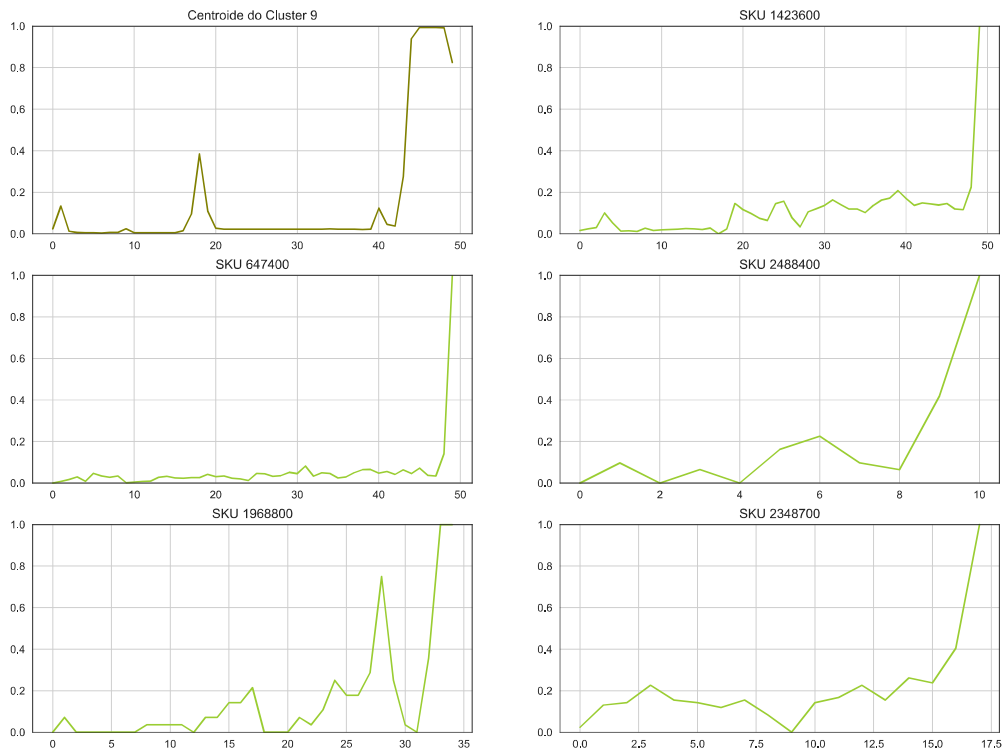
Figura 27 – Centroide e amostra aleatória de 5 produtos do *cluster* 8.



Fonte: Elaborado pelo autor (2021).

O *cluster* 9, mostrado na Figura 28, compreende produtos com poucas vendas em grande parte do período analisado e com um pico de crescimento exponencial nas vendas nos últimos meses.

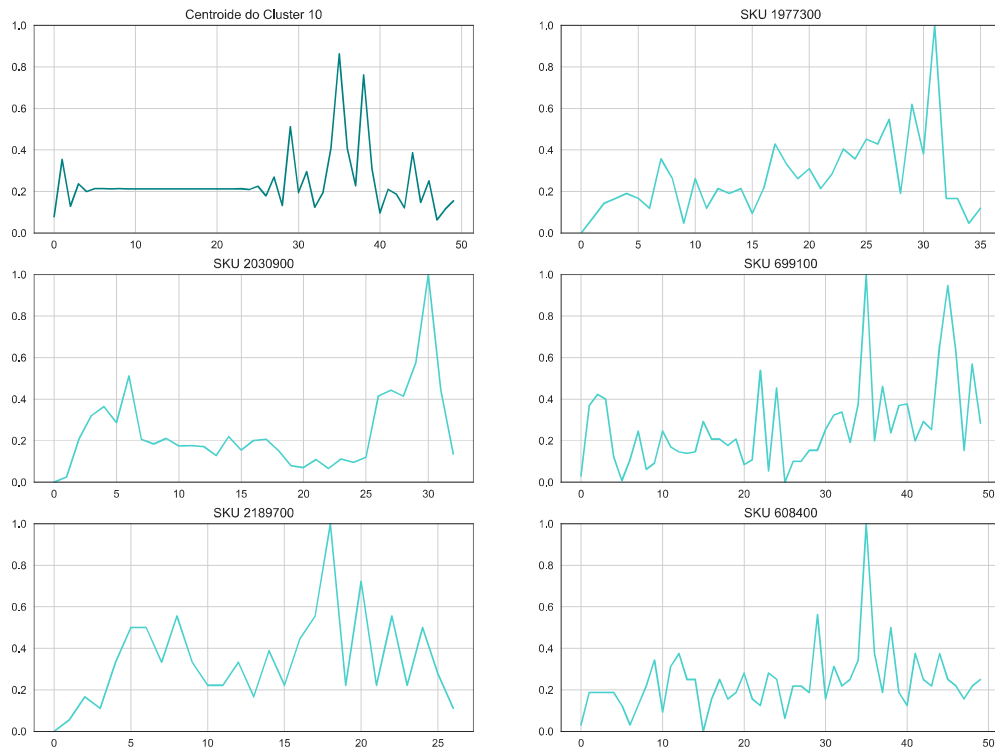
Figura 28 – Centroide e amostra aleatória de 5 produtos do *cluster* 9.



Fonte: Elaborado pelo autor (2021).

Finalmente, a Figura 29 ilustra as características do *cluster* 10. Esse cluster não apresenta sazonalidade e nem uma tendência visível de crescimento, porém apresenta flutuações irregulares com picos mais altos a partir da segunda metade do período analisado.

Figura 29 – Centroide e amostra aleatória de 5 produtos do *cluster* 10.



Fonte: Elaborado pelo autor (2021).

4.5 ESCOLHA E IMPLEMENTAÇÃO DOS MÉTODOS PREDITIVOS

Concluída a etapa de clusterização dos produtos, prosseguiu-se para a especificação das famílias de métodos de previsão a serem utilizadas. Diversos métodos de previsão de séries temporais foram desenvolvidos nos últimos anos, cada um com suas próprias vantagens e limitações. Atentando-se às indicações da literatura, nesse trabalho utilizar-se-á as seguintes classes de métodos: Suavização exponencial, Modelos ARIMA, SARIMA e Redes neurais recorrentes.

Dado que cada modelo geralmente envolve um ou mais parâmetros e hiperparâmetros que devem ser definidos, a seguir é apresentada uma breve descrição da abordagem utilizada em cada família de métodos.

4.5.1 Suavização Exponencial

Ao considerar as variações nas combinações dos componentes de tendência e sazonalidade, 12 métodos de suavização exponencial são possíveis, conforme ilustrado na Figura 2. As equações e o conjunto de parâmetros a serem estimados em cada variação são melhor apresentados em Gardner (2006). Para determinar a variação com melhor resultado em determinada série temporal, empregou-se nesse trabalho uma técnica chamada *Grid Search*, onde todas as combinações de tendência e sazonalidade são testadas e aquela com menor erro de previsão é retornada. Para tanto, comparou-se as previsões de cada variação com um conjunto de dados de teste de seis meses. A métrica RMSE foi utilizada para agregar o erro. Os parâmetros foram estimados através do *Maximum Likelihood Estimation* (MLE), cujo tema é melhor abordado em Myung (2003). É importante ressaltar que, uma vez que os dados de venda foram agrupados com frequência mensal, fixou-se 12 como sendo a duração do período sazonal.

4.5.2 Modelos ARIMA e SARIMA

O modelo completo do ARIMA não sazonal depende da estimativa dos hiperparâmetros p , d e q , onde p é a ordem do modelo autorregressivo, d é o grau de diferenciação e q é a ordem do modelo de média móvel. Diversas técnicas podem ser utilizadas para determinar esses valores. Nesse trabalho empregou-se o Critério de informação de Akaike, cujo a teoria é melhor abordado em Ozaki (1977). Uma grande vantagem dessa técnica é que o modelo de melhor ajuste de acordo com a AIC é aquele que explica a maior quantidade de variação usando o menor número possível de variáveis independentes (MONDAL; SHIT; GOSWAMI, 2014). Após a determinação da ordem do modelo, os parâmetros remanescentes são estimados através do MLE.

De maneira análoga, foi utilizado o AIC para definição dos hiperparâmetros p , d , q , P , D e Q do modelo SARIMA. O período sazonal foi fixado como 12 para todas as séries temporais. Por fim, empregou-se o MLE para a estimativa dos demais parâmetros.

4.5.3 Redes Neurais Recorrentes

Os hiperparâmetros que determinam a estrutura de uma rede neural são normalmente: número e arquitetura das camadas ocultas, quantidade de neurônios em cada camada, função de ativação, taxa de abandono, parada antecipada, tamanho do lote, número de épocas e o algoritmo de otimização. Conforme Wu *et. al* (2020), a infinidade de hiperparâmetros para ajustar nas arquiteturas de redes neurais torna impossível verificar todas as combinações. A fim de reduzir a complexidade do problema, costuma-se fixar alguns desses hiperparâmetros enquanto se realiza um *Grid Search* dentro de um conjunto limitados de opções nos demais. O Quadro 4 resume a abordagem adotada na escolha dos hiperparâmetros da RNR nesse trabalho. A escolha dessa abordagem foi baseada em estudos empíricos conduzidos e relatados na literatura por Agarap (2018), Chang *et al.* (2019), Gal e Ghahramani (2016), Lu *et al.* (2016) e Zaremba *et al.* (2014).

A combinação de hiperparâmetros com melhor resultado foi definida como aquela com menor erro de previsão. Para tal fim, foram comparadas as previsões de cada configuração obtida no *Grid Search* com um conjunto de dados de teste de seis meses. É importante ressaltar que o método de inicialização dos pesos na RNR foi definido como aleatório. Dessa forma, tendo em vista que em cada inicialização pode se obter um resultado final diferente, optou-se por inicializar três vezes cada modelo, retornando ao final a inicialização com menor erro de previsão. Empregou-se a métrica RMSE para agregar o erro de previsão obtido.

Quadro 4 – Abordagem de escolha dos hiperparâmetros das Redes Neurais.

Hiperparâmetro	Grau de Liberdade	Valor
Número de camadas ocultas	Fixo	2 camadas
Arquitetura das camadas	Fixo	<i>Long Short-Term Memory</i>
Quantidade de neurônios em cada camada	<i>Grid Search</i>	1ª camada: 50, 100 ou 150 2ª camada: 25, 50 ou 75
Função de ativação	Fixo	<i>Rectified Linear Unit (ReLU)</i>
Taxa de abandono	Fixo	10% em cada camada oculta
Paciência na parada antecipada	Fixo	2 épocas
Critério monitorado na parada antecipada	Fixo	Erro no conjunto de dados de validação
Métrica de erro	Fixo	Erro quadrático médio
Tamanho do lote	<i>Grid Search</i>	1 ou 6 amostras de treinamento
Número máximo de épocas	Fixo	100
Algoritmo de otimização	Fixo	<i>Adaptive Moment Estimation (Adam)</i>

Fonte: Elaborado pelo autor (2021).

A linguagem de programação Python e pacotes de funções pré-existent em suas bibliotecas foram utilizados para implementar os métodos de previsão escolhidos em um programa computacional. O *hardware* do computador utilizado é resumido no Quadro 5.

Quadro 5 – Hardware utilizado.

Componente	Modelo
CPU	Intel® Core™ i5-10600KF
RAM	16,0 GB DDR4 3200Mhz
DISK	SSD 512GB PCIe Gen3x4
GPU	AMD Radeon RX 5600 XT

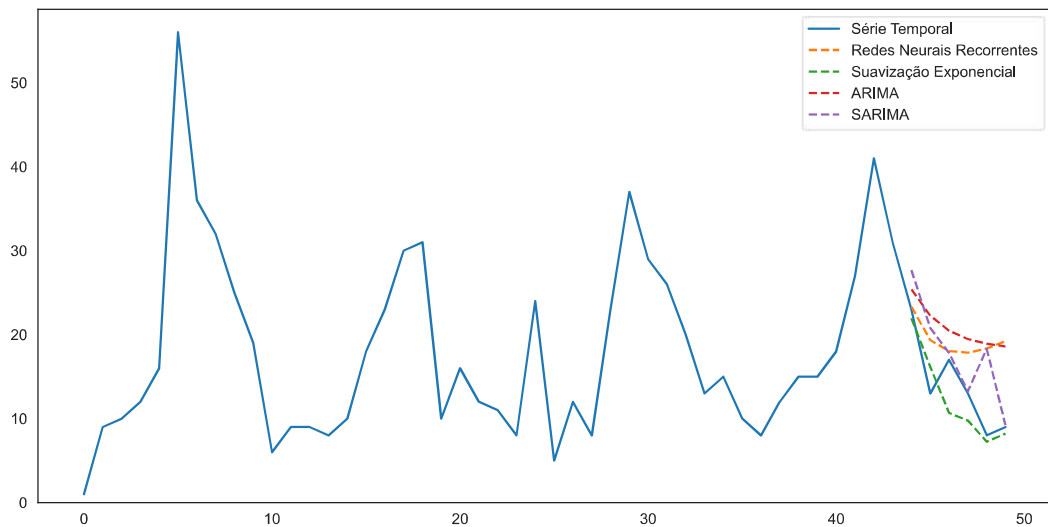
Fonte: Elaborado pelo autor (2021).

4.6 CONCEPÇÃO DAS PREVISÕES E DEFINIÇÃO DO MÉTODO DE SELEÇÃO

Após o desenvolvimento do programa computacional, o mesmo pode ser utilizado para conceber as previsões de dada série temporal. Tendo como exemplo, a Figura 30 apresenta as

previsões obtidas por cada uma das quatro famílias de métodos para o *SKU 813500*. Os hiperparâmetros escolhidos pelo programa computacional em cada família são apresentados no Quadro 6. Adotou-se como período de teste os últimos seis meses de dados e foi utilizado o RMSE como métrica de erro. A Figura 31 ilustra os erros obtidos no conjunto de treino e teste por cada família de métodos preditivos para esse SKU.

Figura 30 – Previsões obtidas pelo programa computacional para o *SKU 813500*.

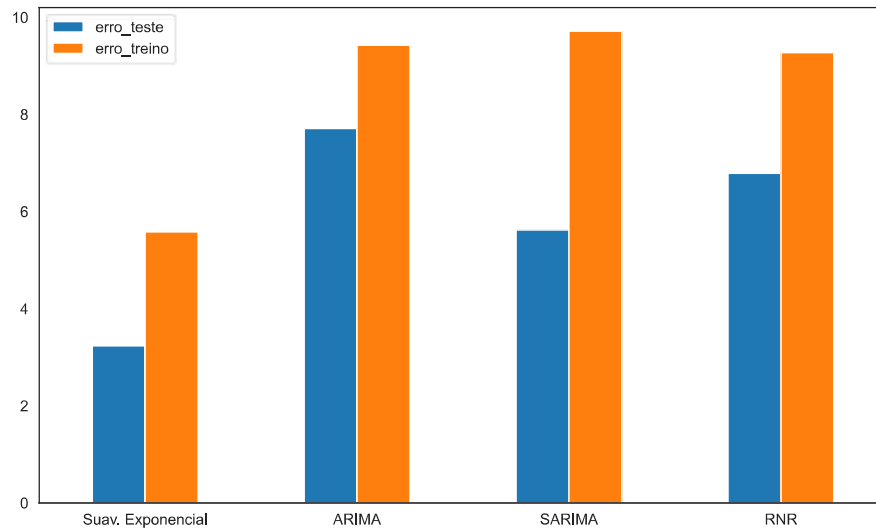


Fonte: Elaborado pelo autor (2021).

Quadro 6– Hiperparâmetros escolhidos pelo programa para o *SKU 813500*.

Família de Método	Hiperparâmetros escolhidos
Suav. Exponencial	Perfil de tendência: Exponencial; Perfil de sazonalidade: Multiplicativo; Periodicidade sazonal: 12.
ARIMA	$p = 1; d=0; q=0$.
SARIMA	$p = 1; d=0; q=0;$ $P=2; D=0; Q=0; m=12$.
Redes Neurais Recorrentes	1ª camada: 100 neurônios; 2ª camada: 50 neurônios; Tamanho do lote: 1.

Fonte: Elaborado pelo autor (2021).

Figura 31 – Erros obtidos no conjunto de treino e teste para o *SKU 813500*.

Fonte: Elaborado pelo autor (2021).

Analisando a Figura 31 em conjunto com o Quadro 6, é possível constatar que a suavização exponencial com perfil de tendência exponencial e sazonalidade multiplicativa foi o método preditivo que apresentou o menor RMSE no conjunto de dados de teste para a o *SKU 813500*. O custo computacional obtido em cada etapa executada pelo programa computacional para o *SKU 813500* é apresentado na Tabela 1.

Tabela 1 – Custo computacional por etapa para o *SKU 813500*.

Etapa	Tempo de processamento (segundos)
<i>Grid Search</i> na Suavização Exponencial	1,13
Treino da variação escolhida na Suavização Exponencial	0,10
Definição dos hiperparâmetros pelo AIC no ARIMA	0,24
Treino do modelo ARIMA com os hiperparâmetros escolhidos	0,03
Definição dos hiperparâmetros pelo AIC no SARIMA	2,26
Treino do modelo SARIMA com os hiperparâmetros escolhidos	0,06
<i>Grid Search</i> nas Redes Neurais Recorrentes	40,33
Treino da RNR com a configuração escolhida	5,59

Fonte: Elaborado pelo autor (2021).

Observa-se na Tabela 1 que as Redes Neurais Recorrentes apresentaram um custo computacional elevado em comparação com as demais famílias de métodos. Por sua vez, o ARIMA foi a classe de métodos que apresentou o menor custo computacional, custando apenas 0,2682 segundo de processamento no total.

De acordo com Fildes (1989), a seleção do melhor método preditivo pode ser baseada em qualquer uma das duas abordagens distintas: seleção individual ou seleção agregada. A seleção individual do melhor método preditivo, onde todos os métodos disponíveis são testados em cada série e o melhor modelo é eleito individualmente, pode resultar em ganhos substanciais na acurácia de previsão. Entretanto, o custo computacional pode ser excessivamente elevado quando há necessidade de se prever um grande número de séries. Neste trabalho optou-se por testar e comparar as duas metodologias para, dessa forma, determinar a mais vantajosa.

Dado que no método da seleção individual deve-se testar todos os métodos preditivos em todas as séries temporais de interesse, utilizou-se o programa computacional desenvolvido na etapa anterior para conceber as previsões para todos os 1.193 *SKUs* classificados como “A”.

A eleição do melhor método preditivo em cada série foi baseada no menor erro de previsão. Para tanto, adotou-se novamente os últimos seis meses de dados como período de teste e empregou-se o RMSE como métrica de erro.

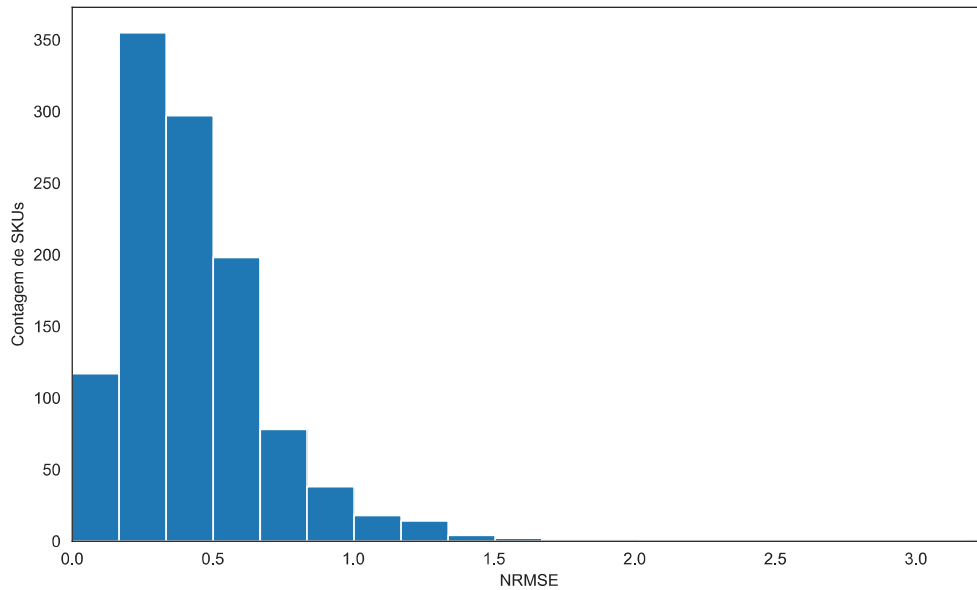
Em seguida, com o intuito de mensurar o erro médio de previsão obtido em todas as 1.193 séries temporais, faz-se necessário normalizar os erros individuais para que a escala das unidades não interfira na análise. Para tanto, pode-se utilizar a métrica NRMSE que consiste na divisão do RMSE pelo valor médio do conjunto de dados de teste (FANG et al., 2008). O NRMSE é frequentemente expresso como uma porcentagem, onde valores mais baixos indicam menor variância residual. Dessa forma, o NRMSE pode ser interpretado de maneira análoga ao MAPE. A grande vantagem do NRMSE em relação ao MAPE é que, como os erros são elevados ao quadrado antes de serem calculados, o NRMSE atribui um peso relativamente alto aos grandes desvios. A Tabela 2 sumariza os resultados obtidos no método da seleção individual. A Figura 32 ilustra o histograma dos NRMSE obtidos. Por sua vez, a Figura 33 ilustra a quantidade de *SKUs* em que cada família de métodos preditivos foi eleita como melhor modelo.

Tabela 2 – Resultados obtidos da aplicação da seleção individual nas 1193 séries temporais priorizadas.

Métrica	Valor
Custo computacional total	74709 segundos (20h45min)
Custo computacional médio por SKU	62,6 segundos
Média dos NRMSE obtidos	43,1%
Desvio padrão dos NRMSE obtidos	26,8%
Menor NRMSE obtido	0.00%
25º percentil dos NRMSE obtidos	24,4%
50º percentil dos NRMSE obtidos	37,8%
75º percentil dos NRMSE obtidos	55,7%
Maior NRMSE obtido	321,6%

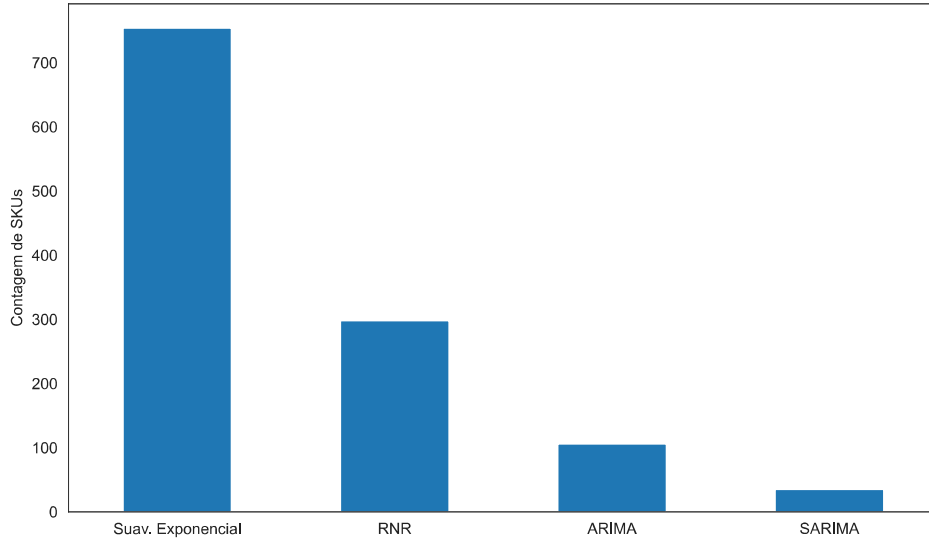
Fonte: Elaborado pelo autor (2021).

Figura 32 – Histograma dos NRMSE obtidos na seleção individual.



Fonte: Elaborado pelo autor (2021).

Figura 33 – Contagem de *SKUs* em que cada família de métodos preditivos foi eleita como melhor modelo na seleção individual.



Fonte: Elaborado pelo autor (2021).

Analisando a Figura 32 é possível observar que a suavização exponencial foi eleita como a melhor família de métodos preditivos para 754 *SKUs*, o que representa cerca de 63,20% do total de produtos classificados como “A”. Por sua vez, um modelo SARIMA foi eleito como melhor método preditivo apenas em 35 produtos.

No método da seleção agregada determina-se o modelo, ou a família de métodos preditivos, com melhor performance para a população como um todo. Uma vez determinado esse modelo, economiza-se o custo de aplicar os demais métodos preditivos disponíveis em cada série temporal. A fim de obter melhores resultados na seleção agregada, utilizar-se-á os *clusters* formados na etapa anterior que reúnem séries temporais de formatos semelhantes, o que se deve à capacidade do DTW de lidar com deslocamentos no tempo. Assim, buscar-se-á o melhor método preditivo para cada *cluster*, ao invés de buscar um único método para toda a população.

Como os centroides, ilustrados na Figura 19, permitem recuperar uma forma média que imita a forma dos membros dos *clusters*, pode-se utilizar a suas séries temporais para determinar o método preditivo a ser utilizado em cada grupo. A grande vantagem dessa abordagem é que a mesma limitará os testes em apenas uma única série temporal por *cluster*. Outra abordagem possível seria testar os métodos preditivos em uma amostra de cada *cluster* e, dessa forma, determinar o método com melhor performance por inferência. Quanto maior o tamanho da amostra, melhor será o resultado na segunda abordagem e também maior o seu custo computacional. Posto que o objetivo da seleção agregada é economizar processamento, optou-se nesse trabalho pela primeira abordagem, uma vez que ela é a mais econômica.

Assim, o modelo a ser utilizado em cada *cluster* pode ser definido como aquele que apresentar o menor erro de previsão na série temporal de cada centroide. Para isto, aplicou-se o conjunto de métodos preditivos em cada centroide e adotou-se os últimos seis meses de dados como período de teste. A métrica NRMSE foi empregada para agregar o erro. A Tabela 3 apresenta a primeira e a segunda família de métodos preditivos com menor NRMSE em cada centroide, assim como os respectivos NRMSEs obtidos. O Quadro 7 exhibe os hiperparâmetros escolhidos pelo programa computacional para o método com menor erro de previsão em cada centroide.

Tabela 3– Primeira e segunda família de métodos preditivos com menor NRMSE em cada centroide.

Centroide do <i>cluster</i>	Família de métodos Rank 1	NRMSE Rank 1	Família de métodos Rank 2	NRMSE Rank 2
1	Suav. Exponencial	15,2%	RNR	16,9%
2	Suav. Exponencial	12,1%	ARIMA	12,3%
3	ARIMA	13,7%	Suav. Exponencial	14,5%
4	Suav. Exponencial	9,8%	RNR	13,1%
5	Suav. Exponencial	6,8%	ARIMA	8,3%
6	Suav. Exponencial	16,4%	RNR	18,8%
7	RNR	24,5%	Suav. Exponencial	25,2%
8	Suav. Exponencial	13,8%	RNR	16,0%
9	Suav. Exponencial	52,3%	ARIMA	69,3%
10	Suav. Exponencial	12,9%	RNR	13,5%

Fonte: Elaborado pelo autor (2021).

Quadro 7 – Hiperparâmetros escolhidos para a família de métodos com menor NRMSE em cada centroide.

Centroide do <i>Cluster</i>	Família de métodos com menor NRMSE	Hiperparâmetros otimizados pelo <i>Grid Search</i>
1	Suav. Exponencial	Perfil de tendência: Sem tendência; Perfil de sazonalidade: Sem sazonalidade.
2	Suav. Exponencial	Perfil de tendência: Aditiva; Perfil de sazonalidade: Sem sazonalidade.
3	ARIMA	$p = 0; d=1; q=1.$
4	Suav. Exponencial	Perfil de tendência: Sem tendência; Perfil de sazonalidade: Aditivo; Periodicidade sazonal: 12.
5	Suav. Exponencial	Perfil de tendência: Exponencial; Perfil de sazonalidade: Aditivo; Periodicidade sazonal: 12.
6	Suav. Exponencial	Perfil de tendência: Aditiva; Perfil de sazonalidade: Sem sazonalidade.
7	RNR	1ª camada: 100 neurônios; 2ª camada: 50 neurônios; Tamanho do lote: 6.
8	Suav. Exponencial	Perfil de tendência: Sem tendência; Perfil de sazonalidade: Multiplicativo; Periodicidade sazonal: 12.
9	Suav. Exponencial	Perfil de tendência: Exponencial amortecida; Perfil de sazonalidade: Aditivo; Periodicidade sazonal: 12.
10	Suav. Exponencial	Perfil de tendência: Sem tendência; Perfil de sazonalidade: Sem sazonalidade.

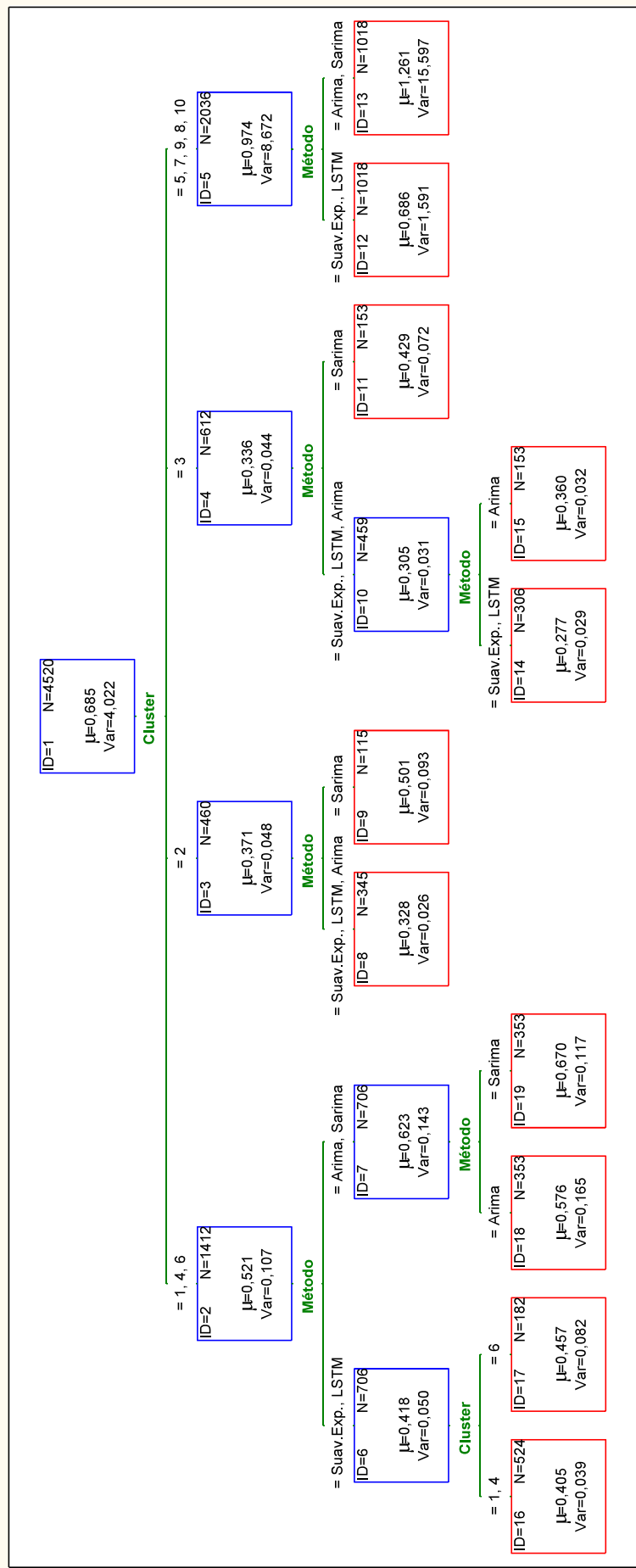
Fonte: Elaborado pelo autor (2021).

Com base nas informações apresentadas na Tabela 3, constata-se que a suavização exponencial foi eleita como a melhor família de métodos preditivos em 8 dos 10 centroides. Observa-se também que os erros no centroide do *cluster* 9 foram consideravelmente elevados em comparação com os erros obtidos nos demais centroides. Tendo em vista que esse *cluster* é

caracterizado por um crescimento exponencial nas vendas dos últimos períodos, conforme ilustrado na Figura 28, esse resultado pode indicar uma incapacidade dos métodos preditivos escolhidos em prever de maneira eficaz tal comportamento.

A escolha do método preditivo a ser aplicado cada *cluster* é uma etapa crítica na seleção agregada. Logo, torna-se desejável obter uma estimativa da assertividade dessa escolha. A partir dos resultados da seleção individual, pode-se desenvolver um modelo de árvore de regressão no qual a variável alvo é o NRMSE e as variáveis explicativas categóricas são as famílias de métodos preditivos aplicadas e o *cluster* respectivo de cada *SKU*. Para tal fim, optou-se nesse trabalho por utilizar o método CHAID. A Figura 34 apresenta o resultado dessa árvore de regressão, na qual o método “LSTM” refere-se às RNR.

Figura 34 - Árvore de regressão do NRMSE em função das famílias de métodos preditivos aplicados e do *cluster* de cada respectivo SKU na seleção individual.



Fonte: Elaborado pelo autor (2021).

Com base nas informações apresentadas na Figura 34, é possível constatar que para os *clusters* 1, 2, 4, 5, 6, 8, 9 e 10 a suavização exponencial faz parte do conjunto de métodos preditivos que proporciona a menor média de NRMSE para aquele *cluster*. Portanto, os métodos com menor erro nas séries temporais dos centroides vão de acordo com os métodos com menor erro apresentados na árvore de regressão para esses *clusters*. De forma similar, a RNR também faz parte do conjunto de métodos que proporciona menor média de NRMSE para o *cluster* 7. No entanto, observa-se que para o *cluster* 3, no qual o ARIMA foi o método escolhido pelo centroide, a Figura 34 indica que a menor média de NRMSE para esse *cluster* é obtido utilizando-se as RNRs ou a suavização exponencial. Ainda com base na árvore, é possível observar que empregar o modelo ARIMA resultará em um aumento médio de 8,3% no erro de previsão das séries do *cluster* 3.

Além da análise das médias expostas na árvore de regressão, é importante analisar as variâncias nas folhas. Observa-se que no caso dos *clusters* 5, 7, 9, 8 e 10, mesmo na folha com menor média de NRMSE, a variância do erro de previsão é expressivamente alta em comparação com a média. Isso indica que indica que os valores observados tendem a estar distantes desse valor médio. Por outro lado, nas folhas com menor erro médio para os demais *clusters*, percebe-se que a variância é relativamente baixa, o que sugere que os erros obtidos tendem a estar mais concentrados em torno da média.

Apesar do resultado discordante no *cluster* 3, pode-se considerar que, de maneira geral, os métodos preditivos com menor erro nas séries temporais do centroide fornecem uma estimativa satisfatória dos métodos preditivos que apresentam menor erro para aquele *cluster* como um todo. Assim, essa abordagem demonstra um grande potencial de aplicação, tendo em vista que o custo computacional nessa técnica é mínimo.

Concluída a validação dos métodos preditivos escolhidos, prosseguiu-se com o teste da aplicação desses métodos nas séries temporais de interesse. Nesta fase, duas abordagens distintas podem ser adotadas com relação à escolha dos hiperparâmetros. Na primeira abordagem, adota-se os hiperparâmetros escolhidos para o centroide em todas as séries temporais daquele *cluster*. A vantagem dessa abordagem é que não é necessário buscar os melhores hiperparâmetros individualmente, economizando esforço computacional. Em contrapartida, na segunda abordagem realiza-se a otimização individual dos hiperparâmetros, o que pode resultar em ganhos na acurácia de previsão. Com o intuito de decidir sobre a utilização ou não dos hiperparâmetros selecionados para o centroide nas demais séries de cada *cluster*,

efetuou-se uma análise com relação aos custos computacionais por etapa obtidos na seleção individual, onde os resultados são apresentados na Tabela 4.

Tabela 4 – Custo computacional médio por série por etapa na seleção individual.

Etapa	Tempo de processamento (segundos)
<i>Grid Search</i> na Suavização Exponencial	1,66
Treino da variação escolhida na Suavização Exponencial	0,07
Definição dos hiperparâmetros pelo AIC no ARIMA	0,67
Treino do modelo ARIMA com os hiperparâmetros escolhidos	0,04
Definição dos hiperparâmetros pelo AIC no SARIMA	1,68
Treino do modelo SARIMA com os hiperparâmetros escolhidos	0,03
<i>Grid Search</i> na Redes Neurais Recorrentes	54,87
Treino da RNR com os hiperparâmetros escolhidos	6,51

Fonte: Elaborado pelo autor (2021).

Observa-se na Tabela 4 que a maior parcela do custo computacional total na seleção individual se concentra no *Grid Search* nas RNR. Dessa forma, a fim de obter maiores economias na seleção agregada, optou-se nesse trabalho por adotar os hiperparâmetros escolhidos para o centroide nas séries do *cluster 8*, onde as RNR foram escolhidas como a família de métodos preditivos a ser empregada. Nos demais *clusters*, realizar-se-á a otimização dos hiperparâmetros individualmente.

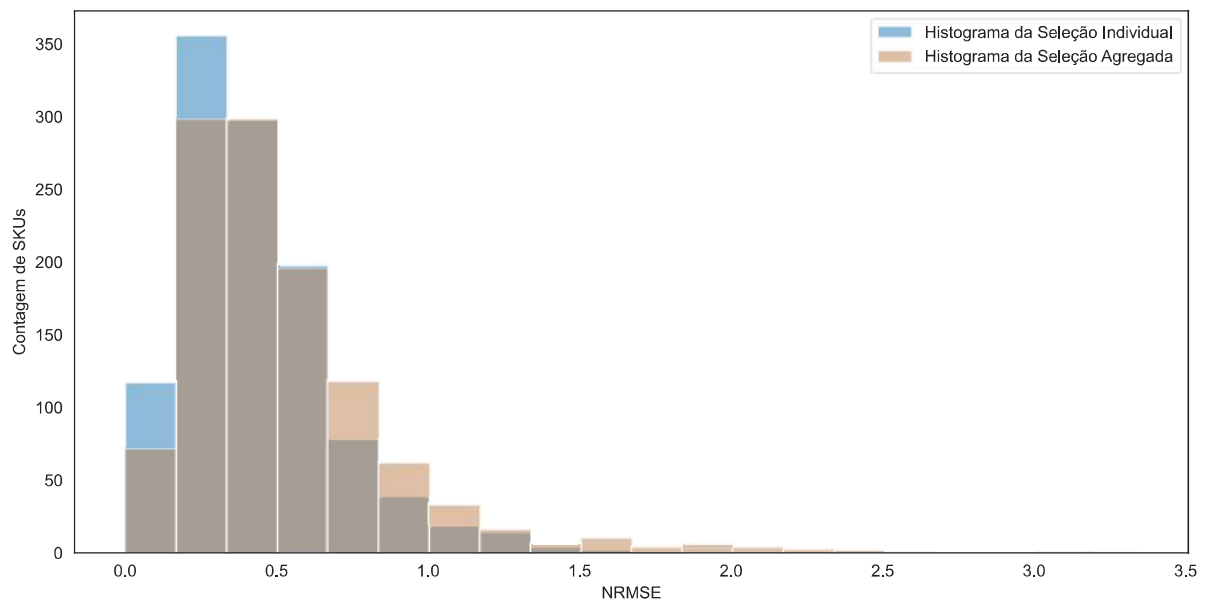
A Tabela 5 apresenta os resultados obtidos da aplicação da seleção agregada no conjunto de produtos classificado como “A”. A Figura 35 ilustra a comparação dos histogramas dos NRMSE obtidos na seleção individual e na seleção agregada.

Tabela 5 – Resultados obtidos da aplicação da seleção agregada nos 1193 produtos priorizados.

Métrica	Valor
Custo computacional total	2.493 segundos ou 41min33s
Custo computacional médio por SKU	2,09 segundos
Média dos NRMSE	52,6%
Desvio padrão dos NRMSE	36,6%
Menor NRMSE	0.0%
25º percentil dos NRMSE	29,0%
50º percentil dos NRMSE	43,0%
75º percentil dos NRMSE	64,9%
Maior NRMSE	334,2%

Fonte: Elaborado pelo autor (2021).

Figura 35 – Comparativo dos histogramas dos NRMSE obtidos na seleção vs. na seleção agregada.



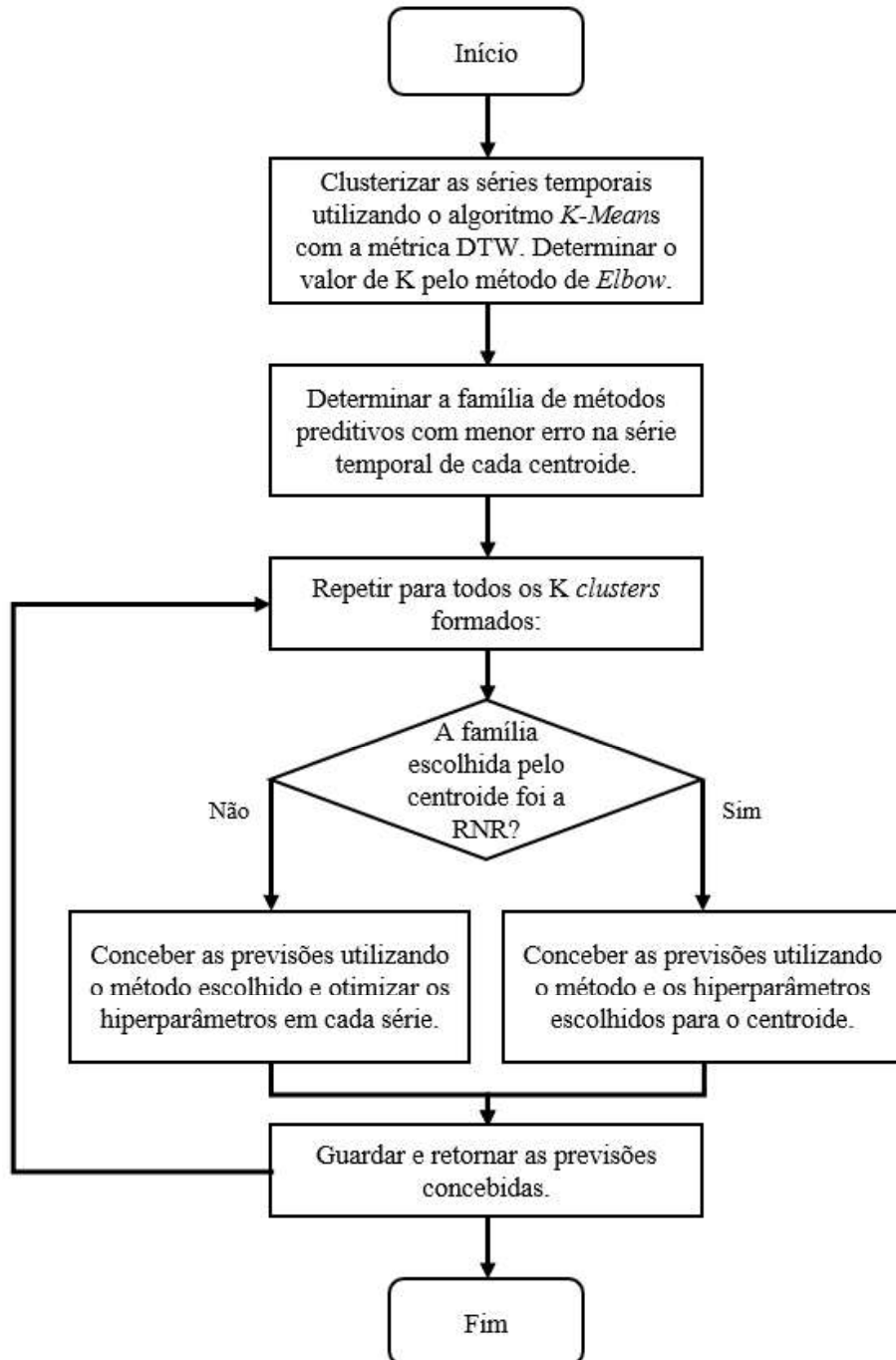
Fonte: Elaborado pelo autor (2021).

Pode-se constatar, a partir das informações apresentadas Tabela 5, que o custo computacional total na seleção agregada foi 96,7% menor do que na seleção individual, o que representa uma economia significativa. Em contrapartida, houve um aumento de 9,5% na média dos NRMSE. Analisando a Figura 35, observa-se também que há um leve aumento na frequência de produtos com NRMSE maior que 67% na seleção agregada.

A escolha da melhor abordagem deve levar em consideração o *trade-off* “Custo computacional VS. Acurácia de previsão” ponderando as necessidades de quem fará uso das

previsões. Neste trabalho, tendo em vista os resultados, optou-se por utilizar a seleção agregada. Dessa forma o modelo de previsão proposto para a empresa é resumido no fluxograma ilustrado na Figura 36.

Figura 36 – Fluxograma do modelo de previsão proposto.



Fonte: Elaborado pelo autor (2021).

Sugere-se que modelo proposto seja aplicado com frequência semanal, inserindo os dados atualizados conforme novas vendas são registradas. Apesar do modelo ser concebido para prever a demanda agregada mensalmente, adaptações simples podem ser realizadas para converter as previsões mensais em diárias. Nesse trabalho, aplicou-se o modelo proposto apenas nos produtos classificados como “A”, porém essa restrição é sugerida apenas quando há um número elevado de séries temporais a ser previsto e almeja-se custo computacional reduzido.

4.7 COMPARAÇÃO COM O MÉTODO ATUAL DA EMPRESA

Para estimar o ganho de acurácia advindo da aplicação do modelo proposto, faz-se necessário comparar suas previsões com as do modelo atualmente adotado na empresa objeto de estudo. Para embasar suas decisões de compra, a varejista utiliza atualmente a técnica de média móvel com janela de tempo fixa de seis meses para conceber as previsões de todos os *SKUs*. Uma das vantagens dessa técnica é que ela é simples e de fácil aplicação, porém é adequada apenas para dados univariados, sem tendência ou sazonalidade. A Tabela 6 apresenta o comparativo das estatísticas descritivas dos NRMSE obtidos no modelo proposto em comparação com método corrente. Para tal finalidade, utilizou-se os últimos seis meses como conjunto de dados de teste e a métrica NRMSE para agregar o erro.

Tabela 6 - Comparativo das estatísticas descritivas dos NRMSE obtidos no modelo proposto vs. método corrente nos produtos classificados como “A”.

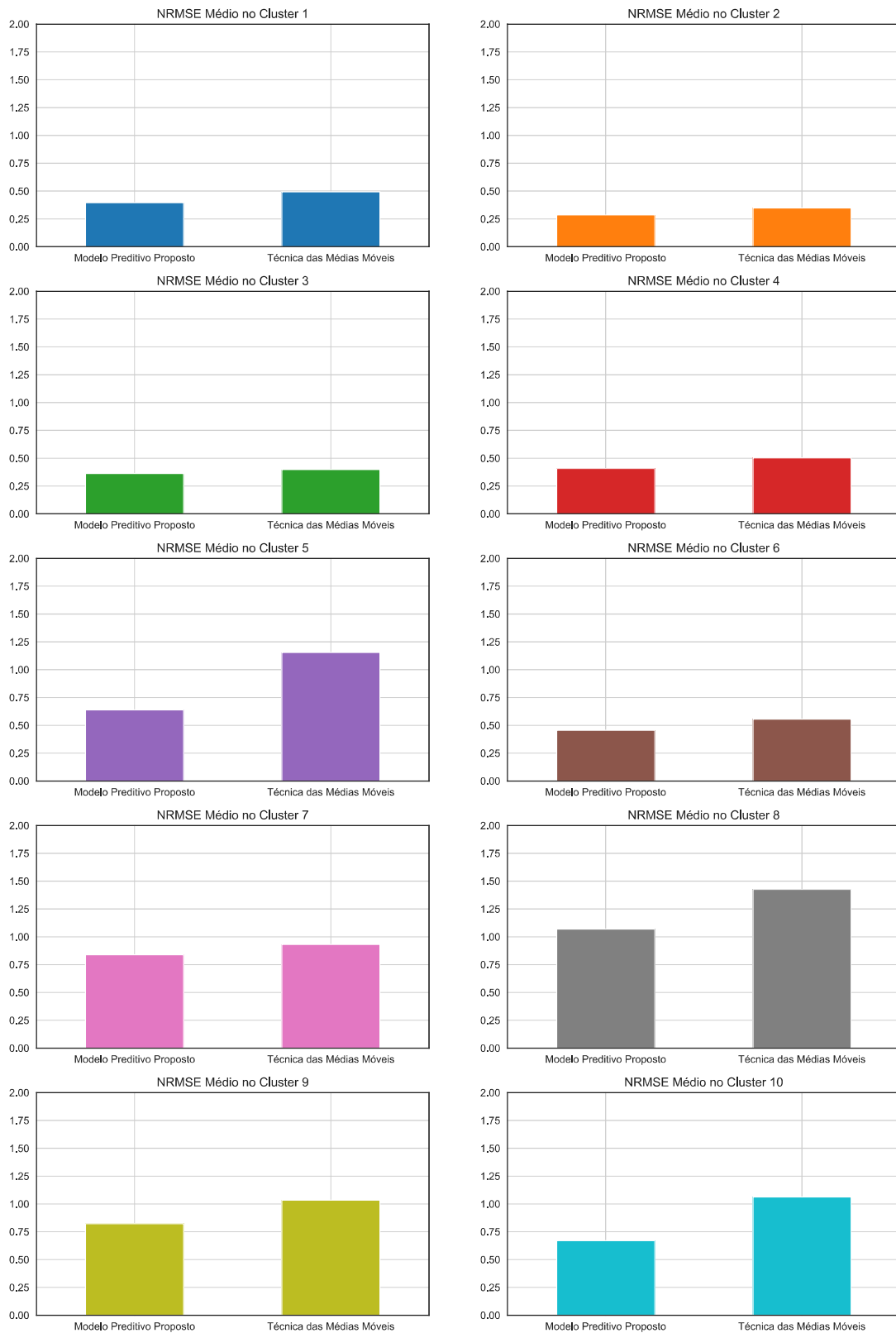
Estatística descritiva	Modelo Preditivo Proposto (%)	Técnica das Médias Móveis (%)
Média	52,6	69,3
Desvio Padrão	36,6	86,5
Menor Valor	0,00	7,7
25º percentil	29,0	35,3
50º percentil	43,0	51,6
75º percentil	64,9	81,3
Maior Valor	334,2	2122,4

Fonte: Elaborado pelo autor (2021).

Com base nos dados apresentados na Tabela 6, pode-se constatar que o NRMSE médio no modelo de previsão proposto é 16,7% menor em comparação com a técnica das médias móveis. Verifica-se também que o NRMSE máximo e o desvio padrão dos NRMSE obtidos são consideravelmente menores no modelo proposto em comparação com a técnica atual da

empresa. A Figura 37 exhibe o comparativo do NRMSE médio obtido no modelo de previsão proposto e na técnica das médias móveis por *cluster*.

Figura 37 – Comparativo do NRMSE médio obtido em cada cluster no modelo proposto vs. método corrente.



Fonte: Elaborado pelo autor (2021).

Observa-se na Figura 37 que os *clusters* 5, 8 e 10 foram os que apresentaram as maiores reduções no erro médio de previsão com a aplicação do modelo proposto. No entanto, obteve-

se nos *clusters* 7, 8 e 9 erros médios de previsão acima dos 75% no conjunto de dados de teste. Tal resultado pode indicar uma inadequação do conjunto de métodos preditivos selecionado em prever as séries temporais desses *clusters* em específico.

É interessante ressaltar que a técnica das médias móveis foi implicitamente contemplada no modelo proposto nas variações do modelo ARIMA com configuração $ARIMA(0,0,q)$, onde q representa o comprimento da janela de tempo da média móvel. Percebe-se também que caso fosse escolhido o método da seleção individual na fase anterior seria possível reduzir o erro médio em aproximadamente 26,2% em comparação com a técnica utilizada atualmente pela empresa. Em compensação, haveria um custo computacional consideravelmente maior.

5 CONCLUSÃO

Este capítulo visa trazer as considerações finais deste trabalho, no qual revisar-se-á os resultados atingidos ao longo do desenvolvimento da pesquisa e também as limitações e sugestões para futuras estudos relacionados à área.

5.1 ATENDIMENTO DOS OBJETIVOS DA PESQUISA

O presente trabalho teve como objetivo desenvolver um modelo de previsão de demanda no varejo do setor de saúde e bem-estar que melhorasse a acurácia de predição e fosse equilibrado em termos de custo computacional e desempenho. Em comparação com a técnica atualmente adotada no objeto de estudo, o modelo proposto foi capaz de reduzir o erro médio de previsão em 16,7%, valor calculado utilizando a métrica NRMSE. Verificou-se também que o erro máximo e o desvio padrão dos erros obtidos foram, respectivamente, 1788% e 50% menores no modelo sugerido. Isto posto, pode-se presumir que o modelo desenvolvido é capaz de aumentar a acurácia de previsão da empresa. Ademais, o custo médio de processamento do modelo proposto foi de apenas 2,09 segundos por *SKU*. Assim, pode-se considerar que o mesmo é equilibrado em termos de custo computacional e desempenho.

Para que se tornasse possível o alcance do objetivo geral da pesquisa, foram traçados quatro objetivos específicos. O primeiro deles, foi o de revisar a literatura sobre o assunto, a partir do qual definiu-se o roteiro metodológico. O segundo objetivo envolvia a coleta e tratamento dos dados. Este objetivo foi atingido através da extração e consolidação em formato tabular dos dados históricos de venda, conforme tópico 4.2. Por meio desses dados estimou-se que, no período de janeiro de 2016 até fevereiro de 2020, foram comercializados 14.816 produtos diferentes.

O terceiro objetivo específico era priorizar os produtos com relação à sua participação no faturamento da empresa. Esse objetivo foi concretizado no tópico 4.3, com a classificação ABC dos produtos de acordo com o princípio de Pareto. Observou-se que os 1193 produtos pertencentes à classificação “A” representaram juntos 83% do faturamento total.

Outro objetivo alcançado era escolher o conjunto de métodos preditivos a ser utilizado e implementá-lo em um programa computacional. A partir de indicações da literatura, optou-se por empregar as seguintes classes de métodos preditivos: Suavização Exponencial, Modelos ARIMA, SARIMA e Redes Neurais Recorrentes. O método de escolha e otimização dos

hiperparâmetros em cada classe foi detalhado no tópico 4.5. Subsequentemente, desenvolveu-se o programa computacional utilizando a linguagem de programação Python e pacotes de funções pré-existentes em suas bibliotecas.

O último objetivo específico atingido era definir o método preditivo a ser utilizado em cada produto e comparar as suas previsões com a técnica atual da empresa. Para tanto, ambas as abordagens, seleção individual e seleção agregada, foram testadas e comparadas no conjunto de produtos classificados como “A”, de acordo com tópico 4.6. Com o intuito de aprimorar os resultados da seleção agradável, utilizou-se o algoritmo *K-Means* com a métrica DTW para formar grupos de produtos com comportamentos históricos similares, conforme tópico 4.4. Uma vez que os centroides produzidos permitem recuperar uma forma média que imita a forma dos membros dos *clusters*, utilizou-se as suas séries temporais para determinar a classe de métodos preditivos a ser utilizada em cada agrupamento. Desta maneira, obteve-se um custo computacional 96,7% menor na seleção agregada comparado com a seleção individual, o que representou uma economia expressiva. Em contrapartida, houve um aumento de 9,5% na média dos NRMSE. Apesar do ligeiro aumento no erro de previsão, optou-se por utilizar o método da seleção agregada, tendo em vista a redução no custo computacional proporcionado.

Posteriormente, comparou-se as previsões do modelo desenvolvido com a técnica atualmente adotada no objeto de estudo, segundo tópico 4.7. Os resultados demonstraram reduções consideráveis no erro de previsão advindos da aplicação do modelo proposto. Por fim, pode-se considerar que, de maneira geral, a combinação de técnicas de clusterização de séries temporais com a abordagem da seleção agregada possibilitou reduções consideráveis no custo computacional sem aumentar em demasia o erro de previsão. Apesar do resultado discordante no *cluster* 3 em comparação com a árvore de classificação, a seleção do método preditivo para cada *cluster* através dos centroides demonstrou ser uma técnica com grande potencial de aplicação.

5.2 LIMITAÇÕES DO TRABALHO E OPORTUNIDADES E SUGESTÕES PARA FUTURAS PESQUISAS

Ainda que os objetivos desta pesquisa foram atingidos, observou-se algumas limitações ao longo da execução deste trabalho. Por intermédio dessas limitações pode-se listar oportunidades e sugestões para pesquisas futuras. A primeira a ser citada é que foi utilizada

apenas uma técnica de clusterização de séries temporais, sendo essa o algoritmo *K-Means* com a métrica DTW. Poder-se-ia utilizar outras técnicas de clusterização, métricas de similaridade de séries temporais e/ou abordagens para a escolha do número de *clusters* e analisar seus efeitos nos resultados da seleção agregada. Outro possível ponto a ser explorado em futuros trabalhos é a escolha do conjunto de métodos preditivos a serem utilizados e a estratégia de otimização de seus hiperparâmetros, tendo em vista que outras configurações podem conduzir a melhores previsões. Sugere-se também à futuras pesquisas a incorporação de variáveis exógenas nos modelos preditivos, posto que estudos empíricos conduzidos e relatados na literatura indicam que essas variáveis podem melhorar consideravelmente a acurácia das estimativas.

Ademais, futuras pesquisas podem visar a aplicação do modelo preditivo proposto em uma amostra maior de empresas de diversos portes e setores, buscando analisar a efetividade do modelo desenvolvido em diferentes cenários. Além de uma amostra maior de empresas, poder-se-ia analisar as consequências de eventos disruptivos, como a pandemia de COVID-19, nas previsões de demanda do varejo do setor de saúde e bem-estar.

Em vista dos resultados alcançados e dos procedimentos executados, pode-se considerar que o estudo ofereceu contribuições pertinentes na área abrangida, ajudando a enriquecer a literatura existente e fomentar o aprofundamento da pesquisa no tema.

Por fim, ressalta-se a importância do Trabalho de Conclusão de Curso no processo de formação do estudante, uma vez que possibilita a aplicação e domínio dos saberes adquiridos ao longo da graduação.

6 REFERÊNCIAS

- ABDOLI, G.; MEHRARA, M.; EBRAHIM ARDALANI, M. Comparing the Prediction Accuracy of Lstm and Arima Models for Time-Series With Permanent Fluctuation. **Gênero & Direito**, v. 9, n. 2, p. 314–339, 2020.
- ABEPRO. **ABEPRO - Associação Brasileira de Engenharia de Produção | A Profissão**. Disponível em: < <http://portal.abepro.org.br/a-profissao/>>. Acesso em: 25 de abril de 2021.
- AGARAP, Abien Fred. Deep learning using rectified linear units (relu). **arXiv preprint arXiv:1803.08375, 2018**.
- AGHABOZORGI, S.; SEYED SHIRKHORSHIDI, A.; YING WAH, T. Time-series Clustering: A Decade Review. **Information Systems**, v. 53, p. 16–38, 2015.
- AL-MAMARY, Y.; SHAMSUDDIN, A.; ABDUL HAMID, N. The Impact of Management Information Systems Adoption in Managerial Decision Making: A Review. **Management Information Systems**, v. 8, n. 4, p. 010–017, 2013.
- AL-SHIAB, M. The Predictability of the Amman Stock Exchange Using the Univariate Autoregressive Integrated Moving Average (ARIMA) Model. **Journal of Economic and Administrative Sciences**, v. 22, n. 2, p. 17–35, dez. 2006.
- ANASARI, R. Expectation of Tourism Demand in Iraq by Using Artificial Neural Network. **International Journal of Social Science Research and Review**, v. 2, n. 3, p. 8–17, 2019.
- ARAVANIS, T. I.; ARAVANIS, T. C. I.; PAPADOPOULOS, P. N. Fault Diagnosis in Direct Current Electric Motors Via an Artificial Neural Network. [s.l.] **Springer International Publishing**, v. 1000, 2019.
- ARZAMASTSEV, A. A.; TROITZSCH, K. G. A Parallel Algorithm for Selecting Activation Functions of an Artificial Network. **Arbeitsberichte aus dem Fachbereich Informatik**, n. 12, 2011.
- ATEV, S.; MILLER, G.; PAPANIKOLOPOULOS, N. P. Clustering of vehicle trajectories. **IEEE Transactions on Intelligent Transportation Systems**, v. 11, n. 3, p. 647–657, 2010.
- ATHANASOPOULOS, G.; HYNDMAN, R. J. Modelling and Forecasting Australian Domestic Tourism. **Tourism Management**, v. 29, n. 1, p. 19–31, 2008.
- AUNG, Yi Yi; MIN, Myat Myat. Hybrid Intrusion Detection System Using K-Means and Classification and Regression Trees Algorithms. In: **2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA)**. IEEE, 2018. p. 195-199.
- BAKAR, N. A.; ROSBI, S. Data Clustering Using Autoregressive Integrated Moving Average

(ARIMA) Model for Islamic Country Currency: An Econometrics Method for Islamic Financial Engineering. **The International Journal of Engineering and Science**, v. 06, n. 06, p. 22–31, 2017.

BATTY, M.; TORRENS, P. M. Modelling complexity : The limits to prediction. **Cybergeog**, 4 dez. 2001.

BEE DAGUM, E.; BIANCONCINI, S. Time Series Components. In: Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation. **Statistics for Social and Behavioral Sciences**. p. 29–57, 2016.

BERGMEIR, C.; BENÍTEZ, J. M. On the Use of Cross-validation for Time Series Predictor Evaluation. **Information Sciences**, v. 191, p. 192–213, 2012.

BORGES, Admir Roberto et al. **Marketing de varejo: as estratégias adotadas pelos supermercados de vizinhança para conquistar e fidelizar clientes**. [s.l.] Universidade Federal dDe Santa Catarina, 2001.

BOTCHKAREV, A. Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio. **SSRN Electronic Journal**, p. 1–16, 2018.

BROCKWELL, P. J.; DAVIS, R. A. **Time Series: Theory and Methods**. New York: Springer New York, 1991.

BROWN, R. G.; MEYER, R. F. The Fundamental Theorem of Exponential Smoothing. **Operations Research**, v. 9, n. 5, p. 673–685, 1961.

BROWNE, M. W. Cross-Validation Methods. **Journal of mathematical psychology**, v. 44, n. 1, p. 108-132, 2000..

CANIATO, F.; KALCHSCHMIDT, M.; RONCHI, S. Integrating Quantitative and Qualitative Forecasting Approaches: Organizational Learning in an Action Research Case. **Journal of the Operational Research Society**, v. 62, n. 3, p. 413–424, 2011.

CAO, M.-T. et al. Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. **Advanced Engineering Informatics**, v. 46, p. 101182, out. 2020.

CARVALHO, Joao Paulo. On the semantics and the use of fuzzy cognitive maps in social sciences. In: **International Conference on Fuzzy Systems**. IEEE, 2010. p. 1-6.

CAUCHICK MIGUEL, P. A. Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução. **Production**, v. 17, n. 1, p. 216–229, 2007.

CAWLEY, G. C.; TALBOT, N. L. C. On Over-Fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. **Journal of Machine Learning Research**, v. 11, p. 2079–2107, 2010.

- CHAI, T.; OCEANIC, N. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE in the Literature. n. June 2014, p. 3–7, 2015.
- CHANG, Zihan; ZHANG, Yang; CHEN, Wenbo. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. **Energy**, v. 187, p. 115804, 2019.
- CHAOJI, V.; RASTOGI, R.; ROY, G. Machine Learning in the Real World. **Proceedings of the VLDB Endowment**, v. 9, n. 13, p. 1597–1600, 2015.
- CHEN, C. F.; CHANG, Y. H.; CHANG, Y. W. Seasonal ARIMA Forecasting of Inbound Air Travel Arrivals to Taiwan. **Transportmetrica**, v. 5, n. 2, p. 125–140, 2009.
- CHU, F. L. Forecasting Tourism Demand in Asian-Pacific Countries. **Annals of Tourism Research**, v. 25, n. 3, p. 597–615, 1998.
- CORDEIRO, C.; NEVES, M. M. How Bootstrap Can Help in Forecasting Time Series with More than One Seasonal Pattern. **AIP Conference Proceedings**, v. 1479, n. 1, p. 1712–1715, 2012.
- COX, J. F.; BLACKSTONE, J. H. **APICS dictionary**. [s.l.] Amer Production & Inventory, 2002.
- DAMON, J.; GUILLAS, S. The Inclusion of Exogenous Variables in Functional Autoregressive ozone forecasting. **Environmetrics**, v. 13, n. 7, p. 759–774, 2002.
- DANGERFIELD, B. J.; MORRIS, J. S. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. **International Journal of Forecasting**, v. 8, n. 2, p. 233–241, 1992.
- DEMATOS, G. et al. Feedforward versus recurrent neural networks for forecasting monthly japanese yen exchange rates. **Financial Engineering and the Japanese Markets**, v. 3, n. 1, p. 59–75, 1996.
- DIEBOLD, F. X.; LOPEZ, J. A. 8 Forecast evaluation and combination. **Handbook of Statistics**, v. 14, p. 241–268, 1996.
- DU, S. et al. Traffic flow forecasting based on hybrid deep learning framework. **Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2017**, v. 2018- Janua, p. 1–6, 2017.
- DUFOUR, L. et al. Test set validation for home electrical signal disaggregation. **Proceedings - 2014 8th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2014**, p. 415–420, 2014.
- DZIUBA, S.; KADŁUBEK, M.; INGALDI, M. Selected Problems of Management of the

Supply Subsystem in the Enterprise of the Foundry Sector: Determination of the Purchasing Sources and the Levels of Unit Material Consumption. **METAL 2017 - 26th International Conference on Metallurgy and Materials, Conference Proceedings**, v. 2017- Janua, p. 2115–2121, 2017.

EHRENTHAL, J. C. F.; HONHON, D.; VAN WOENSEL, T. Demand Seasonality in Retail Inventory Management. **European Journal of Operational Research**, v. 238, n. 2, p. 527–539, 2014.

FALK, M. et al. **A First Course on Time Series Analysis-Examples with SAS**. [s.l: s.n.].

FANG, Q. et al. Modeling Nitrogen and Water Management Effects in a Wheat-Maize Double-Cropping System. **Journal of Environmental Quality**, v. 37, n. 6, p. 2232–2242, 2008.

FILDES, R. Evaluation of Aggregate and Individual Forecast Method Selection Rules. **Management Science**, v. 35, n. 9, p. 1056–1065, 1989.

FILDES, R.; KINGSMAN, B. Incorporating Demand Uncertainty and Forecast Error in Supply Chain Planning Models. **Journal of the Operational Research Society**, v. 62, n. 3, p. 483–500, 2011.

FILDES, R.; PETROPOULOS, F. Simple Versus Complex Selection Rules for Forecasting Many Time Series. **Journal of Business Research**, v. 68, n. 8, p. 1692–1701, 2015.

FLORES, B. E. A Pragmatic View of Accuracy Measurement in Forecasting. **Omega**, v. 14, n. 2, p. 93–98, 1986

GAL, Yarin; GHARAMANI, Zoubin. A theoretically grounded application of dropout in recurrent neural networks. **Advances in neural information processing systems**, v. 29, p. 1019-1027, 2016.

GARBADE, M. J. **Understanding K-means Clustering in Machine Learning**.

GARDNER, E. S. Exponential Smoothing: The State of the Art. v. 4, n. August 1984, p. 1–28, 1985.

GARDNER, E. S. Exponential smoothing: The state of the art-Part II. **International Journal of Forecasting**, v. 22, n. 4, p. 637–666, 2006.

GENÇAY, R.; LIU, T. Nonlinear Modelling and Prediction with Feedforward and Recurrent networks. **Physica D: Nonlinear Phenomena**, v. 108, n. 1–2, p. 119–134, 1997.

GIL, A. C. (ORG). **Delineamento da Pesquisa**. [s.l: s.n.]. v. 264

GRAVES, A. **Supervised Sequence Labelling with Recurrent Neural Networks**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. v. 385

GOYAL, M.; VOHRA, R. Applications of Data Mining in Higher Education. **International Journal of Computer Science Issues**, v. 9, n. 2, p. 113–120, 2012.

HANSUN, S. A New Approach of Brown's Double Exponential Smoothing Method in Time Series Analysis. **Balkan Journal of Electrical and Computer Engineering**, v. 4, n. 2, 2016.

HBNER, A. H.; KUHN, H.; STERNBECK, M. G. Demand and Supply Chain Planning in Grocery Retail: An Operations Planning Framework. **International Journal of Retail & Distribution Management**, v. 41, n. 7, p. 512–530, 2013.

HERTZ, J. A. **Introduction to the Theory of Neural Computation**. [s.l.] CRC Press, 2018.

HOLT, C. C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. **International Journal of Forecasting**, v. 20, n. 1, p. 5–10, 2004.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**.

ISLAM, M. A. et al. Energy demand forecasting. In: **Energy for Sustainable Development**. [s.l.] Elsevier, 2020. p. 105–123.

IZAKIAN, H.; PEDRYCZ, W.; JAMAL, I. Fuzzy Clustering of Time Series Data Using Dynamic Time Warping Distance. **Engineering Applications of Artificial Intelligence**, v. 39, p. 235–244, 2015.

JACOBS, F. Robert; CHASE, Richard B.; LUMMUS, Rhonda R. **Operations and supply chain management**. New York, NY: McGraw-Hill/Irwin, 2014.

Jahr, K. et al. Decentralized Autonomous Fault Detection in Wireless Structural Health Monitoring Systems Using Structural Response Data. n. July, p. 20–22, 2015.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial Neural Networks: A tutorial. **Computer**, v. 29, n. 3, p. 31–44, 1996.

JANG, M. et al. Dynamic Time Warping-Based K-Means Clustering for Accelerometer-Based Handwriting Recognition. **Studies in Computational Intelligence**, v. 363, p. 21–26, 2011.

JEFFARES, A. **K-means: A Complete Introduction**. Towards data science, 2019. Disponível em: <<https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c/>>. Acesso em: 20 de maio de 2021.

JOZEFOWICZ, R.; ZAREMBA, W.; SUTSKEVER, I. An Empirical Exploration of Recurrent Network Architectures. **32nd International Conference on Machine Learning, ICML 2015**, v. 3, p. 2332–2340, 2015.

JUNIOR, Albino Mileski. Análise de métodos de previsão de demanda baseados em series temporais em uma empresa do sector de perfumes e cosméticos. 2007.

KANG, Yanfei et al. D\ej'a vu: forecasting with similarity. arXiv preprint arXiv:1909.00221, 2019.

KALEKAR, P. Time Series Forecasting Using Holt-Winters Exponential Smoothing. **Kanwal Rekhi School of Information Technology**, n. 04329008, p. 1–13, 2004. KALEKAR, P. Time series forecasting using Holt-Winters exponential smoothing. **Kanwal Rekhi School of Information Technology**, n. 04329008, p. 1–13, 2004.

KANUNGO, T. et al. An Efficient K-means Clustering Algorithms: Analysis and Implementation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 7, p. 881–892, jul. 2002.

KASSAMBARA, A. **Determining The Optimal Number Of Clusters: 3 Must Know Methods.**

KASTURI, S. N. **Underfitting and Overfitting in machine learning and how to deal with it !!!**

KASTURI, S. N. **Underfitting and Overfitting in machine learning and how to deal** KAUR, M.; GULATI, H.; KUNDRA, H. Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications. **International Journal of Computer Applications**, v. 99, n. 12, p. 1–3, 2014.

KIM, H. S. Development of Seismic Response Simulation Model for Building Structures with Semi-active Control Devices Using Recurrent Neural Network. **Applied Sciences (Switzerland)**, v. 10, n. 11, 2020.

KODINARIYA, T. M.; MAKWANA, P. R. Review on Determining Number of Cluster in K-Means Clustering. **International Journal of Advance Research in Computer Science and Management Studies**, v. 1, n. 6, p. 2321–7782, 2013.

LEE, C. K. M.; PALANIAPPAN, S. Effective Asset Management for Hospitals with RFID. **2014 IEEE International Technology Management Conference, ITMC 2014**, 2014.

LI, H. et al. Adaptively Constrained Dynamic Time Warping for Time Series Classification and Clustering. **Information Sciences**, v. 534, p. 97–116, 2020.

LI, P.; SUN, M.; PANG, M. **Prediction of Taxi Demand Based on ConvLstm Neural Network.** [s.l.] Springer International Publishing, 2018. v. 11305 LNCS

LI, S.; LI, R. Comparison of Forecasting Energy Consumption in Shandong, China Using the ARIMA Model, GM Model, and ARIMA-GM Model. **Sustainability (Switzerland)**, v. 9, n. 7, 2017.

LIIV, I. Inventory Classification Enhancement with Demand Associations. **2006 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2006**, p. 18–22, 2006.

LIPOVETSKY, S. Pareto 80/20 Law: Derivation Via Random Partitioning. **International Journal of Mathematical Education in Science and Technology**, v. 40, n. 2, p. 271–277, mar. 2009.

LIU, L.; LIU, J. Reconstructing gene Regulatory Networks Via Memetic Algorithm and LASSO Based on Recurrent Neural Networks. **Soft Computing**, v. 24, n. 6, p. 4205–4221, 2020.

LOH, W. K.; MANE, S.; SRIVASTAVA, J. Mining temporal patterns in popularity of web items. **Information Sciences**, v. 181, n. 22, p. 5010–5028, 2011.

LOH, W. Y. Fifty years of classification and regression trees. **International Statistical Review**, v. 82, n. 3, p. 329–348, 2014.

LU, Z.; SINDHWANI, V.; SAINATH, T. N. **Learning compact recurrent neural networks**. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). **Anais...IEEE**, mar. 2016Disponível em: <<http://ieeexplore.ieee.org/document/7472821/>>

LYU, Q. et al. Modelling high-dimensional sequences with LSTM-RTRBM: Application to polyphonic music generation. **IJCAI International Joint Conference on Artificial Intelligence**, v. 2015-Janua, n. Ijcai, p. 4138–4139, 2015.

MAHARAJ, Elizabeth Ann. Cluster of time series. **Journal of Classification**, v. 17, n. 2, p. 297-314, 2000.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. **Forecasting: Methods and Applications, 3rd Edition**. [s.l: s.n.].

MCGOUGH, T.; TSOLACOS, S. Forecasting Commercial Rental Values Using ARIMA Models. **Journal of Property Valuation and Investment**, v. 13, n. 5, p. 6–22, 1995.

MENTZAS, J.; LINARDOPOULOS, I.; ASSIMAKOPOULOS, V. An architecture for intelligent assistance in the forecasting process. **Proceedings of the Annual Hawaii International Conference on System Sciences**, v. 3, p. 167–176, 1995.

MILANOVIĆ, M.; STAMENKOVIĆ, M. CHAID Decision Tree: Methodological Frame and Application. **Economic Themes**, v. 54, n. 4, p. 563–586, 2016.

MESAROS, A. et al. SOUND EVENT ENVELOPE ESTIMATION IN POLYPHONIC MIXTURES Irene Mart ´ Department of Computer Science , Universitat de Val ` encia Laboratory of Signal Processing , Tampere University of Technology. p. 935–939, 2019.

MITRAN, T.; MEENA, R. S.; CHAKRABORTY, A. **Geospatial Technologies for Crops and Soils: An Overview**. [s.l: s.n.].

MONDAL, Prapanna; SHIT, Labani; GOSWAMI, Saptarsi. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. **International Journal of Computer Science, Engineering and Applications**, v. 4, n. 2, p. 13, 2014.

MONIKA, K.; GUPTA, S.; SUSHIL KUMAR, S. Comparative Analysis of Fraud Detection

Technique on Automobile Insurance Claims Using Different Model Evaluation Method. **Alochana Chakra Journal**, 2020.

MORAN, K. R. et al. Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. **Journal of Infectious Diseases**, v. 214, n. Suppl 4, p. S404–S408, 2016.

MÜLLER, M. Chapter 4: Dynamic Time Warping. **Information Retrieval for Music and Motion**, p. 69–84, 2007.

MYUNG, I. J. Tutorial On Maximum Likelihood Estimation. **Journal of Mathematical Psychology**, v. 47, n. 1, p. 90–100, 2003.

OANCEA, B.; CIUCU, ŞT. C. Time Series Forecasting Using Neural Networks. **Proceedings of the International Conference on APL : The Language and its Applications, APL 1994**, p. 86–94, jan. 2014.

OGASAWARA, E. et al. Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time Series. **Proceedings of the International Joint Conference on Neural Networks**, 2010.

ORRÙ, G. et al. Machine Learning Increases Diagnosticity in Psychometric Evaluation of Alexithymia in Fibromyalgia. **Frontiers in Medicine**, v. 6, n. January, p. 1–7, 2020.

OSTERTAGOVÁ, E.; OSTERTAG, O. Forecasting Using Simple Exponential Smoothing Method. **Acta Electrotechnica et Informatica**, v. 12, n. 3, p. 62–66, 2013.

Overfitting and Underfitting. fast.ai reference, [s/d]. Disponível em: <<https://www.fastaireference.com/overfitting/>>. Acesso em: 21 de março de 2021.

OZAKI, T. On the Order Determination of ARIMA Models. **Applied Statistics**, v. 26, n. 3, p. 290, 1977.

OZAKI, T. On the Order Determination of ARIMA Models. **Applied Statistics**, v. 26, n. 3, p. 290, 1977.

PADILHA, T. C. C.; MARINS, F. A. S. Sistemas ERP: características, custos e tendências. **Production**, v. 15, n. 1, p. 102–113, 2005.

PATRO, S. G. K.; SAHU, K. K. Normalization: A Preprocessing Stage. **Iarjset**, p. 20–22, 2015.

PAUL, S. et al. Comparison between Conventional Network and ANN with Case Study. **International Research Journal of Engineering and Technology(IRJET)**, v. 4, n. 8, p. 1795–1803, 2017.

PERMANASARI, A. E.; HIDAYAH, I.; BUSTONI, I. A. SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. **Proceedings - 2013 International Conference on Information Technology and Electrical Engineering: “Intelligent and Green Technologies for Sustainable Development”**, ICITEE 2013, n. 2, p. 203–207, 2013.

PURNIMA, B.; ARVIND, K. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. **International Journal of Computer Applications**, v. 105, n. 9, p. 17–24, 2014.

RACCA, Enrica. **Neural networks in optical domain**. 2019. Tese de Doutorado. Politecnico di Torino.

RAI, P.; SINGH, S. A Survey of Clustering Techniques. **International Journal of Computer Applications**, v. 7, n. 12, p. 1–5, 2010.

RAJU, V. N. G. et al. Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. **Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020**, n. Icssit, p. 729–735, 2020.

SAKOE, H.; CHIBA, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 26, n. 1, p. 43–49, fev. 1978.

SALAM, A.; PANAHIFAR, F.; BYRNE, P. J. Retail supply chain service levels: the role of inventory storage. **Journal of Enterprise Information Management**, v. 29, n. 6, p. 887–902, 2016.

SAMADI, S.; WILSON, C. A. M. E.; MORADKHANI, H. Uncertainty analysis of statistical downscaling models using Hadley Centre Coupled Model. **Theoretical and Applied Climatology**, v. 114, n. 3–4, p. 673–690, 2013.

SARANYA, C.; MANIKANDAN, G. A study on normalization techniques for privacy preserving data mining. **International Journal of Engineering and Technology**, v. 5, n. 3, p. 2701–2704, 2013.

SATYANARAYANA, N.; RAMALINGASWAMY, C.; RAMADEVI, Y. Survey of Classification Techniques in Data Mining. **IJSET -International Journal of Innovative Science, Engineering & Technology**, v. 1, n. 9, p. 268–278, 2014.

SCHWARZKOPF, A. B.; TERSINE, R. J.; MORRIS, J. S. Top-down versus bottom-up forecasting strategies. **International Journal of Production Research**, v. 26, n. 11, p. 1833–1843, 1988.

SCHOEMAKER, Paul JH et al. Forecasting and scenario planning: The challenges of uncertainty and complexity. **Blackwell handbook of judgment and decision making**, p. 274–296, 2004.

SHAADAN, N. et al. Time Series Model for Carbon Monoxide (Co) At. v. 4, n. 2, p. 246–260, 2019.

SHAH, I. et al. Forecasting One-Day-Ahead Electricity Prices for Italian Electricity Market Using Parametric and Nonparametric Approaches. **IEEE Access**, v. 8, p. 123104–123113,

2020.

SHAHABUDDIN, Syed. Why Forecasts Are Wrong?. **The Journal of Business Forecasting**, v. 6, n. 3, p. 16, 1987.

SHALABI, L. AL; SHAABAN, Z.; KASASBEH, B. Data Mining : A Preprocessing Engine Luai Al Shalabi , Zyad Shaaban and Basel Kasasbeh Applied Science University , Amman , Jordan. **Journal of Computer Science**, v. 2, n. 9, p. 735–739, 2006.

SHANKAR, R. Shiva et al. Prediction of employee attrition using datamining. In: **2018 iee international conference on system, computation, automation and networking (icscan)**. IEEE, 2018. p. 1-8.

SHCHERBAKOV, M. V. et al. A survey of forecast error measures. **World Applied Sciences Journal**, v. 24, n. 24, p. 171–176, 2013.

SIAMI-NAMINI, S.; TAVAKOLI, N.; SIAMI NAMIN, A. A Comparison of ARIMA and LSTM in Forecasting Time Series. **Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018**, p. 1394–1401, 2019.

SILVA, D. E. L. da; MENEZES, M. E. M. Metodologia da Pesquisa e Elaboração de Dissertação. **Portal**, v. 29, n. 1, p. 121, 2005.

SILVER, E. A. Operations Research in Inventory Management: a Review and Critique. **Operations Research**, v. 29, n. 4, p. 628–645, 1981.

SILVER, E. A. Inventory management: An overview, Canadian publications, practical applications and suggestions for future research. **Infor**, v. 46, n. 1 SPEC. ISS., p. 15–28, 2008.

SINGH, S. **How to Report Forecast Accuracy to Management.**

SNAPP, S. **How MAE is Calculated for Forecast Error Measurement.**

STACHEWSKI, A. L. **Varejo: micro, pequenas e médias empresas foram as que mais sofreram com a crise - Pequenas Empresas Grandes Negócios | Varejo.**

SYNTETOS, A. A.; KEYES, M.; BABAI, M. Z. Demand categorisation in a European spare parts logistics network. **International Journal of Operations and Production Management**, v. 29, n. 3, p. 292–316, 2009.

TAMASHIRO, H. R. DA S.; MERLO, E. M.; DA SILVEIRA, J. A. G. Comportamento do Consumidor e os Atributos que Sinalizam as Preferências dos Formatos de Varejo no Setor de Higiene Pessoal, Perfumaria e Cosméticos. **Revista Brasileira de Marketing**, v. 10, n. 2, p. 54–83, 2011.

THOMPSON, P. A. An MSE statistic for comparing forecast accuracy across series. **International Journal of Forecasting**, v. 6, n. 2, p. 219–227, 1990.

VALLES, D. Early Warning Embedded System of Dangerous Temperature Using Single Early

Warning Embedded System of Dangerous Temperature Using Single exponential smoothing for Firefighters Safety. n. August, p. 41–44, 2018.

VAN DER LAAN, Mark J.; GRUBER, Susan. Collaborative double robust targeted maximum likelihood estimation. **The international journal of biostatistics**, v. 6, n. 1, 2010.

VAN WIJK, J. J.; VAN SELOW, E. R. Cluster and calendar based visualization of time series data. **Proceedings of the IEEE Symposium on Information Visualization**, p. 4–9, 1999.

VAROTTO, L. F. História do varejo. **GV-executivo**, v. 5, n. 1, p. 86, 2006.

VERRUCK, F.; BAMPI, R. E.; MILAN, G. S. Previsão de Demanda em Operações de Serviço: Um Estudo em Uma Empresa do Setor de Transportes. **Simpósio de Administração da Produção Logística e Operações Internacionais**, 2009.

WHITE, B. FORECASTING MILK OUTPUT IN ENGLAND AND WALES*. **Journal of Agricultural Economics**, v. 38, n. 2, p. 223–234, maio 1987.

WIDIARTA, H.; VISWANATHAN, S.; PIPLANI, R. Forecasting aggregate demand: An analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. **International Journal of Production Economics**, v. 118, n. 1, p. 87–94, 2009.

WU, L.; PERIN, G.; PICEK, S. I Choose You: Automated Hyperparameter Tuning for Deep Learning-based Side-channel Analysis. **Cryptology ePrint Archive**, n. Report 2020/1293, p. 1–23, 2020.

WU, Z. et al. Hybrid improved empirical mode decomposition and BP neural network model for the prediction of sea surface temperature. **Ocean Science**, v. 15, n. 2, p. 349–360, 2019.

YANG, C. T. et al. An implementation of cloud-based platform with R packages for spatiotemporal analysis of air pollution. **Journal of Supercomputing**, v. 76, n. 3, p. 1416–1437, 2020.

YANG, S. N.; CHANG, L. C. Regional inundation forecasting using machine learning techniques with the internet of things. **Water (Switzerland)**, v. 12, n. 6, 2020.

YOOSEFDOOST, A. et al. Comparison between Performance of Statistical and Low Cost ARIMA Model with GFDL, CM2.1 and CGM 3 Atmosphere-Ocean General Circulation Models in Assessment of the Effects of Climate Change on Temperature and Precipitation in Taleghan Basin. **American Journal of Water Resources**, v. 5, n. 4, p. 92–99, 2017.

YUAN, C.; YANG, H. Research on K-Value Selection Method of K-Means Clustering Algorithm. **J**, v. 2, n. 2, p. 226–235, 2019.

ZAREMBA, Wojciech; SUTSKEVER, Ilya; VINYALS, Oriol. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329, 2014.

ZOU, J.; HAN, Y.; SO, S. S. Overview of Artificial Neural Networks. **Methods in Molecular Biology**, v. 458, p. 15–23, 2008.