

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

Augusto Hochleitner

Análise comparativa entre métodos estatísticos tradicionais e baseados em inteligência artificial no problema de previsão de demanda

Florianópolis

2021

Augusto Hochleitner

Análise comparativa entre métodos estatísticos tradicionais e baseados em inteligência artificial no problema de previsão de demanda

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

Orientador: Prof. Ricardo Faria Giglio, Dr.

Florianópolis

2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Hochleitner, Augusto

Análise comparativa entre métodos estatísticos tradicionais e baseados em inteligência artificial no problema de previsão de demanda / Augusto Hochleitner ; orientador, Ricardo Faria Giglio, 2021.

94 p.

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

Inclui referências.

1. Engenharia de Produção Mecânica. 2. Previsão de
demanda . 3. Séries temporais. 4. Aprendizado de máquina.
I. Faria Giglio, Ricardo . II. Universidade Federal de
Santa Catarina. Graduação em Engenharia de Produção Mecânica.
III. Título.

Augusto Hochleitner

Análise comparativa entre métodos estatísticos tradicionais e baseados em inteligência artificial no problema de previsão de demanda

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de “Engenheiro Mecânico com habilitação em Produção” e aprovado em sua forma final pelo Curso de Engenharia de Produção Mecânica

Florianópolis, 09 de setembro de 2021.

Prof.(a) Mônica Maria Mendes Luna, Dra.
Coordenadora do Curso

Banca Examinadora:

Prof. Ricardo Faria Giglio, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. Eduardo Ferreira da Silva, Dr.
Avaliador
Universidade Federal de Santa Catarina

Prof. Sérgio Fernando Mayerle, Dr.
Avaliador
Universidade Federal de Santa Catarina

Este trabalho é dedicado aos meus colegas de classe e aos meus queridos pais.

AGRADECIMENTOS

Agradeço aos meus pais, Jan e Rosana, que sempre me apoiaram e deram todas as condições para que eu me dedicasse aos estudos.

Ao meu orientador, Ricardo Faria Giglio, por me apresentar a área do trabalho e apoiar meus estudos acerca do tema, participando ativamente durante toda a execução do trabalho.

Aos professores, pelas correções e ensinamentos que me permitiram apresentar um melhor desempenho no meu processo de formação profissional ao longo do curso.

Aos meus colegas de curso, com quem convivi intensamente durante os últimos anos, pelo companheirismo e pela troca de experiências que me permitiram crescer não só como pessoa, mas também como formando.

A Universidade Federal de Santa Catarina, essencial no meu processo de formação profissional, por tudo o que aprendi ao longo dos anos do curso.

RESUMO

A previsão de séries temporais é uma questão que se mostra fundamental para os mais diversos setores, pois pode facilitar tomadas de decisão estratégicas diversas. Aspectos como a gestão dos estoques, da cadeia logística de suprimentos e do planejamento e controle da produção de uma empresa são alguns exemplos de processos que se beneficiam amplamente por uma previsão de demanda adequada. Por conta disso, métodos de previsão de séries temporais baseados no uso de inteligência artificial estão em acelerada ascensão, com a premissa que serão mais efetivos que os métodos estatísticos tradicionais. A partir desta premissa este trabalho se propõe a comparar uma seleção de modelos quanto a capacidade de predição e a acurácia, em um cenário específico, que inclui a previsão de múltiplas séries temporais paralelas, uma situação em que os modelos de inteligência artificial tendem a ser superiores. Para realizar estas análises, um total de sete modelos de previsão de demanda foram aplicados em dois conjuntos de dados, e cada conjunto foi analisado em dois níveis de agregação diferentes, a fim de avaliar a capacidade dos modelos na previsão de uma série temporal simples e de séries paralelas, em um cenário com forte sazonalidade e outro com baixa sazonalidade. Foram selecionados os métodos ARIMA, Prophet e Holt-Winters para representar os métodos estatísticos tradicionais, os métodos Decision Tree, Random Forest, XGBoost para representar os métodos de inteligência artificial e uma rede neural convolucional como método mais moderno. O resultado obtido ao fim do trabalho foi que a rede neural apresentou o melhor resultado para a previsão de uma única série temporal com alto nível de agregação mas obteve um baixo desempenho para a previsão de múltiplas séries paralelas, cenário em que os modelos Random Forest e XGBoost superaram os demais por uma margem considerável. Quanto à comparação entre os modelos tradicionais e os modelos de inteligência artificial, concluiu-se que, para os dados do trabalho, o segundo grupo superou o primeiro em todos os cenários, que apresentou erros pelo menos 35% maiores na maioria dos casos.

Palavras-chave: Previsão de demanda. Séries temporais. Aprendizado de máquina.

ABSTRACT

Time series forecasting is an important problem to various sectors, for the help it can provide to strategic decisions. Aspects such as the management of inventories, supply chains and the production planning and control are favored from an adequate demand forecast. For that reason, time series forecasting methods based on artificial intelligence are in fast development, with the promise to be more effective than the traditional statistical methods. Based on that promise, this work proposes to compare a selection of models as to which offers the best prediction in a specific scenario, that includes the prediction of multiple parallel time series, a situation where the traditional models tend to be ineffective. To realize this analysis seven models were selected and used to predict two databases, with two different aggregation levels for each database, aiming to analyze the models capacity to predict a simple time series, as well as multiple parallel time series, in scenarios with both strong and weak seasonality. The traditional methods selected were ARIMA, Holt-Winters and Prophet, and the artificial intelligence methods selected were Decision Tree, Random Forest, XGBoost and a neural network, as the most modern method. The result from this work was that the neural network had the best result for the scenarios of a single time series with high aggregation level, but had a high error on the scenarios of multiple time series, in which the others machine learning methods overcame the others by a considerable margin. As for the comparison between the tradicional and the machine learning methods, the conclusion is that, for the databases used, the latter overcame the first in all scenarios, with errors at least 35% lower in most cases.

Keywords: Demand Forecasting. Time series. Machine Learning.

LISTA DE FIGURAS

Figura 1 - PCP nas empresas.	22
Figura 2 - Processo de implementação de um modelo de previsão de demanda.	24
Figura 3 - Fatores de uma série temporal.	25
Figura 4 - Efeitos de <i>overfitting</i> e <i>underfitting</i>	28
Figura 5 - Exemplo de árvore de decisão.	34
Figura 6 - Estrutura de uma rede neural.	38
Figura 7 - Etapas da pesquisa.	41
Figura 8- Tabela das vendas Rossmann.	43
Figura 9 - Tabela das lojas Rossmann.	44
Figura 10 - Valores históricos de venda da rede Rossmann.	46
Figura 11 - Decomposição série temporal das vendas da rede Rossmann.	47
Figura 12 - Auto correlação vendas Rossmann.	48
Figura 13 - Valores históricos de venda Rossmann (3 maiores lojas).	49
Figura 14 - Histograma dos coeficientes de correlação das lojas Rossmann.	50
Figura 15 - Valores históricos de venda da empresa têxtil.	51
Figura 16 - Decomposição série temporal das vendas da empresa têxtil.	52
Figura 17 - Auto correlação vendas têxtil.	53
Figura 18 - Valores históricos de venda da empresa têxtil (3 maiores pares grupo-região). ...	54
Figura 19 - Histograma dos coeficientes de correlação entre os pares grupo-região.	55
Figura 20 - Previsão Rossmann agregada ARIMA.	57
Figura 21 - Previsão Rossmann segregada ARIMA.	58
Figura 22 - Previsão têxtil agregada ARIMA.	59
Figura 23 - Previsão têxtil segregada ARIMA.	60
Figura 24 - Previsão têxtil agregada Holt-Winters.	61
Figura 25 - Previsão têxtil segregada Holt-Winters.	62
Figura 26 - Previsão Rossmann agregada Prophet.	63
Figura 27 - Previsão Rossmann segregada Prophet.	64
Figura 28 - Previsão têxtil agregada Prophet.	65
Figura 29- Previsão têxtil segregada Prophet.	66

Figura 30 - Previsão Rossmann agregada Decision Tree.	67
Figura 31 - Previsão Rossmann segregada Decision Tree.	68
Figura 32 - Previsão têxtil agregada Decision Tree.	69
Figura 33 - Previsão têxtil segregada Decision Tree.	70
Figura 34 - Previsão Rossmann agregada Random Forest.	71
Figura 35 - Previsão Rossmann segregada Random Forest.	72
Figura 36 - Previsão têxtil agregada Random Forest.....	73
Figura 37 - Previsão têxtil segregada Random Forest.	74
Figura 38 - Previsão Rossmann agregada XGBoost.	75
Figura 39 - Previsão Rossmann segregada XGBoost.	76
Figura 40 - Previsão têxtil agregada XGBoost.	77
Figura 41 - Previsão têxtil segregada XGBoost.	78
Figura 42 - Previsão Rossmann agregada Rede Neural.	79
Figura 43 - Previsão Rossmann segregada Rede Neural.	80
Figura 44 - Previsão têxtil agregada Rede Neural.	81
Figura 45 - Previsão têxtil segregada Rede Neural.	82

LISTA DE QUADROS

Quadro 1 - Diferenças entre Random Forest e árvore de decisão simples.	36
Quadro 2 - RMSE de cada modelo.	83
Quadro 3 - Diferenças relativas percentuais entre RMSE.	84

LISTA DE TABELAS

Tabela 1 - Erro da previsão Rossmann segregada ARIMA.	57
Tabela 2 - Erro da previsão Rossmann segregada Prophet.	64
Tabela 3 - Erro da previsão Rossmann segregada Decision Tree.	67
Tabela 4 - Erro da previsão Rossmann segregada Random Forest.	71
Tabela 5 - Erro da previsão Rossmann segregada XGBoost.	75
Tabela 6 - Erro da previsão Rossmann segregada Rede Neural.	79

LISTA DE ABREVIATURAS E SIGLAS

PCP - Planejamento e Controle da Produção.

SKU - *Stock Keeping Unit* (Unidade de Manutenção de Estoque).

RMSE - *Root Mean Squared Error* (Raiz do Erro Médio Quadrado).

ARIMA - *Autoregressive Integrated Moving Average* (Auto Regressivo Integrado de Médias Móveis).

SARIMA - *Seasonal Autoregressive Integrated Moving Average* (Auto Regressivo Integrado de Médias Móveis Sazonais).

RNN - *Recurrent Neural Network* (Redes Neurais Recorrentes).

ANN - *Artificial Neural Network* (Redes Neurais Artificiais).

CNN - *Convolutional Neural Networks* (Redes Neurais Convolucionais).

LSTM - *Long Short Term Memory Networks* (Redes Neurais de Memória de Curto Prazo Longo).

Seq2Seq - *Sequence-to-Sequence* (Sequência para Sequência).

SUMÁRIO

1 INTRODUÇÃO.....	16
1.1 JUSTIFICATIVA.....	17
1.2 OBJETIVOS	17
1.2.1 Objetivo Geral.....	18
1.2.2 Objetivos Específicos.....	18
1.3 DELIMITAÇÕES	18
1.4 LIMITAÇÕES	19
1.5 ESTRUTURA DO TRABALHO.....	19
2. FUNDAMENTAÇÃO TEÓRICA.....	21
2.1 PLANEJAMENTO E CONTROLE DA PRODUÇÃO.....	21
2.2 PREVISÃO DE DEMANDA	23
2.2.1 Séries temporais	25
2.2.2 Métricas de erro e validação cruzada	26
2.2.3 Métodos tradicionais	29
2.2.3.1 Método ARIMA.....	29
2.2.3.2 Método Holt-Winters	30
2.2.3.3 Prophet.....	31
2.2.3.4 Problemas com abordagens tradicionais	32
2.2.4 Machine learning.....	33
2.2.4.1 Decision Tree	34
2.2.4.2 Random Forest	35
2.2.4.3 XGBoost.....	36
2.2.4.4 Redes Neurais	37
3. PROCEDIMENTOS METODOLÓGICOS	39
3.1 DADOS DA PESQUISA	39
3.2 ROTEIRO DO TRABALHO.....	40

4. DESENVOLVIMENTO.....	42
4.1 SELEÇÃO DOS MODELOS	42
4.2 DADOS	42
4.2.1 FARMACÊUTICA (ROSSMANN).....	43
4.2.2 TÊXTIL.....	50
4.3 DEFINIÇÃO BASE PARA COMPARAÇÃO.....	55
4.4 APLICAÇÃO DOS MÉTODOS.....	56
4.4.1 ARIMA	56
4.4.2 Holt-Winters.....	60
4.4.3 Prophet	62
4.4.4 Decision Tree	66
4.4.5 Random Forest	70
4.4.6 XGBoost.....	74
4.4.7 Rede Neural.....	78
5. RESULTADOS	83
6. CONCLUSÕES	87
REFERÊNCIAS.....	89

1 INTRODUÇÃO

A previsão de séries temporais é uma questão que se mostra fundamental para os mais diversos setores, pois podem facilitar tomadas de decisão estratégicas diversas. A previsão de demanda de uma empresa, por exemplo, pode facilitar a decisão de um gestor quanto à programação da produção, e ao mesmo tempo a previsão do comportamento das ações de uma empresa pode fundamentar a decisão de um investidor. Segundo Ali et al. (2009) um modelo de previsão de demanda adequado pode contribuir para o aumento do lucro e rentabilidade de uma empresa, entre outros fatores, por facilitar a gestão dos estoques e reduzir o nível de segurança dos estoques consideravelmente.

Ali et al. (2009) indicam que para períodos com vendas sem promoções é muito difícil superar o desempenho de modelos simples, como os modelos de regressão por médias móveis, e os métodos de *machine learning* em geral conseguem apenas resultados semelhantes. Para períodos com promoções, entretanto, os autores defendem que os métodos de *machine learning* apresentam resultados até cerca de 65% melhores, através do uso de algoritmos de árvore.

O presente trabalho busca analisar e comparar o desempenho, a capacidade de predição e a acurácia de métodos de previsão tradicionais, que utilizam princípios estatísticos básicos, e de métodos robustos, que utilizam artifícios de inteligência artificial e *machine learning*, em português aprendizado de máquina, na previsão de demanda para múltiplas séries temporais paralelas, que são séries temporais semelhantes que podem estar correlacionadas entre si, em um mesmo período de tempo, e verificar se, de fato, os métodos robustos apresentam resultados mais assertivos, bem como verificar se a previsão conjunta de séries paralelas apresenta vantagens sobre a predição individualizada destas séries.

O uso de inteligência artificial, em especial redes neurais artificiais, apresenta um crescimento exponencial de popularidade, devido à sua grande capacidade de previsão de séries temporais e identificação de padrões em conjuntos de dados. Apesar desta popularidade, os modelos de inteligência artificial apresentam uma série de complicações em sua utilização, que podem invalidar os resultados obtidos e inutilizar este modelo; em especial, estes modelos podem identificar padrões não existentes na prática em dados não relacionados, e retornar previsões que não condizem com a realidade (ZHANG, 2007).

Grange (1996) afirma que melhorar a qualidade de previsão é um grande desafio, para o qual Alon, Qi e Sadowski (2001) afirmam que métodos que utilizam *machine learning* se propõem a resolver. Segundo os autores, é fundamental que um modelo de previsão

identifique corretamente padrões como tendência e sazonalidade para realizar uma previsão assertiva, sendo as redes neurais fortes candidatas a superar o desempenho de métodos como a regressão exponencial e do modelo ARIMA de Box-Jenkins.

1.1 JUSTIFICATIVA

A importância do trabalho pode ser percebida ao observar a influência que uma previsão de demanda adequada apresenta para uma empresa, e os benefícios que traz para esta, como as informações fornecidas por esta previsão, que fornecem vantagens competitivas e facilitam o planejamento da produção e a gestão empresarial a nível estratégico.

A importância da pesquisa está ligada ao problema que será analisado, em que se faz necessária a predição de múltiplas séries temporais paralelas. Para um cenário como este, modelos tradicionais de previsão de demanda em geral tendem a apresentar baixo desempenho, e será analisado se os modelos que utilizam artifícios de *machine learning* são capazes de atingir melhores resultados, ao agrupar as múltiplas séries para realizar a previsão conjunta. Segundo Makridakis, Spiliotis e Assimakopoulos (2018) apesar da popularidade dos modelos de *machine learning* ainda não existe um número suficiente de comprovações objetivas em relação a performance destes métodos para afirmar que de fato eles são superiores aos métodos tradicionais, portanto o presente trabalho é proposto com o intuito de agregar para esta discussão.

Além disso, os modelos selecionados serão utilizados para a previsão de dois conjuntos de dados com características muito diferenciadas. Os dados utilizados são provenientes de setores completamente diferentes, desta forma apresentam padrões de sazonalidade e correlação entre os valores a serem previstos muito diferenciados, e os próprios conjuntos se diferem também na quantidade de valores existentes. A partir dos resultados obtidos será possível analisar como cada modelo é influenciado por estas diferenças e estimar quais modelos tendem a apresentar resultados melhores de acordo com as características dos dados que serão analisados.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Esta pesquisa tem como objetivo geral:

- Comparar o desempenho de métodos tradicionais estatísticos com métodos de inteligência artificial, especificamente em um contexto de previsão de múltiplos itens possivelmente correlacionados.

1.2.2 Objetivos Específicos

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

- a) Definir o experimento de comparação em relação à periodicidade, agregação e métricas de erro.
- b) Revisar literatura acerca de métodos estatísticos tradicionais e de métodos de inteligência artificial, selecionar e implementar métodos.
- c) Interpretar resultados obtidos evidenciando a situação de previsão de múltiplos itens possivelmente correlacionados.

1.3 DELIMITAÇÕES

Existe uma grande quantidade de métodos que podem ser utilizados para se realizar a previsão de séries temporais, e diversos estudos e discussões acerca da efetividade dos métodos. Partindo disto, o presente trabalho se propõe a analisar uma seleção de modelos pré-determinados a partir da análise de experiências e estudos já divulgados, que se espera que tenham resultados satisfatórios para grande parte dos casos de uso.

Desta forma, existem princípios e métodos que não serão abordados pois serão analisados apenas métodos de *machine learning* que utilizam uma lógica de Árvore de Decisão. Além disto, não será realizada nenhuma previsão que combine diferentes métodos e as redes neurais aplicadas utilizarão apenas uma estrutura de rede básica, pois a complexidade de elaborar modelos que incluíssem estas características foge do escopo de análise do trabalho.

Os métodos tradicionais estatísticos selecionados foram os métodos ARIMA, um modelo auto regressivo integrado de médias móveis; Holt-Winters, um modelo baseado em suavização exponencial; e o Prophet, um algoritmo baseado em um modelo aditivo. Os dois primeiros foram selecionados por serem os métodos mais tradicionais, amplamente utilizados para previsões de séries temporais por apresentarem resultados satisfatórios para uma grande gama de problemas. Já o terceiro é um método menos tradicional, desenvolvido a menos tempo pela equipe de ciência de dados do Facebook, mas que está sendo amplamente utilizado pela indústria, apresentando também bons resultados em diversas aplicações.

Para a seleção dos métodos de inteligência artificial foi seguido o mesmo princípio que os métodos estatísticos, em que foram selecionados modelos tradicionais já conceituados e um modelo disruptivo em ascensão. Para tanto, foram selecionados três algoritmos baseados no princípio de árvore de decisão (Decision Tree, Random Forest e XGBoost), que apresentam resultados satisfatórios para uma grande gama de problemas, de maneira semelhante aos métodos estatísticos escolhidos e são os mais utilizados no âmbito de aprendizado de máquina. O último princípio selecionado foi o de redes neurais, que vem sendo cada vez mais utilizadas, com a promessa de ser mais assertivo que os outros modelos em uma grande gama de problemas, principalmente por terem uma grande capacidade de assimilar séries temporais com funções não lineares, ao custo de uma maior complexidade de implementação.

1.4 LIMITAÇÕES

Uma limitação na execução do trabalho foi referente a aplicação de um dos métodos selecionados. O modelo Holt-Winter não pode ser utilizado em um dos conjuntos de dados utilizados, pois não havia dados suficientes para aplicação adequada do modelo. O modelo ARIMA, entretanto, foi aplicado conforme o esperado, o que permitiu que a análise proposta no trabalho pudesse ser realizada conforme o planejado inicialmente.

1.5 ESTRUTURA DO TRABALHO

O trabalho está dividido em 6 capítulos. O primeiro deles é o capítulo atual, referente a introdução, em que são apresentadas noções iniciais acerca do tema, as justificativas para sua escolha e os objetivos e limitações para a elaboração do trabalho.

O segundo capítulo é o de fundamentação teórica, no qual são explicados todos os métodos utilizados no trabalho e demais conceitos fundamentais para compreensão do trabalho, com base nas literaturas já publicadas acerca do tema.

O terceiro capítulo é constituído pelos procedimentos metodológicos, no qual são explorados os conjuntos de dados utilizados e é determinado o roteiro seguido no desenvolvimento do trabalho.

No quarto capítulo é descrito todo o desenvolvimento do trabalho, seguindo as etapas levantadas no capítulo anterior, que incluem a seleção dos modelos utilizados no trabalho, a definição de um cenário base para a comparação dos resultados e a aplicação de cada método e armazenamento dos resultados.

No quinto capítulo todos os resultados obtidos durante o desenvolvimento são retomados e agrupados para que seja feita a análise desejada, comparando os erros obtidos em cada cenário e concluindo a comparação de desempenho, objetivo geral do trabalho. Por fim, no sexto capítulo são demonstradas as conclusões e considerações finais do trabalho

2. FUNDAMENTAÇÃO TEÓRICA

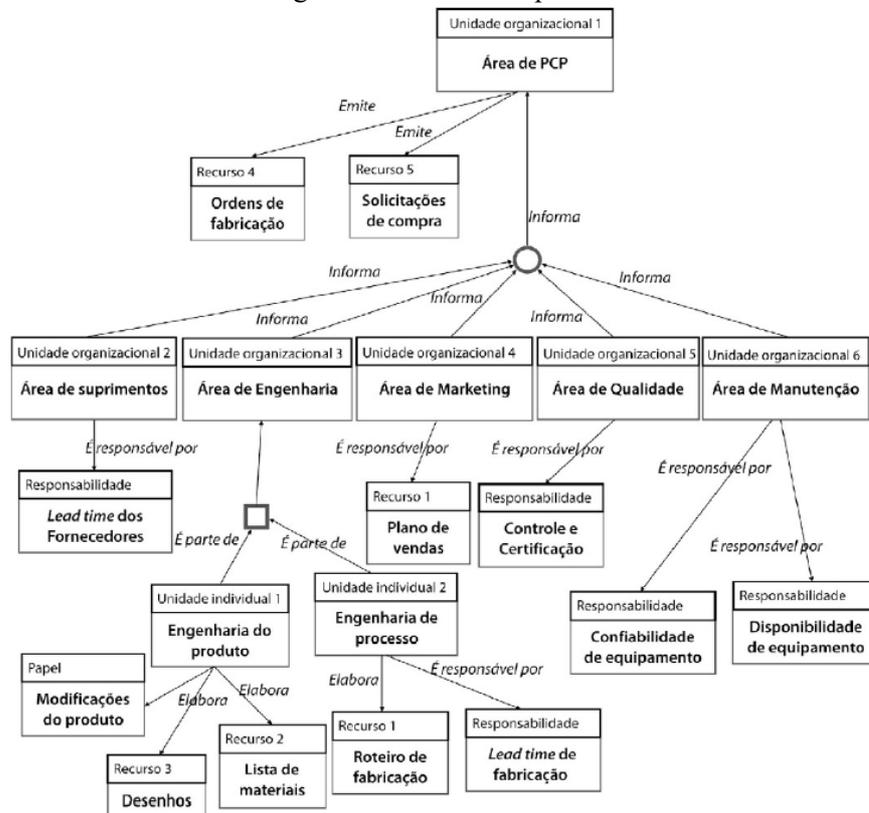
2.1 PLANEJAMENTO E CONTROLE DA PRODUÇÃO

O planejamento e controle da produção (PCP), segundo Guerrini, Belhot e Junior (2014), surgiu durante a transição do sistema artesanal para o sistema de produção em massa, em especial durante a revolução industrial, a partir da obra “A Riqueza das Nações” escrito por Adam Smith em 1776. O PCP atual é composto por uma combinação de diferentes técnicas de resolução de problemas utilizadas na gestão dos processos produtivos originados em torno desta época. Desta forma, o PCP necessita de informações provenientes de diversos setores de uma empresa e fornece uma visão sistêmica dos processos envolvidos.

O PCP se mostra como uma ferramenta fundamental para garantir a sobrevivência de uma empresa em mercados competitivos, pois para que um sistema produtivo desempenhe sua função conforme o esperado é necessário que ele seja planejado em torno de prazos, e é função do PCP realizar este planejamento (TUBINO, 2007).

A área de PCP em uma empresa integra diversos setores que trabalham conjuntamente para atingir seus objetivos, como demonstrado na figura 1, e é cada vez mais valorizada na gestão das empresas pelas vantagens competitivas que gera, como a redução de tempos de espera, redução de estoques e um aumento de responsividade a variações nas demandas (STEVENSON; HENDRY; KINGSMAN, 2004).

Figura 1 - PCP nas empresas.



Fonte: Guerrini, Belhot e Junior (2014)

Entre as atividades descritas pelo PCP citadas por Guerrini, Belhot e Junior (2014) estão a identificação dos sistemas de produção, o planejamento de recursos, a administração de estoque, a programação das atividades e a previsão das vendas, que serve de base para as demais atividades. Esta etapa é o primeiro processo realizado, que define a quantidade do que será produzido, por meio de métodos de predição que em geral se baseiam em dados históricos de vendas para prever as demandas futuras e, segundo Tubino (2007) é a etapa mais importante entre todas, por ser a responsável por definir todas as funções desenvolvidas pelo PCP.

A etapa de previsão de demanda é, portanto, a base do planejamento estratégico da produção de uma empresa, pois é ela que determina o rumo da empresa, permitindo aos administradores planejar suas decisões adequadamente com base nas expectativas para o futuro (TUBINO, 2007), sendo útil em diversos horizontes de planejamento. Em um curto prazo, as previsões de demandas são fundamentais na gestão das atividades da produção e na gestão dos estoques, enquanto no médio prazo as previsões são úteis na elaboração do plano de recursos e do plano de produção e no longo prazo podem servir de base para decisões

acerca de capacidade produtiva instalada, e desenvolvimento de novos produtos (GUERRINI; BELHOT; JUNIOR, 2014).

Em muitas empresas a previsão de demanda é uma responsabilidade dos setores de marketing ou vendas, mas mesmo nesses casos é fundamental que o setor de PCP compreenda este processo, pois ele é a base para todo o planejamento da produção e um entendimento direto do modo como estas previsões foram obtidas facilita a comunicação entre estes setores, e conseqüentemente a gestão dos meios produtivos (TUBINO, 2007).

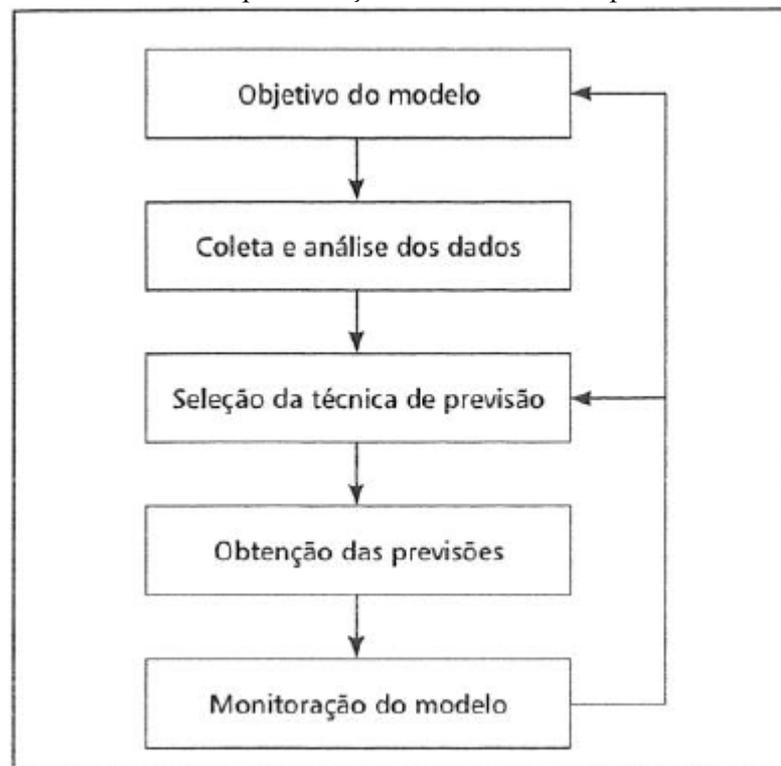
2.2 PREVISÃO DE DEMANDA

A previsão de demanda é parte fundamental para o setor de PCP de uma empresa, como visto anteriormente, mas ainda existem outros estudos não relacionados ao setor que evidenciam a importância desta previsão, como os estudos de Agrawal e Schorling (1996), que afirmam que uma previsão de demandas adequada é crucial para manter a viabilidade de uma empresa, pois sua falta pode acarretar em diversos problemas que afetem diretamente a eficiência e a competitividade de uma empresa, como a gestão dos estoques.

Tubino (2007) classifica a previsão de demandas como uma ciência não exata, que utiliza conjuntamente recursos computacionais e conhecimento tácito de quem realiza a previsão para obter um valor previsto, que é uma aproximação do valor real, o que implica que sempre haverá uma diferença entre estes valores, caracterizando o erro da previsão. O processo de previsão de demanda, portanto, envolve a seleção de modelos de previsão adequados e a otimização de seus parâmetros, com o objetivo de reduzir o erro da previsão, gerando previsões mais assertivas e conseqüentemente com maior capacidade de auxílio ao setor de PCP.

A implementação de um modelo de previsão de demanda em uma empresa se inicia com a definição do objetivo do modelo, que determina a demanda a ser prevista. Em seguida os dados históricos que serão a base do modelo devem ser coletados e analisados e a partir disso a técnica de previsão que será utilizada deve ser selecionada e a previsão pode ser realizada (TUBINO, 2007). Este processo pode ser realizado repetidamente para buscar a redução do erro da previsão a cada tentativa, como explicitado na figura 2.

Figura 2 - Processo de implementação de um modelo de previsão de demanda.



Fonte: Tubino (2007)

A previsão de demandas é realizada principalmente através da previsão de séries temporais, que são sequências de observações em intervalos regulares de tempo (FALK et al., 2006), de forma que modelos procuram calcular a demanda futura com base nos dados históricos de vendas, partindo do princípio que os valores futuros serão projeções dos valores passados, princípio conhecido como hipótese de continuidade (GUERRINI; BELHOT; JUNIOR, 2014).

As técnicas de previsão são classificadas por Tubino (2007) em qualitativas ou quantitativas. As qualitativas são de difícil representação numérica, pois envolvem muitos conhecimentos tácitos de quem realiza as previsões, e são constituídos por técnicas como pesquisas de mercado, analogias históricas ou análises dos ciclos de vendas (GUERRINI; BELHOT; JUNIOR, 2014), enquanto as técnicas quantitativas envolvem análises numéricas das informações históricas, sem utilizar informações subjetivas, podem ser classificados entre métodos de extrapolação e causais (ALI et al., 2009). Segundo Guerrini, Belhot e Junior (2014) estas técnicas só podem ser utilizadas se a informação do passado estiver disponível e puder ser quantificada, o que faz delas as mais importantes na previsão de séries temporais.

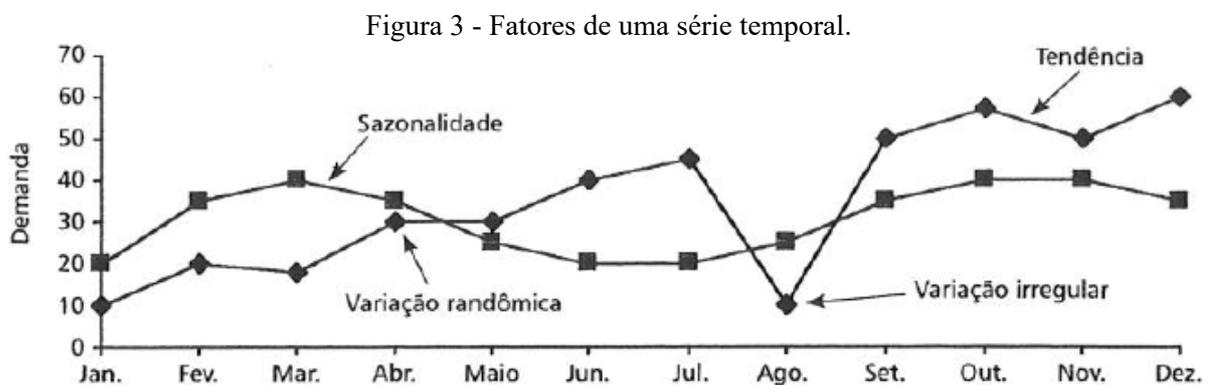
Métodos de extrapolação utilizam apenas os padrões presentes na série temporal prevista para gerar as previsões, por meio de técnicas como o cálculo de médias móveis, de

suavização exponencial ou a abordagem de Box-Jenkins. Por meio destes princípios, os métodos de extrapolação identificam e extrapolam padrões de tendência, sazonalidade e auto correlação nas séries temporais, mas não são capazes de identificar a influência de fatores externos na previsão, como mudanças de preço e promoções (ALON; QI; SADOWSKI, 2001).

Os métodos causais, por outro lado, são constituídos por modelos quantitativos que levam em consideração no cálculo da previsão fatores que se acredita que influenciam nos valores da série temporal, assumindo uma relação de causa-efeito entre os valores da série e estes fatores (GUERRINI; BELHOT; JUNIOR, 2014). Métodos causais podem ser simples como modelos de regressão lineares que levam em consideração variáveis como promoções ou complexos como os modelos de *machine learning* que envolvem árvores de decisão e redes neurais (ALI et al., 2009).

2.2.1 Séries temporais

A previsão de séries temporais tem como objetivo básico a identificação de um padrão na série histórica que possa ser extrapolado para o futuro, de forma a realizar a previsão por meio de operações matemáticas (GUERRINI; BELHOT; JUNIOR, 2014). Estes padrões, segundo Tubino (2007), são a tendência, a sazonalidade, variações irregulares e variações randômicas que são presentes nos dados, como demonstrado na figura 3.



Fonte: Tubino (2007)

Essa figura traz uma série temporal decomposta em duas outras séries temporais, uma delas sendo referente a sazonalidade dos dados, retratada pela linha com quadrados, e outra representando a tendência, sendo afetada pelas variações, retratada pela linha com losangos, de forma que a série original é a soma dos componentes. Observando a linha da tendência é possível perceber as variações randômicas, que fazem com a linha não assuma uma reta, e uma variação irregular, que afasta muito o comportamento em relação a linha original.

Guerrini, Belhot e Junior (2014) defendem que a tendência de uma série temporal é o comportamento mais importante em para a previsão, consistindo em um movimento gradual de longo prazo que direciona os dados (TUBINO, 2007). A sazonalidade observa mudanças cíclicas previsíveis em uma série temporal, causadas principalmente por mudanças nos padrões de consumo durante períodos específicos, por exemplo, para o caso de cobertores, em que a demanda é muito maior durante o inverno. Estes ciclos podem ser observados com diferentes perspectivas temporais, como sazonalidades anuais, mensais ou até diárias (GUERRINI; BELHOT; JUNIOR, 2014).

Quanto às variações, Tubino (2007) defende que as aleatórias devem ser tratadas pela sua média, enquanto as irregulares devem ser desconsideradas na previsão, e os valores que afetam devem ser substituídos pela média da série histórica, pois são resultantes de fatores excepcionais como catástrofes, que não podem ser previstos e portanto não contribuem na previsão da série como um todo.

Para a compreensão do comportamento de séries temporais uma das ferramentas úteis é a análise de correlações. Por meio de coeficientes de correlação, como o coeficiente de Pearson, é possível determinar o quanto duas séries temporais paralelas estão correlacionadas (PODOBNIK; STANLEY, 2008), ou, analisando apenas uma série temporal, é possível perceber como esta série está relacionada com ela mesmo no futuro (TULJAPURKAR; HARIDAS, 2006), o que facilita a visualização dos períodos de sazonalidade da série temporal.

2.2.2 Métricas de erro e validação cruzada.

Uma das técnicas mais utilizadas para a avaliação do desempenho de modelos de previsão é a de validação cruzada, que consiste em reservar parte do conjunto de dados utilizado durante a elaboração do modelo para avaliar seu erro. Os dados são separados em dois grupos, um denominado treino e outro denominado teste, de maneira que o primeiro é o maior dos dois e é utilizado para a elaboração do modelo, estimação de parâmetros e para o

próprio fit do modelo, enquanto o segundo é utilizado para mensuração da previsão (ZHANG, 2007).

A técnica de validação cruzada consiste então em utilizar o grupo de treino para fornecer as informações necessárias para o modelo, e após isso realizar a previsão referente ao grupo de teste, como se os valores reais deste grupo fossem desconhecidos e comparar estes valores obtidos na previsão com os valores reais respectivos por alguma métrica de cálculo de erro, como o RMSE. Por meio desta técnica é possível determinar o erro de diversos métodos de previsão, bem como otimizar os parâmetros do modelo de maneira a retornar um menor erro (ZHANG, 2007).

O valor de erro de uma previsão deve ser calculado através da diferença entre o valor real e o valor previsto da série, por meio de alguma métrica, como o erro acumulado ou o erro quadrático médio (GUERRINI; BELHOT; JUNIOR, 2014). Uma das métricas de erro mais utilizadas em diversas aplicações, segundo Wang e Lu (2018) é a métrica RMSE (*Root Mean Squared Error*), que apresenta entre suas principais vantagens o fato de que os valores são relativos aos dados utilizados para o cálculo, o que faz com que o método seja de fácil aplicação para diversos casos (CHAI; DRAXLER, 2014).

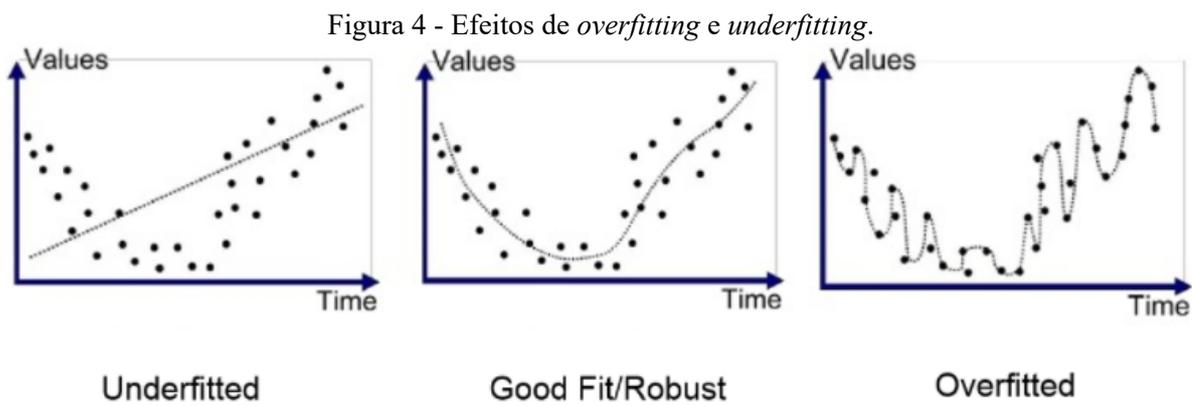
A fórmula utilizada para o cálculo do RMSE pode ser vista abaixo na equação 1, onde \hat{y}_i representa o valor previsto para o período i , y_i representa o valor real para este período e n representa o número de períodos, que é de 52 para todas as previsões, pois corresponde a um ano de valores agregados semanalmente.

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

Uma das técnicas mais utilizadas para reduzir o erro de uma previsão é a de otimização de parâmetros, que busca determinar os melhores parâmetros para o cenário que está sendo analisado (SILVEIRA, 2019). Bergstra et al. (2011) demonstram diferentes metodologias que podem ser utilizadas para a otimização dos parâmetros de um modelo, incluindo a perspectiva utilizada para o desenvolvimento deste trabalho, que envolve determinar um espaço amostral para os parâmetros do modelo, com base em estudos divulgados, e otimizar cada parâmetro dentro deste espaço, uma perspectiva simples, porém

efetiva. A otimização necessita ainda de uma atenção especial, pois uma otimização exacerbada pode acarretar um problema de *overfitting* (YING, 2018).

Entre os problemas mais conhecidos de modelos de previsão do tipo causal, como as redes neurais (ALI et al., 2009), está a questão de *overfitting* e *underfitting*. Estas questões são problemas opostos relativos ao grau especificação de um modelo de previsão durante o período de treino, apresentando respectivamente uma especificação muito alta e muito baixa (ZHANG, 2007). A figura 4 exemplifica os efeitos de *overfitting* e *underfitting* na previsão de séries temporais, comparando com uma previsão ideal.



Fonte: Associação Brasileira de Ciências de Dados (ABRACD)

Segundo Zhang (2007) o problema de *underfitting* ocorre quando o modelo é elaborado e treinado com uma baixa especificação da importância dos parâmetros, o que resulta em previsões com grande margem de erro inclusive para o grupo de teste. Este problema pode ser facilmente percebido durante a implementação de um modelo de previsão, o que faz dele o problema menos crítico entre os dois.

Ainda segundo Zhang (2007), o problema de *overfitting*, por outro lado, é o mais crítico entre os dois, e é tópico de diversas discussões e trabalhos científicos, o que ocorre pois este problema pode não ser percebido durante a elaboração do modelo mas inviabiliza sua utilização. O *overfitting* é caracterizado quando um modelo de previsão identifica padrões particulares para o conjunto de teste utilizado, o que faz com que o modelo se ajuste muito bem ao conjunto de dados observado, mas se mostra ineficaz para prever novos resultados, o que resulta em um erro muito baixo no ambiente de teste, por vezes inclusive durante as etapas de validação cruzada, o que indicaria um modelo bem estruturado, mas que não se repete em outros casos, pois o modelo foi especificado demais para o conjunto de teste, daí o nome *overfitting*. Este problema é mais comum em redes neurais devido a alta flexibilidade

destes modelos (ABURTO; WEBER, 2007), mas deve ser observado e evitado durante a aplicação de qualquer modelo de previsão.

Ayotte et al. (2021) tratam de outro fator que pode levar ao *overfitting* de um modelo que exige atenção durante a etapa de validação cruzada, que é a questão de *data leakage*, ou vazamento de dados, um dos principais erros na área de análise de dados de acordo com os autores, que é definida por Kaufman et al. (2012) como a inserção no treino de teste de valores ou informações pertencentes ao grupo de teste, que não deveriam ser disponíveis.

Em se tratando de séries temporais, a melhor opção para fugir deste problema é abordada por Ali (2008), que afirma que para a previsão de séries temporais é fundamental considerar a perspectiva temporal para a validação cruzada. Neste tipo de aplicação deve-se considerar uma data limite para o treino e teste, de maneira que o modelo de previsão adotado utilize apenas valores anteriores a esta divisão para seu treino e os valores posteriores para avaliação, simulando a aplicação real.

2.2.3 Métodos tradicionais

O primeiro grupo de modelos de previsão abordado neste trabalho são os métodos tradicionais estatísticos, que em sua maioria são caracterizados pelos modelos de extrapolação, de acordo com a classificação de ALI et al. (2009). Os modelos abordados são o ARIMA e o Holt-Winters, que Guerrini, Belhot e Junior (2014) afirmam estar os melhores modelos para previsão de séries temporais não estacionárias, em conjunto com o modelo Prophet, mais recente que os anteriores e que se enquadra como o exemplo mais simples de modelo causal de acordo com a classificação de ALI et al. (2009), sendo um modelo de regressão que leva em consideração fatores externos.

2.2.3.1 Método ARIMA

Box e Jenkins (1970) elaboraram uma abordagem complexa e amplamente utilizada até os dias de hoje para a previsão de séries temporais, que combina componentes autoregressivos e de médias móveis em um mesmo modelo. O primeiro componente calcula o valor de uma série temporal no futuro com base em seus valores passados (YOUSEFDOOST

et al., 2017) enquanto o segundo indica que o erro de um valor da regressão é uma combinação dos erros que ocorreram anteriormente (BAKAR; ROSBI, 2017).

Segundo Zhang (2003) esta abordagem serviu como base para a elaboração de um método de previsão mais recente, que ficou conhecido como ARIMA por agregar aos componentes inicialmente utilizados um componente de integração, que substitui valores de uma série pela sua diferença em relação a valores anteriores para séries não estacionárias (YANG et al., 2020)

Ainda segundo Zhang (2003), o modelo ARIMA apresenta grande flexibilidade e é possível utilizar o modelo para prever séries temporais com características distintas de maneira satisfatória, porém a maior limitação deste modelo está no fato que ele assume apenas uma relação linear entre as variáveis da previsão, ou seja, assume que o valor futuro de uma série é uma função linear de diversas observações e erros aleatórios, o que impede sua aplicação em casos que estas relações não são lineares (ABURTO; WEBER, 2007)

Existe ainda uma variação para o modelo ARIMA que leva em consideração componentes sazonais nas séries temporais, chamado de ARIMA sazonal ou SARIMA, que considera novamente os componentes de auto regressão, integração e médias móveis, mas com coeficientes distintos para os componentes sazonais (PERMANASARI et al., 2013).

2.2.3.2 Método Holt-Winters

O modelo de previsão Holt-Winter é um modelo simples que tem a capacidade de prever séries com tendência e variações sazonais, o que faz deste modelo ideal para o planejamento de produção e controle de estoques. Este modelo é útil por apresentar resultados próximos a outros modelos estatísticos, como a abordagem ARIMA de Box-Jenkins, sendo muito mais simples (CHATFIELD, 1978).

O modelo Holt-Winters consiste em múltiplas suavizações exponenciais, entre 1 e 3, dependendo da aplicação. O modelo com uma suavização é utilizado para séries estacionárias, o com dupla suavização é utilizado para séries com tendência e o modelo com tripla suavização é utilizado para séries com presença de tendência e sazonalidade (KALEKAR, 2004).

O caso mais geral de aplicação é, portanto, a suavização exponencial tripla, para o qual existem duas variações, dependendo do comportamento da sazonalidade que pode ser aditiva ou multiplicativa. As diferenças entre estes modelos estão presentes tanto na forma como as suavizações são combinadas quanto em seus próprios cálculos (KALEKAR, 2004).

Os dados aplicados no presente trabalho apresentaram um comportamento de sazonalidade do tipo aditivo, e as equações que descrevem as suavizações do modelo Holt-Winters para sazonalidade aditiva podem ser vistos a seguir nas equações 2, 3 e 4, representando, respectivamente, os componentes estacionário, de tendência e sazonal do modelo, seguindo os estudos de (KALEKAR, 2004).

$$\bar{R}_t = \alpha(y_t - \bar{S}_{t-L}) + (1 - \alpha)(\bar{R}_{t-1} + \bar{G}_{t-1}) \quad (2)$$

$$\bar{G}_t = \beta(\bar{S}_t - \bar{S}_{t-1}) + (1 - \beta)\bar{G}_{t-1} \quad (3)$$

$$\bar{S}_t = \gamma(y_t - \bar{S}_t) + (1 - \gamma)\bar{S}_{t-L} \quad (4)$$

A partir destas equações tem-se que a previsão é dada pela soma de todos os componentes, como visto na equação 5.

$$y_t = \bar{R}_t + \bar{G}_t + \bar{S}_t \quad (5)$$

Nestas equações temos que y_t corresponde ao valor previsto para um período t e \bar{R}_t , \bar{G}_t e \bar{S}_t correspondem aos valores de previsão dos componentes estacionário, de tendência e sazonal da série, respectivamente, para um período t . Quanto aos componentes utilizados nas equações, α , β e γ correspondem aos três coeficientes de suavização, que podem assumir valores entre 0 e 1.

2.2.3.3 Prophet

Prophet é um modelo de previsão de séries temporais automático, que permite incorporar variáveis externas à série no modelo desenvolvido por Taylor e Letham (2018) para as linguagens de programação R e Python, com o objetivo de ser um modelo simples com parâmetros intuitivos, e tem sua eficiência semelhante a modelos como ARIMA e o modelo de Machine Learning Random Forest, como comprovado por Papacharalampous e

Tyralis (2018) e Samal et al. (2019), respectivamente. Satrio et al. (2020) ainda mostram em seus estudos que o Prophet foi superior ao modelo ARIMA em um curto prazo, mas que perdeu capacidade de predição com o aumento do horizonte de previsão.

Prophet modela o comportamento de uma série temporal como a combinação de três componentes, semelhante ao modelo Holt-Winters, que são somados a um valor médio de erro, correspondente às variações que não se encaixam nos demais componentes. O primeiro destes componentes é a tendência da série, que pode assumir um comportamento de crescimento saturado ou de um modelo linear por partes, e é acrescido pelos valores correspondentes à sazonalidade da série e aos feriados (TAYLOR; LETHAM 2018).

O componente de feriados é o componente especial do Prophet, que permite ao modelo identificar efeitos que não apresentem um intervalo regular, mas que sejam conhecidos, como acontecimentos importantes sem datas fixas, por exemplo finais de campeonatos esportivos, e períodos de promoção, entre outros (TAYLOR; LETHAM 2018).

O modelo Prophet é um modelo de fácil utilização, pois seus parâmetros são simples e automaticamente otimizados, como demonstrado nos estudos de (SATRIO et al., 2020), o que faz com que o modelo seja flexível e útil em diversas aplicações. Apesar de ser mais eficiente para séries temporais com históricos longos e comportamento majoritariamente sazonal, Prophet apresenta resultados satisfatórios para séries diversas (SATRIO et al., 2020).

2.2.3.4 Problemas com abordagens tradicionais

Abordagens tradicionais de previsão de séries temporais, como os métodos ARIMA e Holt-Winters, buscam prever os valores futuros de uma série baseados apenas no histórico de valores conhecidos desta série, o que na prática pode não ser efetivo. Existem diversos fatores que podem interferir com os valores reais, como outras séries que podem estar relacionadas, ou diferentes sazonalidades que os dados podem apresentar conjuntamente. A partir deste cenário, os métodos de *machine learning*, como os por árvore de decisão e as redes neurais propõem solucionar estes problemas (WEN et al., 2018).

Com o desenvolvimento de novos produtos especialmente acelerado nos últimos anos, o ciclo de vida dos produtos tende a ser menor (BAYUS e PUTSIS, 1999), o que prejudica os modelos tradicionais que necessitam de séries temporais relativamente longas para apresentar resultados satisfatórios (ALI et al., 2009). Além disso, as vendas são fortemente influenciadas pela presença de promoções, que aumentam as demandas durante os períodos em que ocorrem, e os modelos tradicionais não são capazes de prever esta influência, o que pode

resultar em previsões incorretas e conseqüentemente a falta de produtos para atender as demandas (ALI et al., 2009).

Estes problemas comuns enfrentados pelos métodos tradicionais podem ser solucionados por meio do uso de métodos de *machine learning*, pois estes métodos, segundo Ali et al. (2009), são capazes de assimilar variáveis além dos valores da série temporal, o que possibilita utilizar informações de produtos diferentes como apoio na previsão, diminuindo o impacto da previsão para produtos com ciclo de vida curto, possibilita fornecer informações adicionais como períodos de promoção e também possibilita a compreensão de séries complexas compostas por relações não lineares. Todos estes fatores fazem com que os métodos de *machine learning* sejam especialmente úteis em cenários dinâmicos.

2.2.4 Machine learning

A inteligência artificial é cada vez mais utilizada para diversas aplicações avançadas e complexas, como na direção de veículos autônomos, reconhecimento de imagens e traduções automáticas, entre diversas outras, e são capazes de vencer jogos como xadrez e pôquer contra jogadores de altíssimo nível (MAKRIDAKIS, 2017). Segundo (MAKRIDAKIS; SPILLOTIS; ASSIMAKOPOULOS, 2018) entre as principais aplicações da inteligência artificial estão os sistemas de previsão, que tem seu sucesso baseado principalmente em algoritmos capazes de melhoria contínua por processos de tentativa e erro e não apenas com base nos aspectos tradicionais baseados em lógica de programação.

Os métodos de *machine learning*, especialmente as redes neurais, são extremamente úteis na previsão de séries temporais, e existe um grande número de novos métodos propostos indicando aumento na eficácia das previsões, porém existem poucos estudos que comprovam esta superioridade, sendo que os existentes, em geral, se limitam a analisar esta superioridade para poucas séries temporais, com horizonte de previsão em geral baixos e com poucos métodos alternativos para comparação (MAKRIDAKIS; SPILLOTIS; ASSIMAKOPOULOS, 2018).

Uma questão importante a ser analisada para a avaliação de modelos de previsão, que por vezes não recebe a devida importância, é a da complexidade dos modelos e os custos desta complexidade. Utilizar dados com grande detalhamento acarreta em elevados custos,

pois é preciso coletar, analisar e armazenar não somente a série temporal que será prevista, mas também todas as variáveis que serão englobadas na previsão, e é benéfico apenas se métodos robustos forem aplicados, pois modelos de regressão simples não são capazes de utilizar variáveis externas para a previsão, o que faz com que modelo de *machine learning* apresentem um custo elevado e é o principal motivo que limita a aplicação destes modelos, mesmo que eles apresentem resultados superiores para a maioria dos casos (ALI et al., 2009).

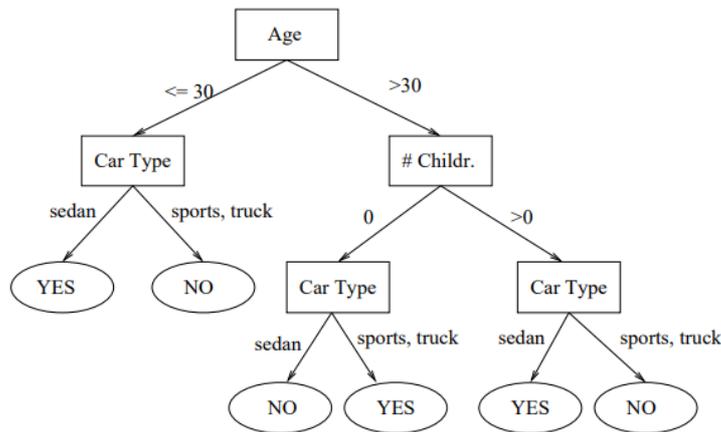
Os modelos de *machine learning* utilizados no trabalho e abordados a seguir são os modelos Decision Tree, Random Forest, XGBoost e uma rede neural, classificados como causais de acordo com a classificação de ALI et al. (2009).

2.2.4.1 Decision Tree

Árvores de decisão são métodos de *machine learning* baseados em aprendizado supervisionado, utilizado para classificação e regressão, que tem como objetivo criar um modelo que realiza a previsão a partir das variáveis conhecidas, seguindo regras simples (DAWOOD, 2019).

A estrutura básica de uma árvore de decisão envolve múltiplos nós e ligações. O nó inicial de uma árvore de decisão engloba todos os dados que serão fornecidos ao modelo e a partir deste nó ramificações são criadas a partir dos valores das variáveis por meio de regras de classificação. A árvore é então formada com múltiplos possíveis nós de saída, denominados de folhas, cada um recebendo um possível valor de resposta, e cada valor que será previsto é direcionado a um nó específico, a partir dos valores de suas variáveis (MAURYA et al., 2020). Um exemplo de árvore de decisão de classificação é visto na figura 5, que busca classificar a variável “Sub”.

Figura 5 - Exemplo de árvore de decisão.



Fonte: Dobra (2002)

Car type	Age	Chd	Sub
sedan	23	0	yes
sports	31	1	no
sedan	36	1	no
truck	25	2	no
sports	30	0	no
sedan	36	0	no
sedan	25	0	yes
truck	36	1	no
sedan	30	2	yes
sedan	31	1	yes
sports	25	0	no
sedan	45	1	yes
sports	23	2	no
truck	45	0	yes

Árvores de decisão utilizadas para regressão são generalizações das árvores utilizadas para classificação, em que os valores previstos são contínuos e não categóricos, o que faz com que os valores dos nós sejam funções ao invés de valores absolutos. O resultado de uma regressão é alcançado da mesma forma que uma classificação, quando se atinge uma folha da árvore (DOBRA, 2002).

Dobra (2002) defende que as árvores de decisão estão entre os modelos mais utilizados devido às suas 3 principais vantagens, que garantem resultados comparáveis aos demais métodos. A primeira delas é a representação intuitiva da árvore, que facilita seu entendimento. A segunda é que as árvores não são paramétricas, o que faz delas aplicáveis para diversas situações e a terceira é que árvores são algoritmos facilmente escaláveis.

As árvores de decisão estão entre os métodos mais utilizados para problemas de regressão, pois são modelos intuitivos que garantem resultados comparáveis aos demais métodos (DOBRA, 2002). Algoritmos baseados em árvores de decisão simples exigem pouco esforço de preparação de dados, o que os torna versáteis para diversas aplicações, e também são simples em termos computacionais, exigindo baixo poder de processamento para sua aplicação, mas não são efetivos para a previsão de conjuntos com grandes quantidades de dados (DAWOOD, 2019).

2.2.4.2 Random Forest

Um grande problema dos algoritmos de árvore de decisão é a questão de *overfitting*, um problema comum que ocorre com maior frequência em árvores muito grandes (DAWOOD, 2019) e é a razão que faz métodos derivados de árvores de decisão mais robustos serem populares, pois evitam estes problemas.

Com o objetivo de reduzir a ocorrência de *overfitting*, Breiman (2001) cita uma abordagem baseada no princípio de árvore de decisão que consiste em uma combinação de diversas árvores individuais e aleatórias, denominada Random Forest. Este modelo apresenta resultados muito superiores em regressões pois o resultado final se torna a média das previsões de cada árvore, o que faz com que o erro diminua com o aumento das árvores.

Algoritmos de tipo Random Forest podem ser utilizados tanto para classificação quanto para regressão e apresentam como principais vantagens do modelo a redução do erro da previsão, redução da possibilidade de *overfitting* e a possibilidade de previsão de conjuntos de dados extensos, a custo de um modelo mais complexo, com maior dificuldade de implementação e interpretação e exigindo mais recursos computacionais (MAURYA et al., 2020). O quadro 1 resume as principais diferenças entre o modelo Random Forest e um modelo simples de árvore de decisão.

Quadro 1 - Diferenças entre Random Forest e árvore de decisão simples.

	Random Forest	Decision Tree
Quantidade de árvores	Múltiplas	Uma
Conjunto de dados	Subconjuntos criados para diferentes árvores	Conjunto utilizado sem subdivisões
Eficácia	Alta	Relativamente menor
Possibilidade de <i>overfitting</i>	Baixa	Alta
Implementação	Relativamente mais complexa	Fácil
Interpretação	Relativamente mais complexa	Fácil
Custo computacional	Relativamente maior	Baixo
Quantidade de dados	Bons desempenhos para conjuntos extensos	Baixo desempenho para conjuntos extensos

Fonte: Adaptado de Maurya et al. (2020).

2.2.4.3 XGBoost

XGBoost é a forma abreviada de *eXtreme Gradient Boosting*, um modelo de previsão extremamente eficiente que vem apresentando desempenho superior a maioria dos modelos

de previsão em diversos desafios da área de Machine Learning, estando entre os modelos mais utilizados por vencedores de competições no site Kaggle (CHEN; GUESTRIN, 2016).

Como o nome completo sugere, o XGBoost é baseado no princípio de *Gradient Boosting*, uma técnica de *machine learning* utilizada para regressão e classificação, que consiste na combinação de diversos modelos de previsão fracos para elaboração de um modelo principal robusto (Friedman, 2002). No caso do XGBoost os modelos utilizados para a combinação são árvores de decisão pequenas (CHEN; GUESTRIN, 2016), o que reduz consideravelmente a chance do modelo acarretar um *overfitting* (DAWOOD, 2019).

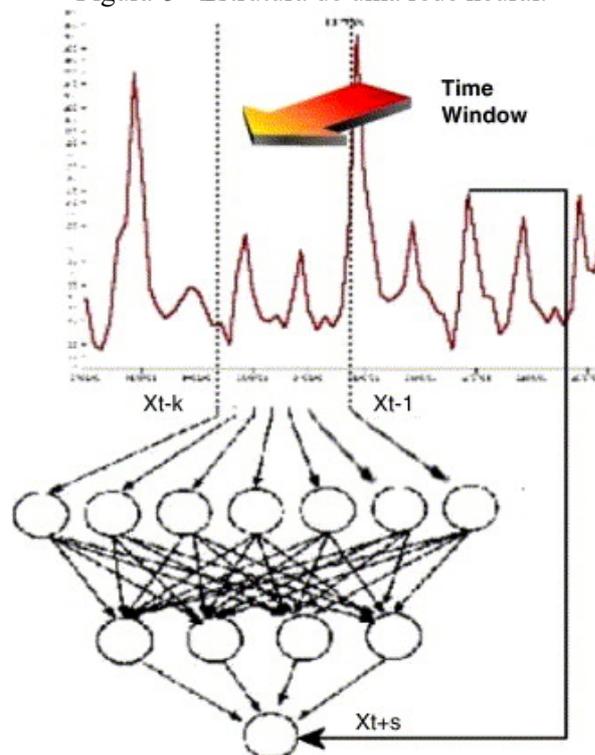
Chen e Guestrin (2016) afirmam que o fator mais importante para a popularidade do modelo XGBoost é a sua escalabilidade e aplicabilidade em diversos cenários, devido a diversas otimizações de algoritmos realizadas durante seu desenvolvimento, o que faz deste modelo um dos mais utilizados em aplicações reais.

2.2.4.4 Redes Neurais

Segundo (ALON, QI, & SADOWSKI, 2001) Redes Neurais Artificiais (ANN) são uma classe de modelos de regressão não linear inspirados nos estudos do sistema nervoso humano, capazes de modelar regressões complexas e não lineares sem necessidade de conhecimento prévio e abstrato a respeito dos dados analisados, o que faz das ANNs extremamente úteis na previsão de séries temporais, por exemplo na previsão do preço de uma ação no mercado financeiro ou na previsão de demandas de uma empresa (ABURTO; WEBER, 2007).

Redes neurais são extremamente complexas, e de difícil interpretação, exigindo uma baixa interação com quem a está utilizando, sendo por vezes tratada como uma “caixa preta”. Uma rede neural é uma rede de unidades, denominadas neurônios, ligadas por conectores, que são treinados de acordo com alguma regra para ajustar os pesos de cada conector de acordo com o comportamento dos dados analisados, de forma a alterar o valor dos neurônios com o objetivo de minimizar o erro da previsão (ABURTO; WEBER, 2007). A figura 6 exemplifica a estrutura de uma rede neural simples, constituída por uma camada com 4 neurônios.

Figura 6 - Estrutura de uma rede neural.



Fonte: Aburto e Weber (2007).

Nas últimas três décadas as redes neurais tiveram um acelerado crescimento em diversas áreas de aplicação, devido a sua promessa de identificar e modelar com eficiência padrões complexos, que não são identificados pelos modelos tradicionais de regressão. O principal contraponto para a implementação de redes neurais é a sua complexidade e

dificuldade de implementação, devido a sua grande flexibilidade, que inclui a necessidade de especificar a arquitetura da rede, incluindo o número de camadas, neurônios e conectores (ALON, QI, & SADOWSKI, 2001).

Historicamente, Redes Neurais Recorrentes (RNN) são os melhores modelos de previsão para séries temporais com características não lineares, principalmente devido a abordagem de *Sequence-to-Sequence* (Seq2Seq) de Graves (2013), que permitiu as RNNs gerarem uma série futura como previsão a partir de uma série temporal conhecida, a partir de previsões com múltiplos níveis. Dentre as diversas estruturas de redes neurais, as Redes Neurais Convolucionais (CNN) são amplamente utilizadas, apresentando resultados marcantes na área de geração de áudio (WEN et al., 2018).

3. PROCEDIMENTOS METODOLÓGICOS

O presente trabalho é uma pesquisa de natureza aplicada, de abordagem quantitativa, com propósito de caráter explicativo e de método experimental, que se enquadra na área de pesquisa operacional da engenharia de produção, na subárea de análise de demanda (ABEPRO, [s.d.]).

3.1 DADOS DA PESQUISA

O presente trabalho foi realizado com base em dois conjuntos de dados distintos, referentes à empresas que atuam em diferentes setores da economia e diferentes países, para que fosse possível analisar a influência do comportamento dos dados nos modelos de previsão aplicados. Desta forma os modelos selecionados foram aplicados para ambos os conjuntos independentemente.

O primeiro conjunto utilizado no trabalho é um conjunto de dados público disponibilizado através da plataforma Kaggle, uma comunidade on-line da área de ciência de dados, amplamente utilizada para competições de Data Science, em que empresas disponibilizam bases de dados próprias para que sejam realizadas análises preditivas. O conjunto utilizado é referente à empresa Rossmann, uma farmacêutica alemã, com mais de 3700 lojas por toda europa, e são referentes aos valores de vendas históricas para 1115 lojas, entre o período de janeiro de 2013 a setembro de 2015.

O segundo conjunto utilizado corresponde a dados privados de uma empresa brasileira do setor têxtil, referentes às vendas entre janeiro de 2015 e outubro de 2020. As previsões e análises realizadas no trabalho serão restritas apenas a uma das 60 coleções presentes neste conjunto, escolhida por ser a responsável pela maior parte das vendas da empresa no período, e por ser a coleção com menos lacunas nos dados utilizados. Para a elaboração do trabalho, entretanto, os dados do ano de 2020 foram desconsiderados, pois a pandemia trouxe grandes alterações extraordinárias nas vendas, e caso este período fosse utilizado para a análise os resultados poderiam ser prejudicados, por se tratar de uma situação extremamente peculiar.

Devido às diferenças entre os dados das empresas é esperado que seja possível analisar a influência das características das séries previstas para o desempenho de cada modelo estudado. Entre as características que podem influenciar os resultados está a quantidade de dados, que é maior no segundo conjunto. Enquanto o primeiro apresenta dados históricos referentes a 33 meses, o segundo apresenta dados referentes a 70 meses. Além

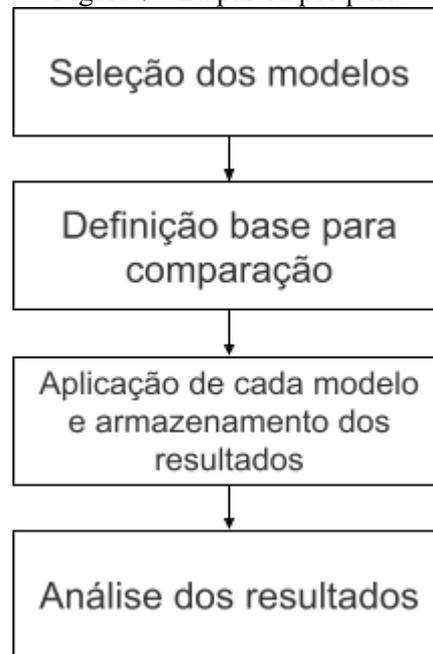
disso, existe uma sazonalidade muito forte nas vendas da indústria têxtil, o que não acontece no setor farmacêutico e pode fazer com que o desempenho de métodos que identifiquem esta sazonalidade seja maior.

Para ambos os conjuntos a previsão foi realizada em dois níveis de agregação distintos, com o intuito de verificar a capacidade de cada modelo de realizar a previsão de múltiplas séries paralelas. Para a empresa farmacêutica foi realizada uma previsão geral da rede Rossmann como um todo, agrupando os valores das 1115 lojas, e outra previsão com as principais lojas da rede, enquanto para o conjunto da empresa têxtil foi realizada uma previsão geral para a coleção selecionada e uma segunda previsão com menor nível de agregação, em que foram previstas as linhas de produtos que constituem a coleção, com distinção da região do Brasil que realizou a compra.

3.2 ROTEIRO DO TRABALHO

As etapas para realização do trabalho foram o levantamento de modelos de previsão tradicionais e dos robustos, a definição de uma base para comparação entre os resultados, a aplicação e avaliação dos métodos de previsão de demanda tradicionais univariados, aplicação e avaliação dos métodos de previsão de demanda robustos e por fim a comparação dos métodos aplicados e dos resultados obtidos. A figura 7 retrata estas etapas de maneira resumida.

Figura 7 - Etapas da pesquisa.



Fonte: Autor (2021)

A seleção dos modelos foi realizada arbitrariamente no início da pesquisa. Foram selecionados modelos de maneira que a pesquisa aborda modelos tradicionais e modelos modernos. Foram utilizados os modelos ARIMA, Holt-Winters e Prophet como representantes dos modelos tradicionais de regressão e os algoritmos Decision Tree, Random Forest, XG Boost e redes neurais como representantes dos modelos de *machine learning*.

Para que os resultados obtidos em cada previsão possam ser comparados adequadamente é necessário estabelecer um cenário base para os resultados, que envolve estabelecer uma metodologia padrão para a previsão de todos os métodos e uma métrica para cálculo de erro. A metodologia de previsão escolhida foi a de prever o último ano retratado nos dados, em uma perspectiva semanal, e comparar esta previsão com os valores reais pelo cálculo da raiz do erro médio quadrado, métrica chamada de RMSE.

A aplicação de todos os modelos, e todas as preparações de dados e otimização dos modelos necessárias foram realizadas através da linguagem de programação Python, uma linguagem de programação de alto nível orientada a objetos de código aberto criada em 1990 que é amplamente utilizada por grandes empresas como Google e Microsoft (BORGES, 2014).

4. DESENVOLVIMENTO

4.1 SELEÇÃO DOS MODELOS

A seleção dos modelos de previsão foi realizada de maneira arbitrária através do estudo de trabalhos e pesquisas divulgados. Os métodos tradicionais estatísticos selecionados foram os métodos ARIMA, Holt-Winters e Prophet. Os dois primeiros foram selecionados por serem os métodos mais tradicionais, que tendem a apresentar resultados satisfatórios para uma grande gama de problemas. Já o terceiro é um método menos tradicional, desenvolvido a menos tempo pela equipe de ciência de dados do Facebook, que está sendo amplamente utilizado pela indústria, e apresenta resultados satisfatórios em diversas aplicações.

A seleção dos métodos de inteligência artificial foi realizada seguindo o mesmo princípio de que os métodos estatísticos, em que foram selecionados modelos tradicionais já conceituados e um modelo disruptivo em ascensão. Para tanto, foram selecionados três algoritmos baseados no princípio de árvore de decisão (Decision Tree, Random Forest e XGBoost), e um modelo de rede neural, princípio que vem sendo cada vez mais utilizado, com a promessa de ser mais assertivo que os outros modelos em uma grande gama de problemas, principalmente por ter uma grande capacidade de assimilar séries temporais com funções não lineares, ao custo de uma maior complexidade de implementação.

4.2 DADOS

O presente trabalho foi realizado com base em dois conjuntos de dados distintos, sendo o primeiro um conjunto público referente a farmacêutica alemã Rossmann e o segundo pertencente a uma empresa têxtil localizada em Santa Catarina. Estes conjuntos apresentam grandes diferenças entre si, que vão desde a quantidade de dados e o horizonte de tempo até a sazonalidade dos dados, e que permitem análises quanto à capacidade dos métodos de previsão em determinadas situações. Os valores de venda da Rossmann, por exemplo, apresentam variações menores em relação à média, por ser proveniente de uma demanda que varia pouco ao longo do ano, comportamento bem diferente das vendas do setor têxtil, em que determinadas coleções são vendidas apenas durante algumas estações do ano.

4.2.1 FARMACÊUTICA (ROSSMANN)

Os dados da empresa farmacêutica são obtidos através da plataforma Kaggle, são referentes aos valores de vendas de 1115 lojas entre janeiro de 2013 e setembro de 2015 e são fornecidos através de duas tabelas, que podem ser vistas nas figuras 8 e 9.

Figura 8- Tabela das vendas Rossmann.

	DayOfWeek	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday	
Date	Store							
2015-07-31	1	5	5263	555	1	1	0	1
	2	5	6064	625	1	1	0	1
	3	5	8314	821	1	1	0	1
	4	5	13995	1498	1	1	0	1
	5	5	4822	559	1	1	0	1
...
2013-01-01	1111	2	0	0	0	0	a	1
	1112	2	0	0	0	0	a	1
	1113	2	0	0	0	0	a	1
	1114	2	0	0	0	0	a	1
	1115	2	0	0	0	0	a	1

1017209 rows × 7 columns

Fonte: Autor (2021)

Figura 9 - Tabela das lojas Rossmann.

Store	StoreType	Assortment	CompetitionDistance	CompetitionOpenSinceMonth	CompetitionOpenSinceYear	Promo2	Promo2SinceWeek	Promo2SinceYear	PromoInterval
1	c	a	1270.0	9.0	2008.0	0	NaN	NaN	NaN
2	a	a	570.0	11.0	2007.0	1	13.0	2010.0	Jan,Apr,Jul,Oct
3	a	a	14130.0	12.0	2006.0	1	14.0	2011.0	Jan,Apr,Jul,Oct
4	c	c	620.0	9.0	2009.0	0	NaN	NaN	NaN
5	a	a	29910.0	4.0	2015.0	0	NaN	NaN	NaN
...
1111	a	a	1900.0	6.0	2014.0	1	31.0	2013.0	Jan,Apr,Jul,Oct
1112	c	c	1880.0	4.0	2006.0	0	NaN	NaN	NaN
1113	a	c	9260.0	NaN	NaN	0	NaN	NaN	NaN
1114	a	c	870.0	NaN	NaN	0	NaN	NaN	NaN
1115	d	c	5350.0	NaN	NaN	1	22.0	2012.0	Mar,Jun,Sept,Dec

1115 rows x 9 columns

Fonte: Autor (2021)

A primeira tabela é referente às vendas da rede, em que cada linha é referente a venda de uma loja em um dia específico, ou seja, cada dia do período entre janeiro de 2013 e setembro de 2015 aparece 1115 vezes, uma para cada loja da rede no conjunto de dados. Cada coluna desta tabela fornece uma informação específica para cada loja e dia especificados, como explicitado a seguir:

- “DayOfWeek”: dia da semana, em que 1 é referente a segunda-feira e 7 é referente a domingo.
- “Sales”: valor total das vendas.
- “Customers”: número total de clientes.
- “Open”: informa se a loja está aberta, indicado pelo número 1, ou fechada, indicada pelo número 0.
- “Promo”: informa se está acontecendo alguma promoção, indicado pelo número 1, ou não, indicado pelo número 0.
- “StateHoliday”: indica o acontecimento de feriados nacionais, nos quais, em geral, todas as lojas da rede são fechadas. “0” indica que a data não é um feriado, “a” indica um feriado público, “b” indica o feriado de Páscoa e “c” indica o feriado do Natal.
- “SchoolHoliday”: indica se a loja foi afetada por escolas públicas estando fechadas, indicado pelo número 1, ou não, indicado pelo número 0.

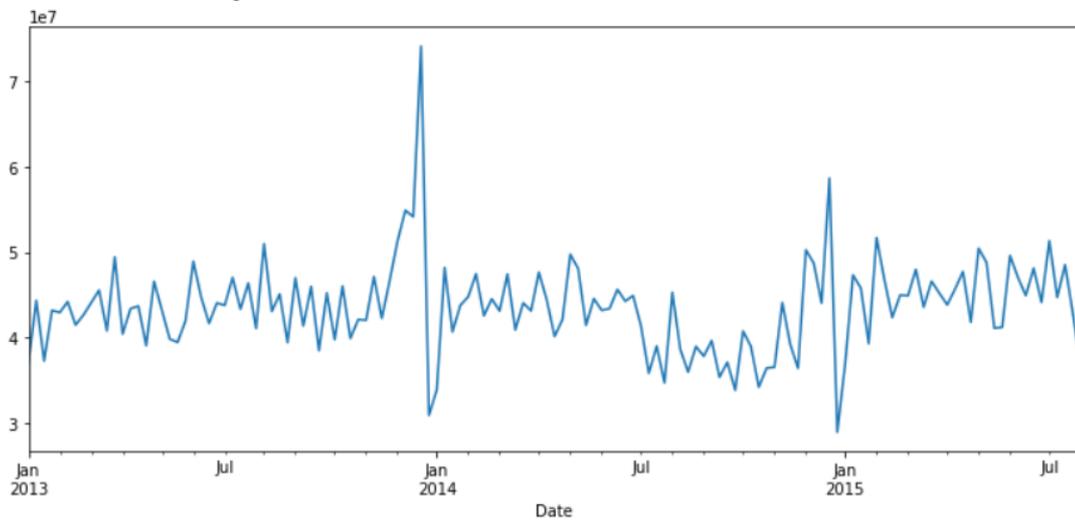
A segunda tabela traz informações referentes a cada loja, que são constantes durante todo o período, sendo elas:

- “StoreType”: classifica as lojas da rede em 4 grupos, indicados pelas letras “a”, “b”, “c” e “d”. Não é fornecida uma explicação sobre as características reais dos grupos.
- “Assortment”: classifica o nível de sortimento presente na loja, onde “a” é um sortimento básico, “b” é um sortimento extra e “c” é um sortimento estendido.
- “CompetitionDistance”: distância em metros até a loja concorrente mais próxima.
- “CompetitionOpenSinceMonth” : indica o mês em que a loja concorrente mais próxima foi inaugurada.
- “CompetitionOpenSinceYear”: indica o ano em que a loja concorrente mais próxima foi inaugurada.
- “Promo2”: informa se a loja participa de uma promoção contínua que ocorre para algumas lojas. O número 1 indica que a loja participa da promoção e o número 0 indica que não participa.
- “Promo2SinceWeek”: indica a semana em que a loja iniciou sua promoção contínua.
- “Promo2SinceYear”: indica o ano em que a loja iniciou sua promoção contínua.
- “PromoInterval”: indica os meses em que a promoção contínua é revisada.

Para o desenvolvimento do trabalho este conjunto de dados foi utilizado em duas perspectivas de granularidade diferentes, sendo uma delas a previsão de todas as lojas da rede agrupadas e a outra a previsão paralela das principais lojas da rede.

A primeira perspectiva é a série agrupada, com a maior granularidade dos dados, em que as vendas de cada loja foram somadas para possibilitar uma análise geral das vendas da rede Rossmann. Os valores foram agrupados semanalmente, de forma que cada valor da série corresponde ao valor total de vendas das 1115 lojas para uma semana específica. A figura 10 traz esta série temporal agrupada das lojas no período de análise

Figura 10 - Valores históricos de venda da rede Rossmann.

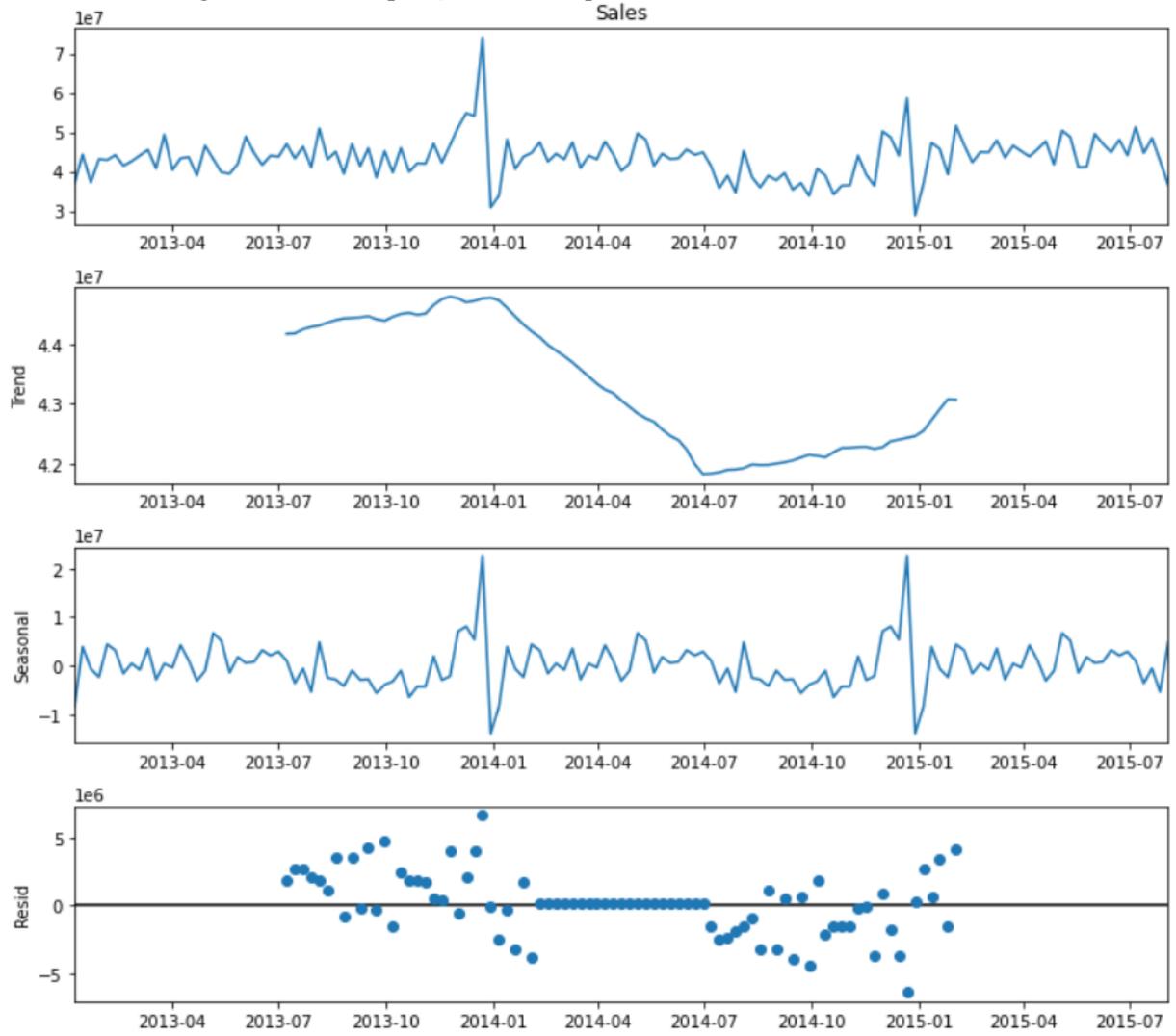


Fonte: Autor (2021)

À primeira vista, a análise deste gráfico permite perceber que os valores apresentam um comportamento serrilhado, em que as vendas se mantêm relativamente próximas à média com variações aparentemente aleatórias durante quase todo o período retratado. A principal exceção a essa característica pode ser observada nos meses de dezembro de 2013 e 2014, em que ocorrem em sequência um valor muito acima da média e outro muito abaixo.

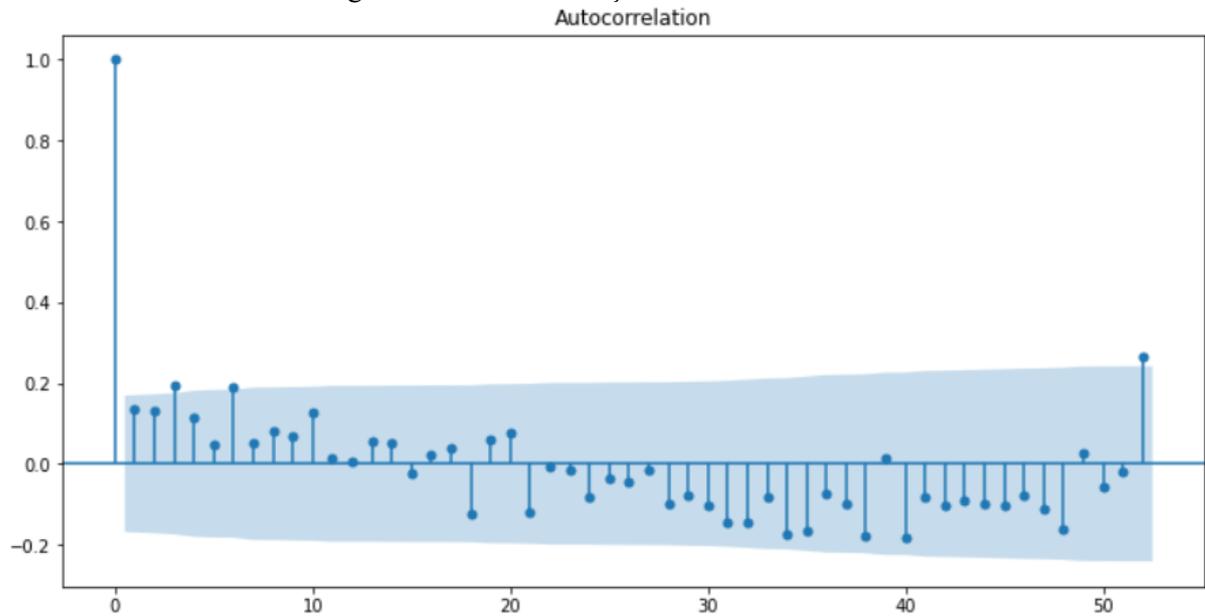
Para melhor visualizar as características dos dados a figura 11 traz a decomposição desta série temporal, de forma que é possível analisar separadamente os principais componentes da série, a tendência, a sazonalidade e o resíduo e a figura 12 mostra a auto correlação desta série.

Figura 11 - Decomposição série temporal das vendas da rede Rossmann.



Fonte: Autor (2021)

Figura 12 - Auto correlação vendas Rossmann.

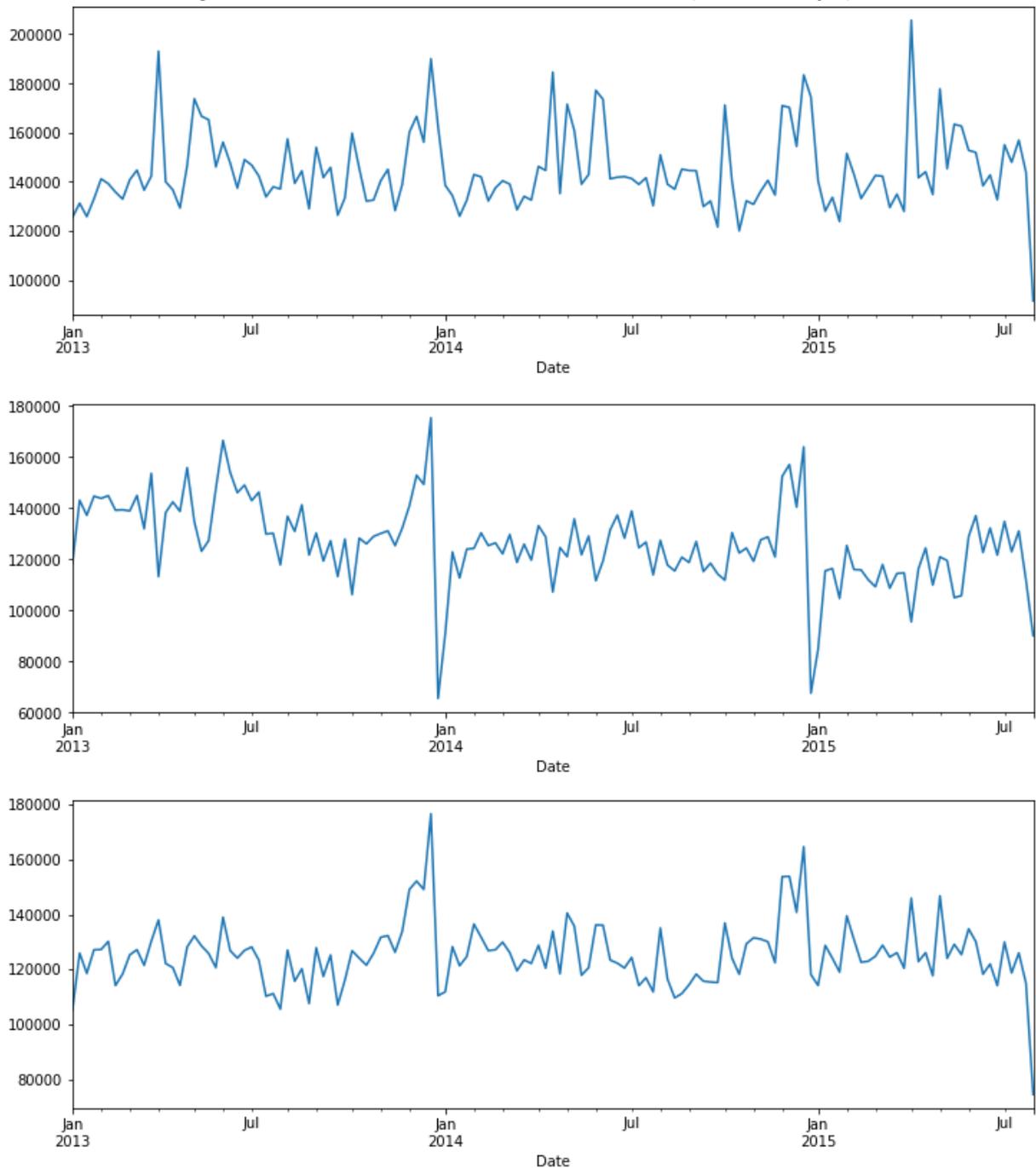


Fonte: Autor (2021)

A partir destes gráficos é possível perceber que de fato os valores não apresentam uma sazonalidade evidente, além de que existe muito ruído presente nos valores, e uma baixa auto correlação na série, o que pode dificultar a sua previsão, em especial para os métodos tradicionais que são fortemente dependentes de princípios estatísticos de regressão.

A segunda perspectiva de previsão foi a previsão paralela para as 75 lojas da rede com maior valor de vendas totais para o período. Esta perspectiva apresenta uma menor granularidade, de forma que as séries temporais analisadas correspondem às vendas específicas de cada loja. A figura 13 traz as séries das 3 lojas com maiores valores de vendas no período.

Figura 13 - Valores históricos de venda Rossmann (3 maiores lojas).

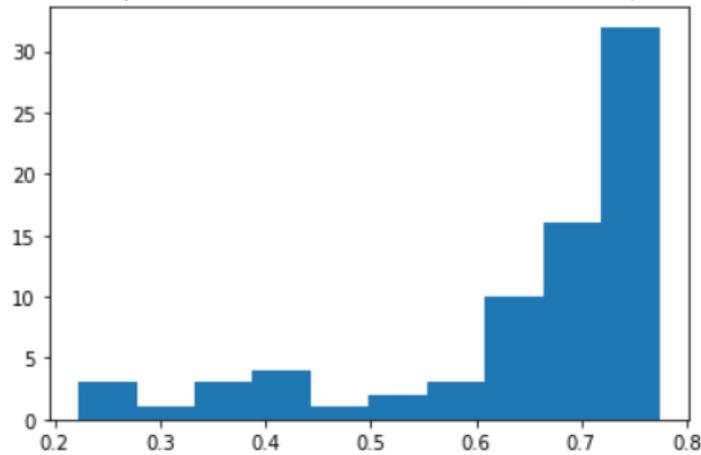


Fonte: Autor (2021)

A partir deste gráfico é possível perceber que as séries temporais específicas para cada loja tendem a apresentar um comportamento similar à série de todas as lojas agrupadas, na questão de apresentarem valores serrilhados, sem sazonalidade evidente. Quanto a relação entre as lojas, a figura 14 mostra um histograma composto pelos coeficientes de correlação entre as séries. É possível perceber que a maioria dos valores apresentam coeficientes acima de 0,6, o que indica uma forte correlação entre as vendas das lojas analisadas, uma grande

vantagem para os métodos de *machine learning*, que realizam a previsão conjunta destas séries.

Figura 14 - Histograma dos coeficientes de correlação das lojas Rossmann.



Fonte: Autor (2021)

4.2.2 TÊXTIL

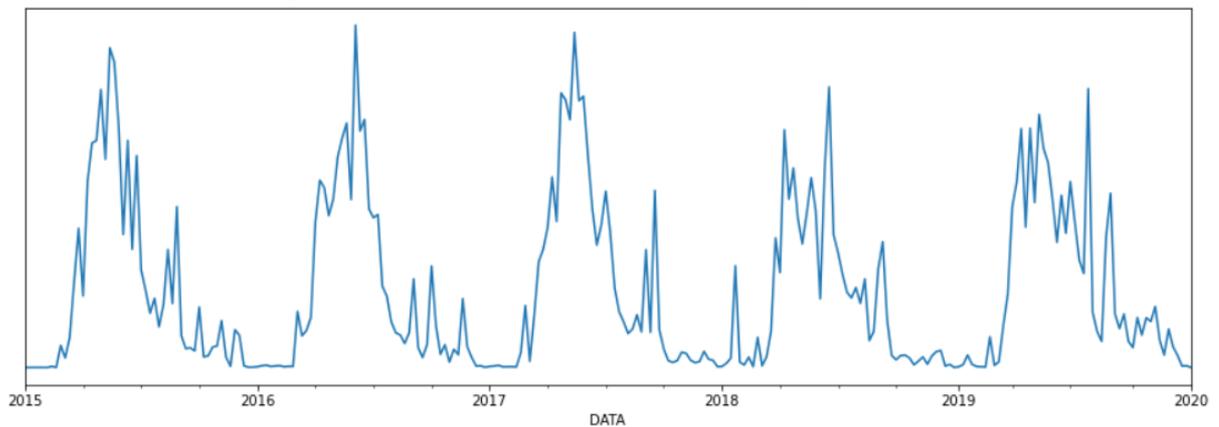
Os dados da empresa têxtil são dados privados referentes às vendas entre janeiro de 2015 e outubro de 2020 para 60 coleções produzidas pela empresa, porém apenas a coleção com o maior número de vendas foi utilizada para o desenvolvimento do trabalho, pois esta era a coleção com maior representatividade entre todas, contribuindo com aproximadamente 18% das vendas totais da empresa no período em questão.

O horizonte de análise, entretanto, foi reduzido para os valores entre janeiro de 2015 e dezembro de 2019, ou seja, desconsiderando o ano de 2020, pois as alterações causadas pela pandemia deste ano poderiam prejudicar a capacidade de previsão por se tratar de um cenário sem precedentes.

O conjunto de dados fornecidos foi elaborado com base em informações de vendas da empresa, de forma que cada linha representa uma venda de um produto. Nesta perspectiva, para cada venda as informações que foram disponibilizadas neste conjunto de dados são: o id específico referente ao SKU do produto vendido, o grupo do qual o produto faz parte, a quantidade de unidades de produtos vendida, o valor da venda, a região onde a venda ocorreu, o canal de vendas e a data da venda.

Para o desenvolvimento do trabalho este conjunto de dados foi utilizado em duas perspectivas de granularidade diferentes, sendo uma delas a previsão geral da coleção analisada, e a outra a previsão paralela de diferentes grupos de produtos pertencentes a esta coleção, com distinção da região onde ocorreu a venda, de maneira análoga ao que foi realizado para as vendas da farmacêutica Rossmann. A figura 15 traz a série temporal para as vendas totais da coleção analisada durante o período de análise.

Figura 15 - Valores históricos de venda da empresa têxtil.

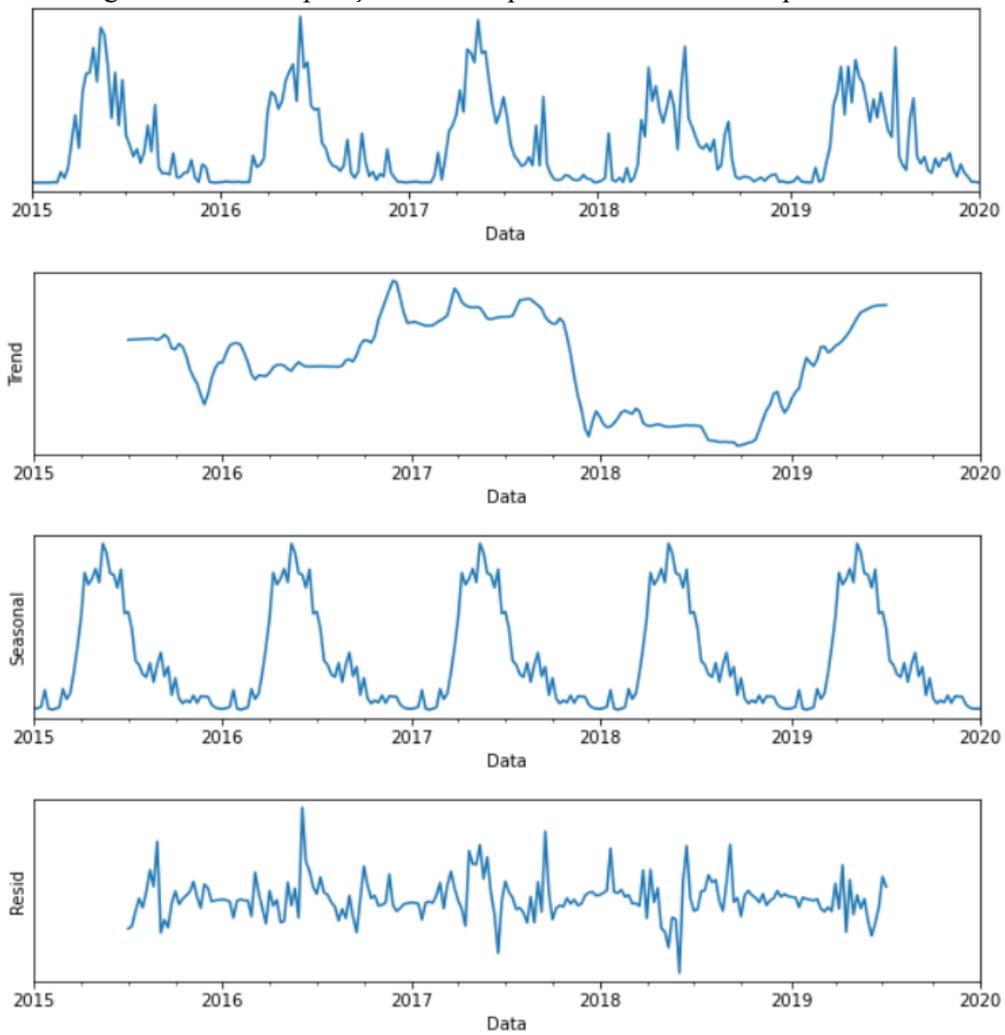


Fonte: Autor (2021)

Este gráfico permite perceber uma forte presença de sazonalidade nas vendas da empresa, devido ao comportamento do setor têxtil. A coleção analisada é uma coleção de roupas de verão e como a maior parte das vendas da empresa é destinada para revendedores, que realizam os pedidos com antecedência para terem as roupas disponíveis no início da estação, as vendas apresentam valores elevados ao final de cada verão.

A figura 16 traz a decomposição desta série temporal de maneira que fica evidente a presença de uma forte sazonalidade nas vendas, além de uma tendência que também é perceptível e valores de resíduo relativamente pequenos.

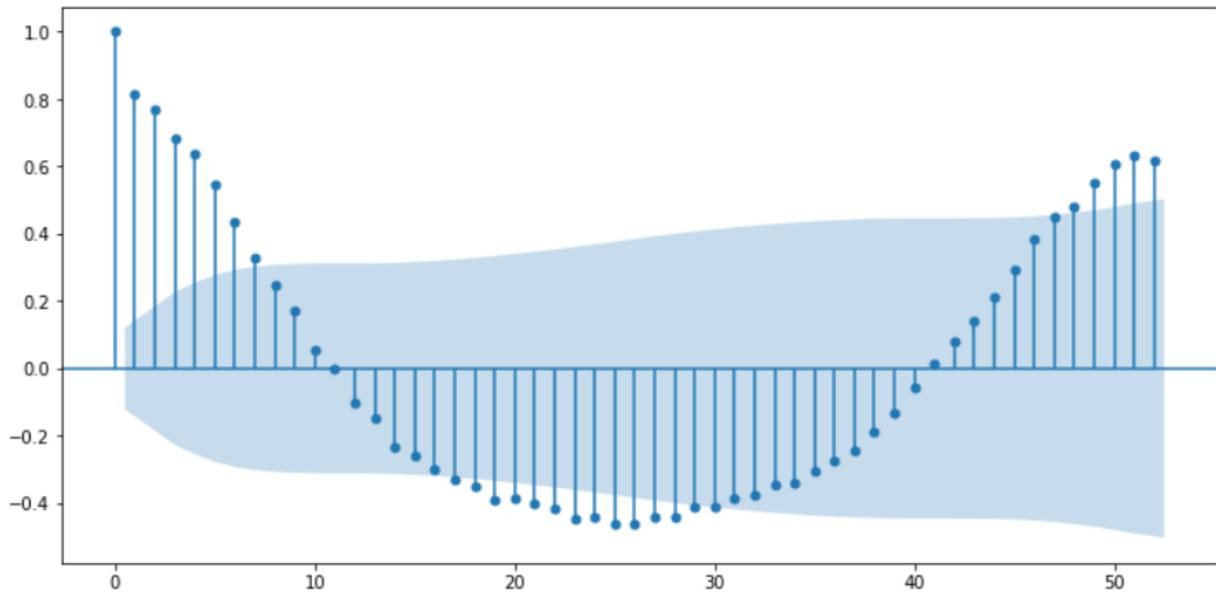
Figura 16 - Decomposição série temporal das vendas da empresa têxtil.



Fonte: Autor (2021)

A figura 17 a seguir traz ainda a auto correlação dos valores de venda da empresa para o período de um ano, e reforça a forte influência da sazonalidade nesta série temporal.

Figura 17 - Auto correlação vendas têxtil.
Autocorrelation

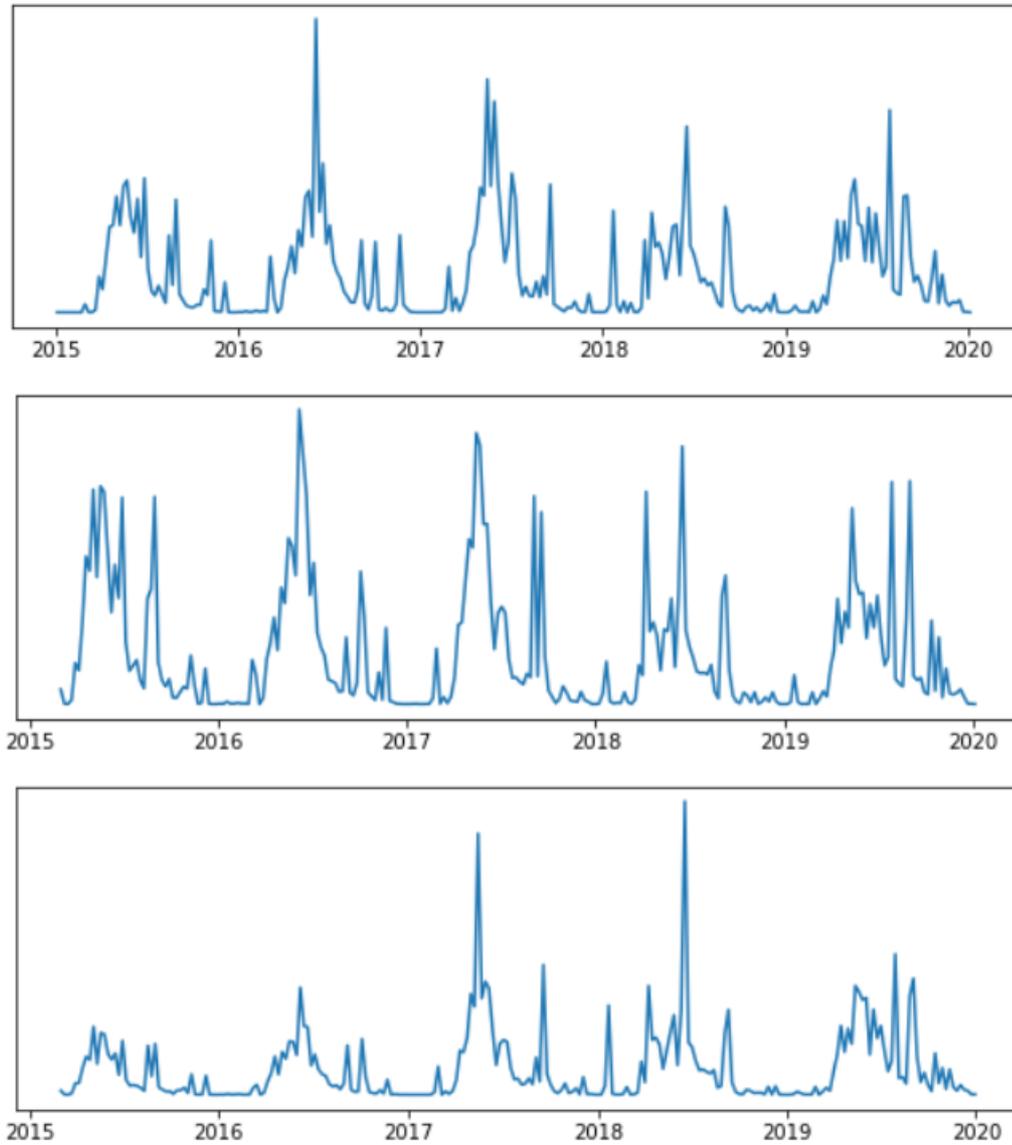


Fonte: Autor (2021)

A partir destes resultados podemos perceber que, ao contrário do que ocorre com a série de vendas da farmacêutica, para a empresa têxtil existe uma vantagem para os métodos tradicionais que se baseiam em modelos de regressão e se beneficiam de comportamentos sazonais.

A segunda perspectiva de previsão para a empresa têxtil foi realizada com o intuito de prever séries multivariadas paralelas. Para isto, foram utilizadas como base as séries temporais para alguns grupos de produtos, sendo que cada grupo foi dividido de acordo com a região em que ocorreu a venda. Assim, foram selecionados os 49 pares de grupo-região que apresentavam maior quantidade de vendas para as previsões. A figura 18 mostra o histórico de vendas para os 3 pares mais representativos, que são, respectivamente, a venda de vestidos, camisetas com bermudas e blusas com shorts, todos para a região sudeste.

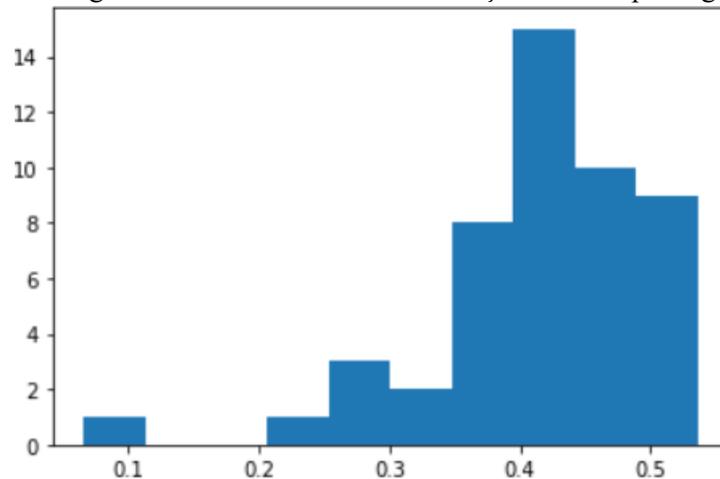
Figura 18 - Valores históricos de venda da empresa têxtil (3 maiores pares grupo-região).



Fonte: Autor (2021)

Estes gráficos permitem visualizar que a venda dos grupos tende a apresentar um comportamento similar ao comportamento da coleção como um todo, porém com maiores variações em relação às médias. A relação entre estas linhas, no entanto, não é tão significativa quanto a relação entre as séries das lojas da rede Rossmann, o que pode ser observado pelos valores de correlação entre estas séries, explicitados no histograma da figura 19.

Figura 19 - Histograma dos coeficientes de correlação entre os pares grupo-região.



Fonte: Autor (2021)

Esta baixa correlação entre as linhas indica que, para este conjunto de dados, os modelos que realizam a previsão conjunta das séries temporais não devem se beneficiar muito com esta capacidade.

4.3 DEFINIÇÃO BASE PARA COMPARAÇÃO

Para que a comparação entre os diferentes métodos seja efetiva é necessário garantir que todos os métodos sejam aplicados e avaliados em um mesmo cenário. Para isto foi definido o cenário utilizado como base para comparação, e uma engenharia de dados foi realizada em ambos os conjuntos para adequá-los a este cenário de base. O cenário foi definido padronizando o nível de agregação de tempo, o horizonte de previsão e o cálculo de erro da previsão.

Os valores de ambos os conjuntos foram agrupados semanalmente, de maneira que cada valor da série temporal corresponde à soma de todas as vendas de uma semana, para obter-se séries com menores variações em média e facilitar as previsões. O horizonte de previsão foi definido como o último ano de dados de cada conjunto e o RMSE foi definido como a métrica de erro a ser utilizada. Desta forma, cada modelo estudado foi aplicado um total de 4 vezes, prevendo a série temporal da empresa Rossmann agregada e desagregada a nível das lojas, e a série temporal da empresa têxtil, também na perspectiva agregada e desagregada, a nível dos pares grupo-região de produto da coleção estudada.

Cada previsão foi feita utilizando os valores anteriores ao último ano conhecido como base de treino, para que os modelos pudessem assimilar as características das séries

temporais, gerando uma série de 52 valores, equivalente ao ano que foi desconsiderado para o treino. O erro é calculado comparando esta série prevista com os valores reais que haviam sido desconsiderados, através da métrica selecionada, de maneira que o intervalo previsto não é utilizado para o treino dos modelos, garantindo que não existam problemas ligados ao vazamento de dados (*data leakage*) do período previsto. Para as previsões com menor agregação o erro total é calculado a partir da média ponderada do erro de cada série paralela em função do seu valor total de venda.

4.4 APLICAÇÃO DOS MÉTODOS

Este capítulo demonstra como cada método foi aplicado durante a execução do trabalho e seus resultados. Os parâmetros utilizados em cada modelos foram otimizados, a partir de intervalos de valores que tendem a apresentar bons resultados, de maneira a reduzir o erro de cada previsão e a demonstração de todos os métodos seguem a seguinte ordem em relação às séries previstas:

1. Rossmann (agregado)
2. Rossmann (segregado por loja)
3. Têxtil (agregado)
4. Têxtil (segregado por par grupo-região)

Todos os gráficos apresentados nesta seção seguem o mesmo padrão de cores para padronizar a visualização. A cor azul demonstra os valores utilizados para treinar o modelo, a cor laranja demonstra os valores reais do período de previsão e a cor verde demonstra a previsão realizada pelo modelo em questão. Quanto à legenda, para as previsões segregadas, a legenda referente a cor verde informa o número identificador da loja ou o par grupo-região o qual o gráfico retrata.

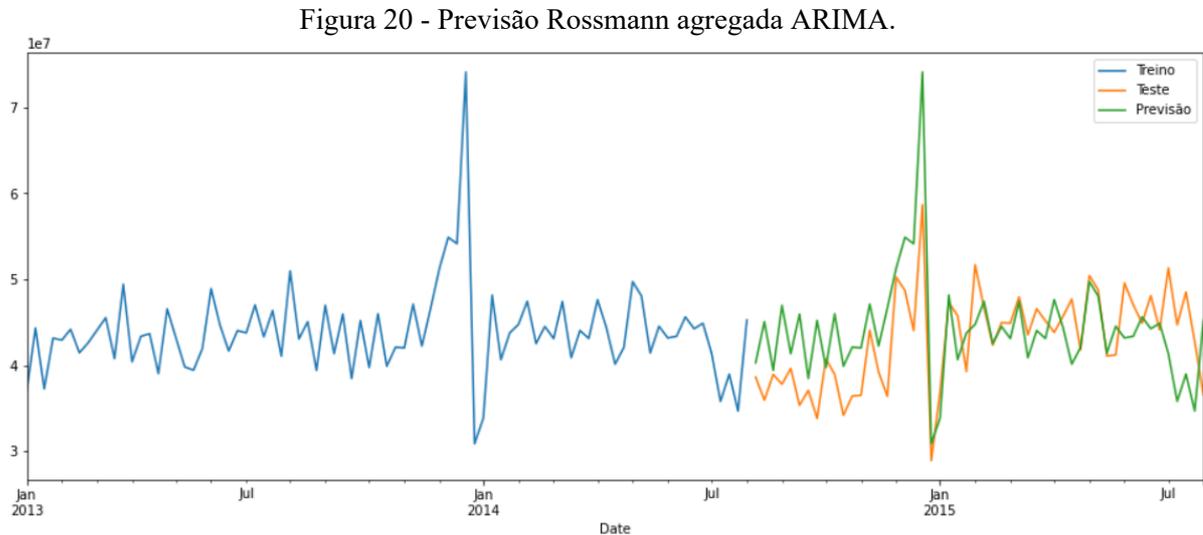
4.4.1 ARIMA

O método ARIMA foi aplicado utilizando o pacote para python pmdarima, que identifica automaticamente os melhores parâmetros para o modelo, considerando também a perspectiva sazonal da série temporal prevista. Este pacote é baseado nos estudos de

Hyndman e Khandakar (2008), e adapta o pacote desenvolvido para a linguagem R para a aplicação em Python.

A previsão para a linha geral da Rossmann foi feita com base em um modelo ARIMA com coeficientes $p=0$, $d=0$, $q=1$, $P=0$, $D=1$, $Q=0$ para uma sazonalidade anual.

Esta previsão apresentou um erro de 5.822.762 e pode ser vista na figura 20.



Fonte: Autor (2021)

A previsão a nível segregado para a Rossmann foi feita com base em um modelo com parâmetros similares aos da linha geral ($p=0$, $d=0$, $q=1$, $P=0$, $D=1$, $Q=0$ para uma sazonalidade anual) que apresentou um erro de 10.076. O erro das 3 lojas com maior participação nas vendas pode ser visto na tabela 1 e sua previsão pode ser vista na figura 21.

Tabela 1- Erro da previsão Rossmann segregada ARIMA.

Loja	Total de vendas	Erro da previsão
262	19.516.842	16.362
817	17.057.867	13.108
562	16.927.322	10.560

Fonte: Autor (2021)

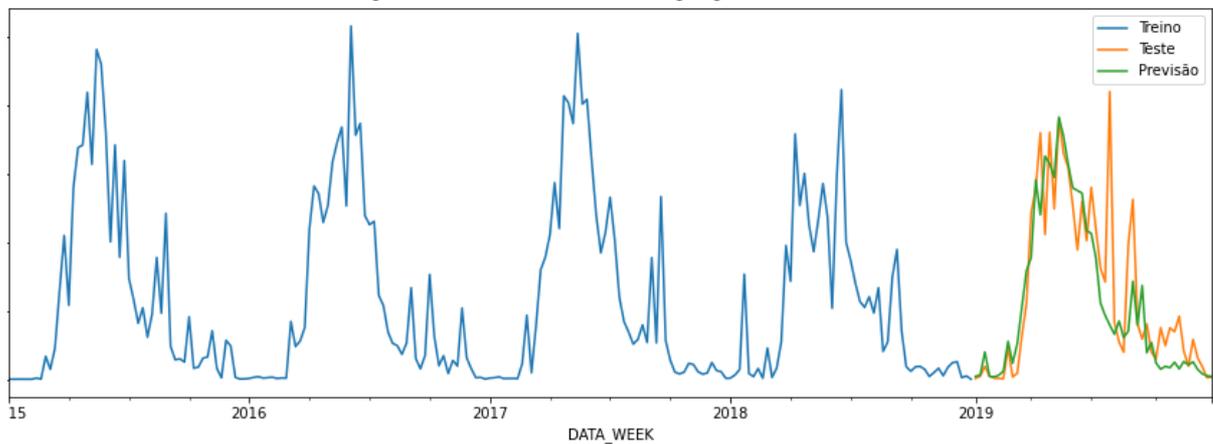
Figura 21 - Previsão Rossmann segregada ARIMA.



Fonte: Autor (2021)

A previsão para a linha geral da empresa têxtil foi feita com base em um modelo ARIMA com coeficientes $p=2$, $d=1$, $q=1$, $P=1$, $D=1$, $Q=1$. Esta previsão apresentou um erro de 33.836 e pode ser vista na figura 22.

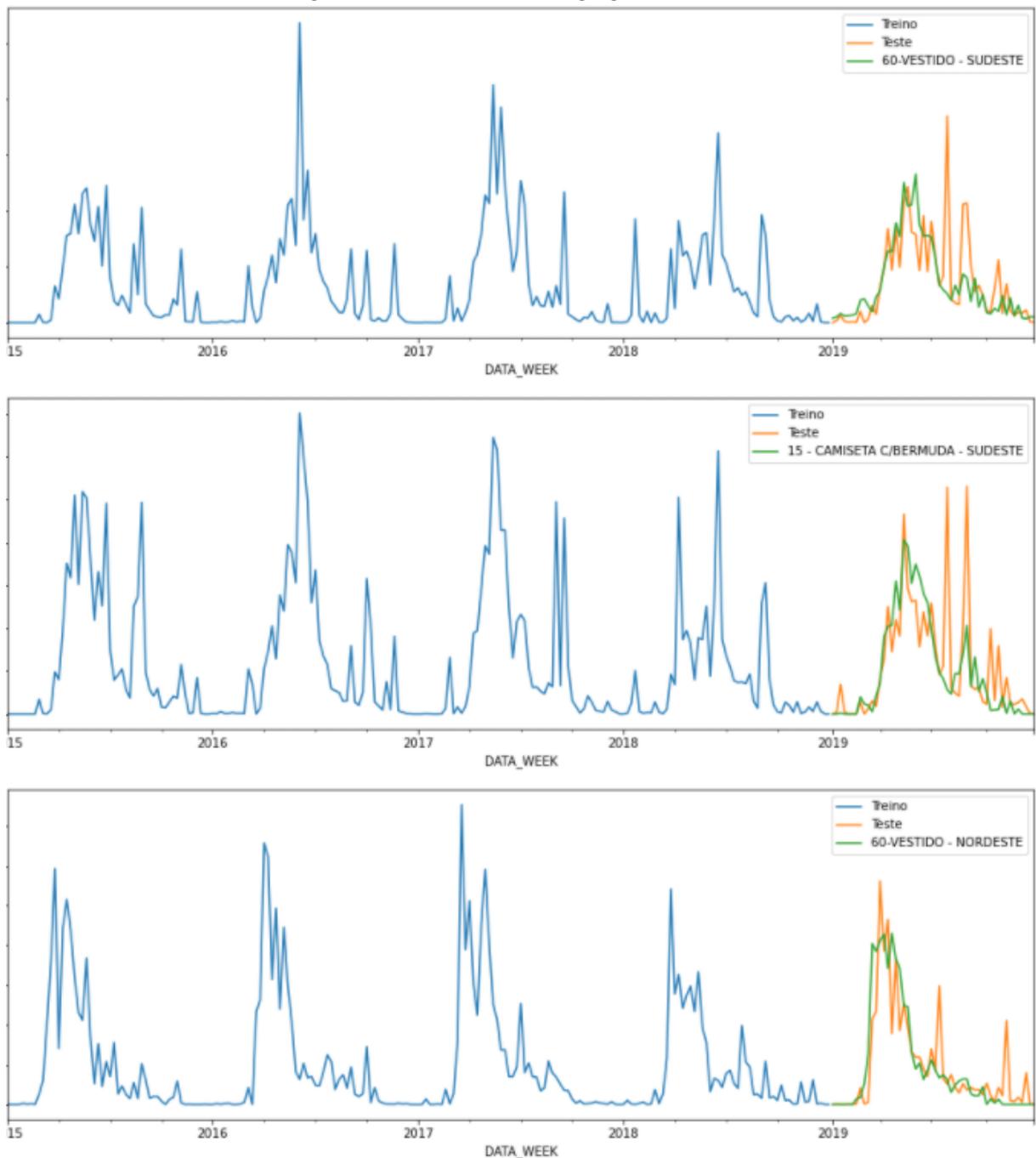
Figura 22 - Previsão têxtil agregada ARIMA.



Fonte: Autor (2021)

A previsão a nível de par grupo-produto da empresa têxtil foi feita com base em um modelo com parâmetros similares aos da linha geral ($p=2$, $d=1$, $q=1$, $P=1$, $D=1$, $Q=1$ para sazonalidade anual), que apresentou um erro de 2.665. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 23.

Figura 23 - Previsão têxtil segregada ARIMA.



Fonte: Autor (2021)

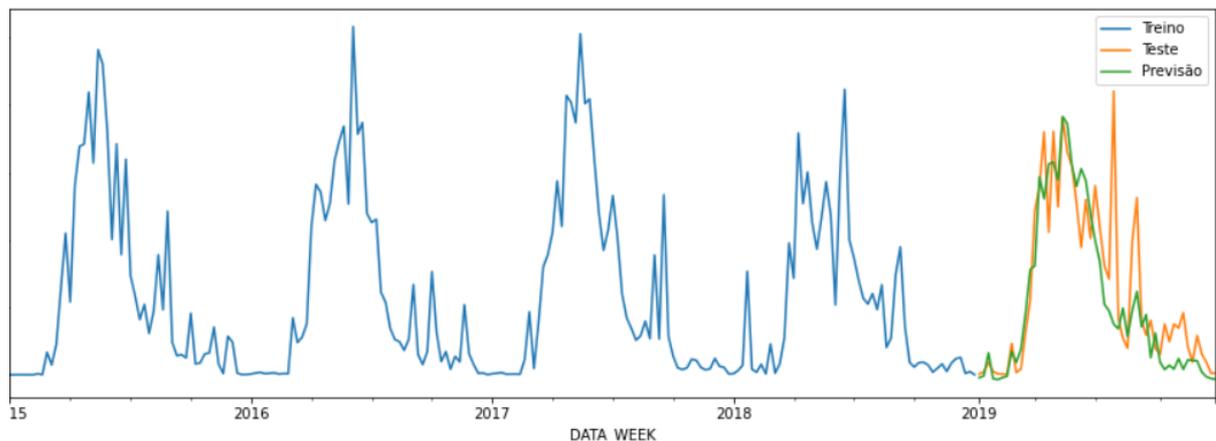
4.4.2 Holt-Winters

O modelo Holt-Winter foi implementado através do pacote *statsmodels* para Python, elaborado por Seabold e Perktold (2010). Este foi o único modelo que não pode ser aplicado para todas as séries de interesse devido às suas restrições, pois este modelo exige ao menos duas vezes o período da sazonalidade para o treino. Ambos os conjuntos de dados do trabalho

apresentam sazonalidade anual, e, portanto, é necessário fornecer uma série com pelo menos dois anos para o modelo, o que é impossível para o conjunto da Rossmann, pois não há dados suficientes. Assim, este modelo foi aplicado apenas para a empresa têxtil.

A previsão para a linha geral foi realizada através de um modelo com os parâmetros “smoothing_level”= $\alpha=0,1$, “smoothing_seasonal”= $\beta=0,1$ e “smoothing_slope”= $\gamma=0,01$, e com sazonalidade aditiva e apresentou um erro de 34.436 e pode ser visualizada na figura 24.

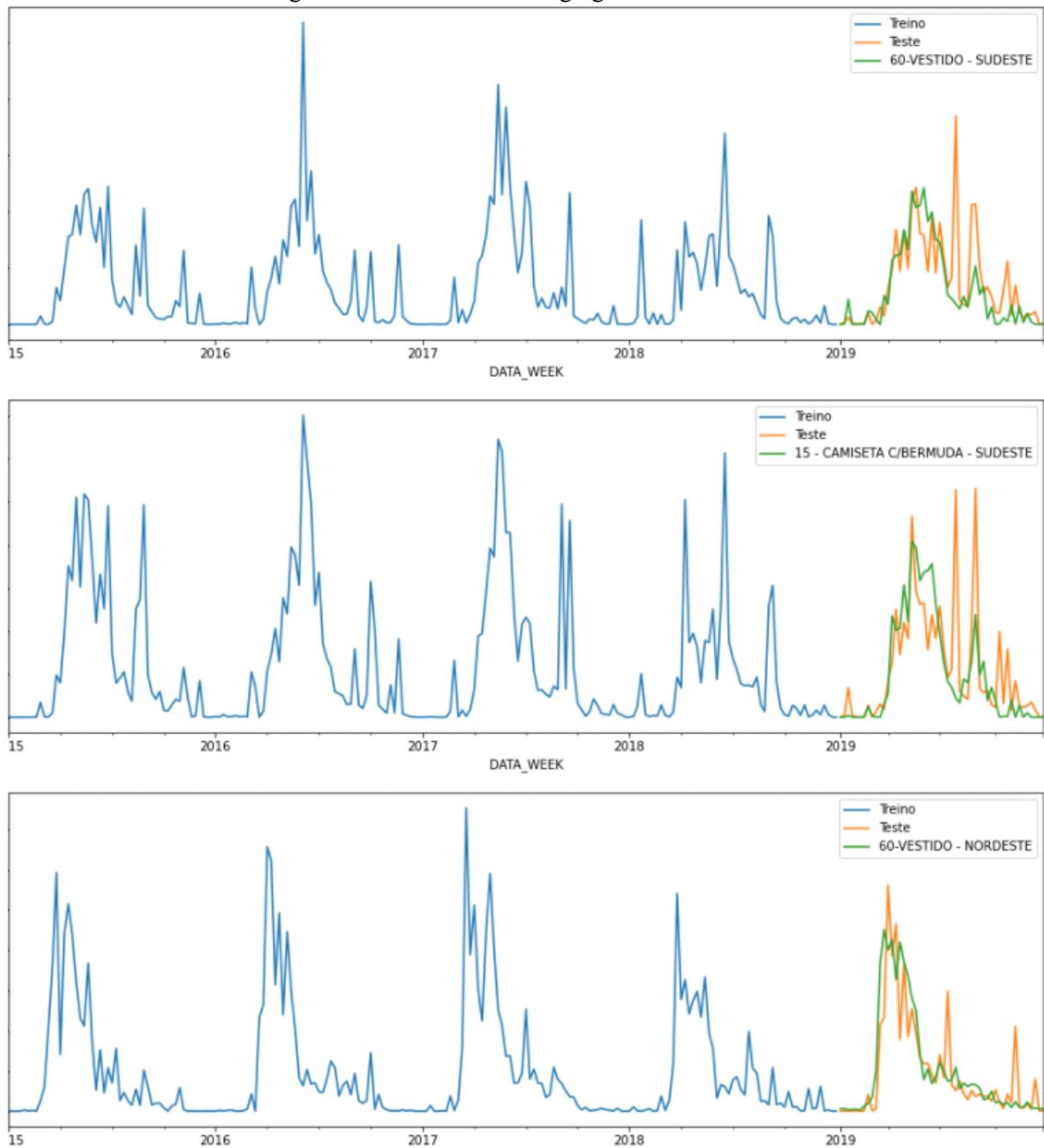
Figura 24 - Previsão têxtil agregada Holt-Winters.



Fonte: Autor (2021)

A previsão a nível segregado utilizou os mesmos parâmetros da linha agregada, e apresentou um erro de 2.615. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 25.

Figura 25 - Previsão têxtil segregada Holt-Winters.



Fonte: Autor (2021)

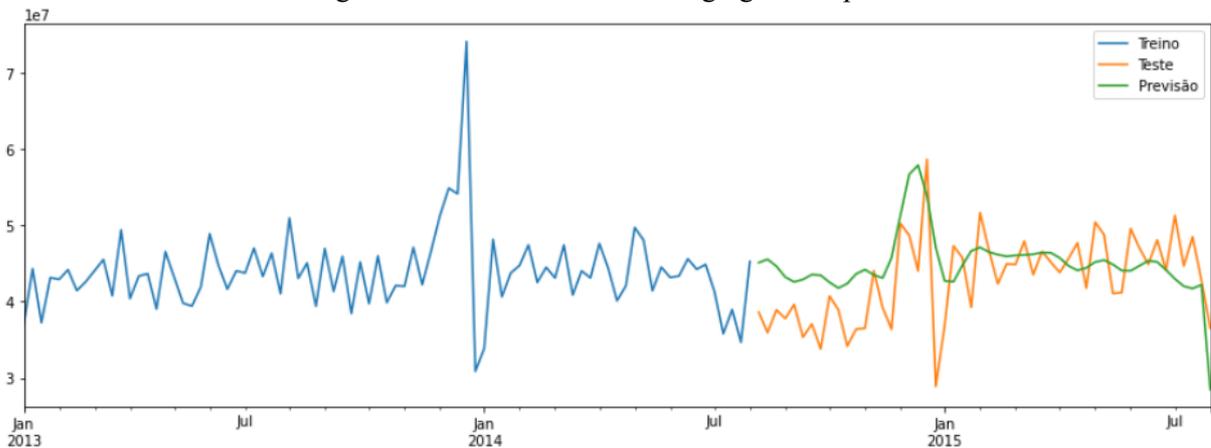
4.4.3 Prophet

O modelo Prophet foi aplicado fazendo uso do pacote *fbprophet*, elaborado com base nos estudos divulgados por Taylor e Letham (2018). Este modelo apresentou uma peculiaridade em relação aos demais, que é a incapacidade da previsão com agregação semanal, o que fez com que as previsões por meio deste modelo tenham sido realizadas em

uma perspectiva diária, e posteriormente agrupadas na perspectiva semanal para a comparação.

A previsão da linha agregada da Rossmann foi realizada com os seguintes parâmetros: “changepoint_prior_scale”= 0,01; “changepoint_range”=0,95; “holidays_prior_scale”=0,1; “seasonality_mode”=multiplicativa; “seasonality_prior_scale”=0,1. Esta previsão apresentou um erro de 5.848.904 e pode ser vista na figura 26.

Figura 26 - Previsão Rossmann agregada Prophet.



Fonte: Autor (2021)

Visualmente fica claro que este modelo não assimila bem as características da série, mesmo seu erro sendo próximo ao erro apresentado pelo modelo ARIMA, o que ocorre pois a previsão pelo modelo Prophet tende a apresentar resultados próximos à média, o que gera um RMSE pequeno, pois essa métrica é caracterizada por punir grandes desvios e assim beneficiar previsões em torno da média.

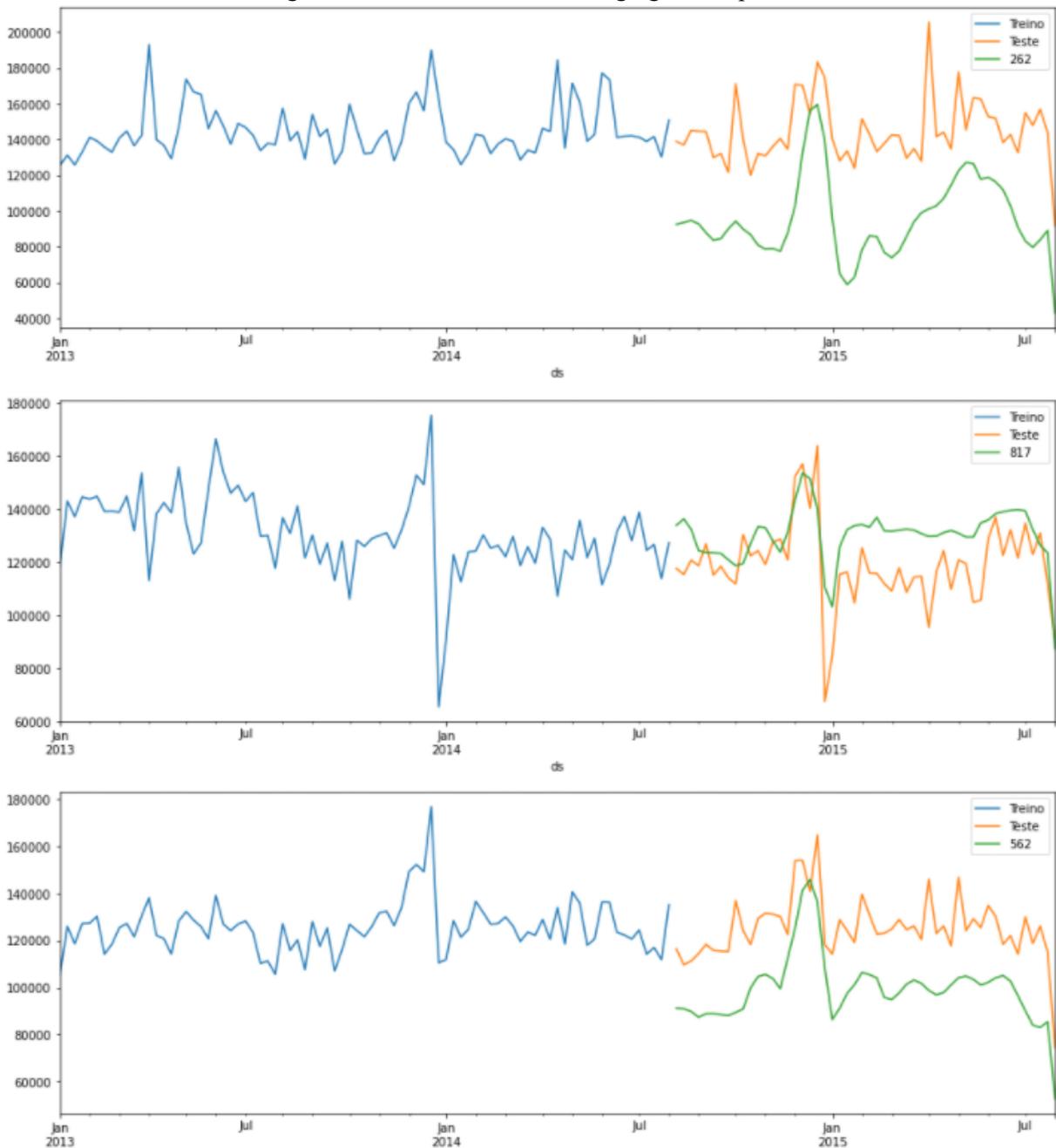
A previsão da linha segregada da Rossmann foi realizada com os mesmos parâmetros que a previsão anterior, e apresentou um erro de 11.075. O erro das 3 lojas com maior participação nas vendas pode ser visto na tabela 2 e sua previsão pode ser vista na figura 27.

Tabela 2 - Erro da previsão Rossmann segregada Prophet.

Loja	Total de vendas	Erro da previsão
262	19.516.842	52.116
817	17.057.867	16.077
562	16.927.322	27.433

Fonte: Autor (2021)

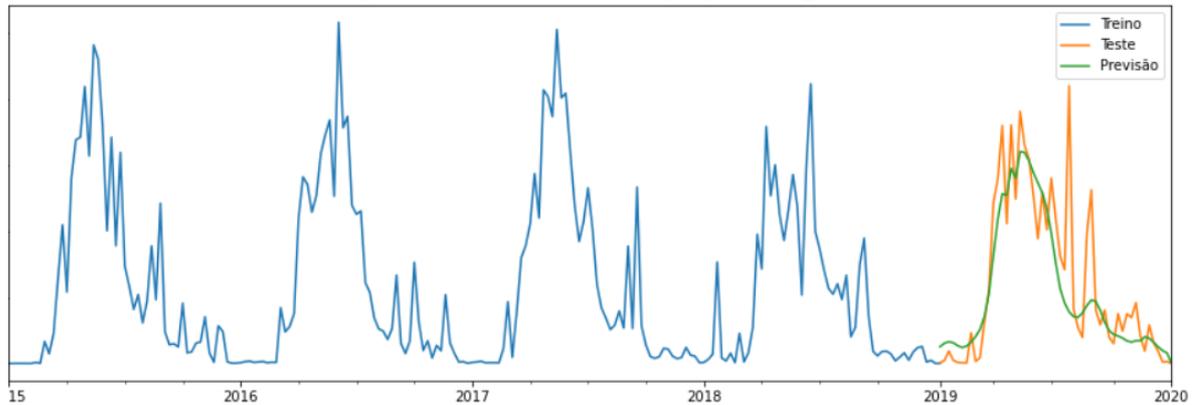
Figura 27 - Previsão Rossmann segregada Prophet.



Fonte: Autor (2021)

A previsão da linha agregada da empresa têxtil foi realizada com os seguintes parâmetros: “changepoint_prior_scale”=0,1; “changepoint_range”=0,95; “holidays_prior_scale”=5; “seasonality_mode”=aditiva; “seasonality_prior_scale”=0,01. Esta previsão apresentou um erro de 33.233 e pode ser vista na figura 28.

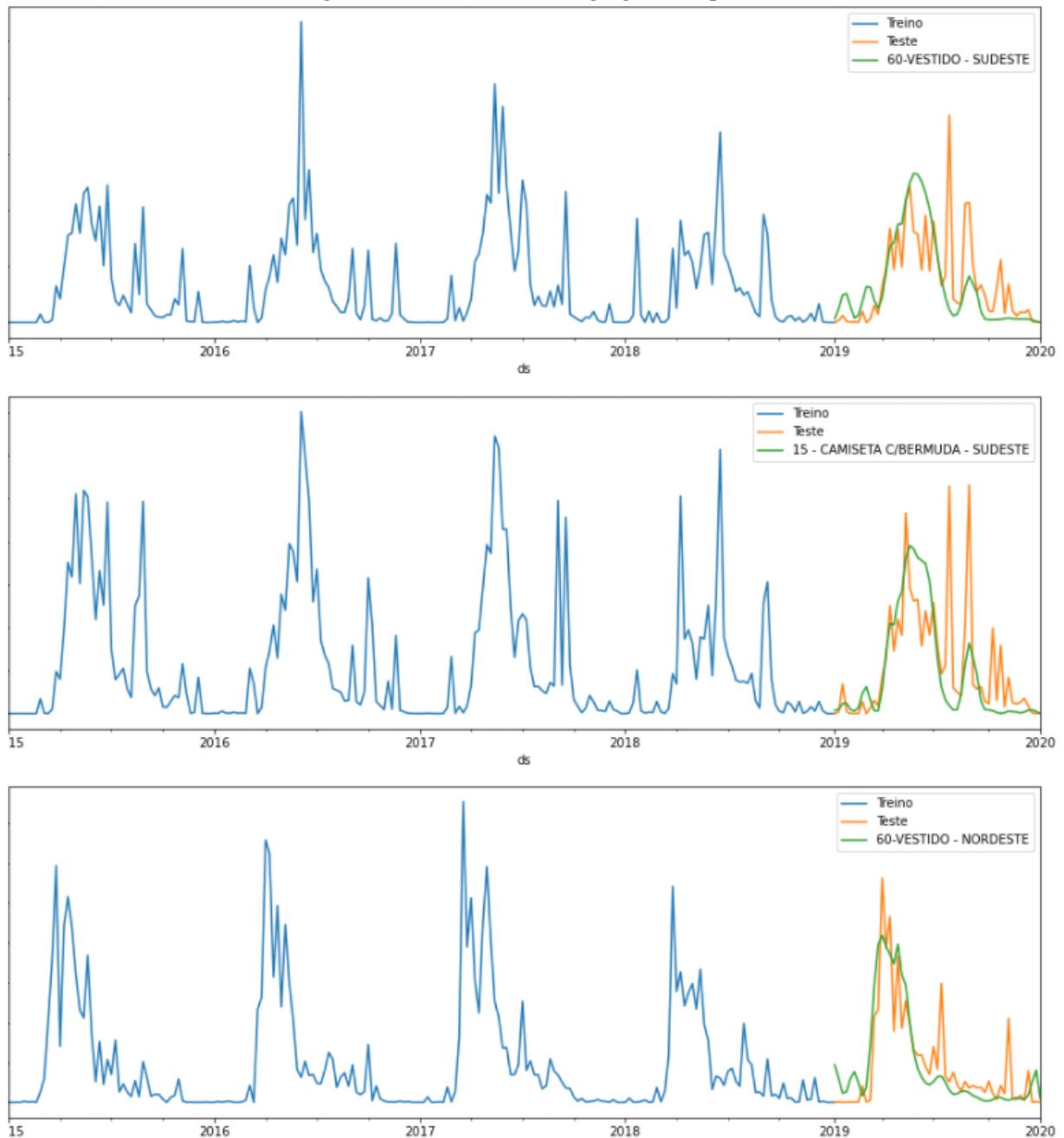
Figura 28 - Previsão têxtil agregada Prophet.



Fonte: Autor (2021)

A previsão a nível segregado utilizou os mesmos parâmetros da linha agregada, e apresentou um erro de 2.880. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 29.

Figura 29- Previsão têxtil segregada Prophet.



Fonte: Autor (2021)

4.4.4 Decision Tree

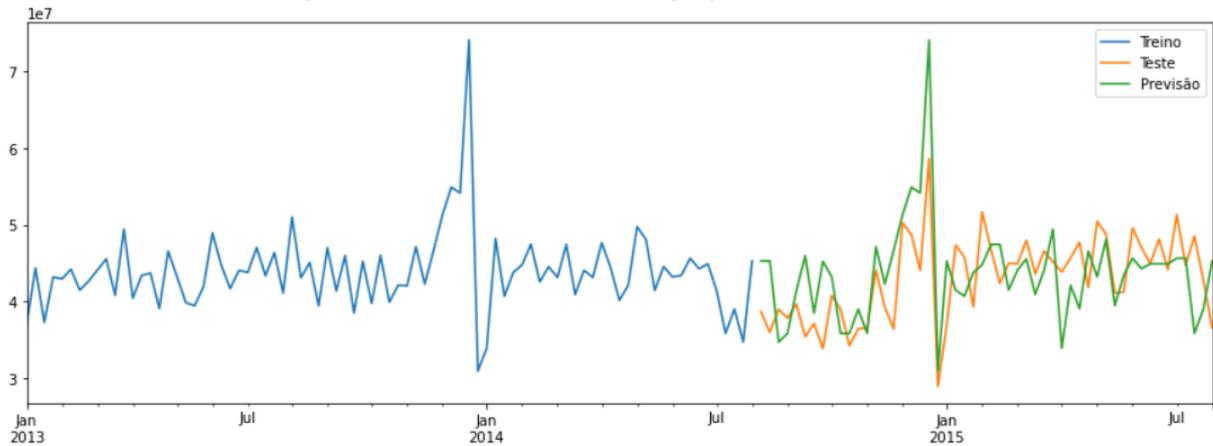
O modelo Decision Tree foi o primeiro modelo de *machine learning* aplicado, e para isso foi utilizado o pacote *scikit-learn*, desenvolvido por Pedregosa et al. (2011).

A previsão da série Rossmann agregada foi utilizada com os seguintes parâmetros:

"max_depth"=*None*; "max_features"=*None*; "max_leaf_nodes"=*None*;
 "min_samples_leaf"=1; "min_samples_split"=2; "min_weight_fraction_leaf"=0;

"random_state"=*None*. Esta previsão resultou em um erro de 5.877.980 e pode ser vista na figura 30.

Figura 30 - Previsão Rossmann agregada Decision Tree.



Fonte: Autor (2021)

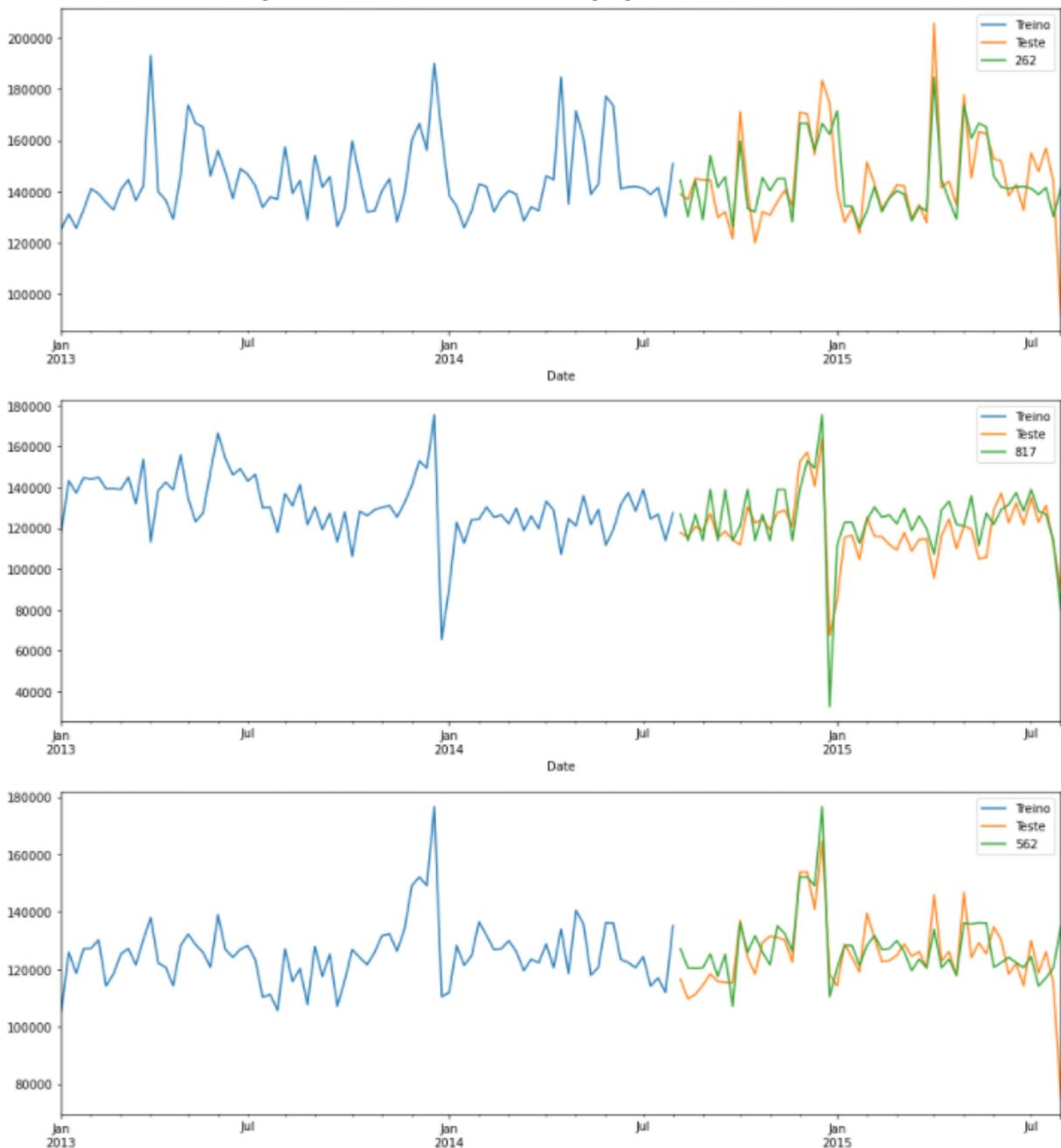
Para a previsão das séries específicas por lojas foram utilizados os mesmos parâmetros que a previsão agregada para a série da Rossmann. Esta previsão apresentou um erro de 10.892. O erro das 3 lojas com maior participação nas vendas pode ser visto na tabela 3 e sua previsão pode ser vista na figura 31.

Tabela 3 - Erro da previsão Rossmann segregada Decision Tree.

Loja	Total de vendas	Erro da previsão
262	19.516.842	12.108
817	17.057.867	11.241
562	16.927.322	10.921

Fonte: Autor (2021)

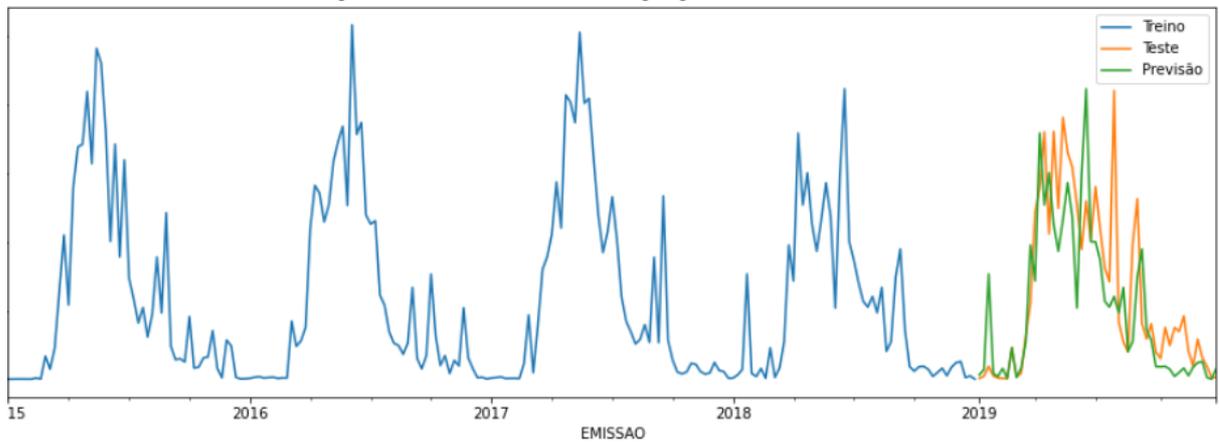
Figura 31 - Previsão Rossmann segregada Decision Tree.



Fonte: Autor (2021)

A previsão para a linha geral da empresa têxtil foi feita pelo modelo Decision Tree com os seguintes parâmetros: "max_depth"=None; "max_features"=None; "max_leaf_nodes"=None; "min_impurity_split"=0,0000001; "min_samples_leaf"=1; "min_samples_split"=2; "min_weight_fraction_leaf"=0; "random_state"=None. Esta previsão apresentou um erro de 41.076 e pode ser vista na figura 32.

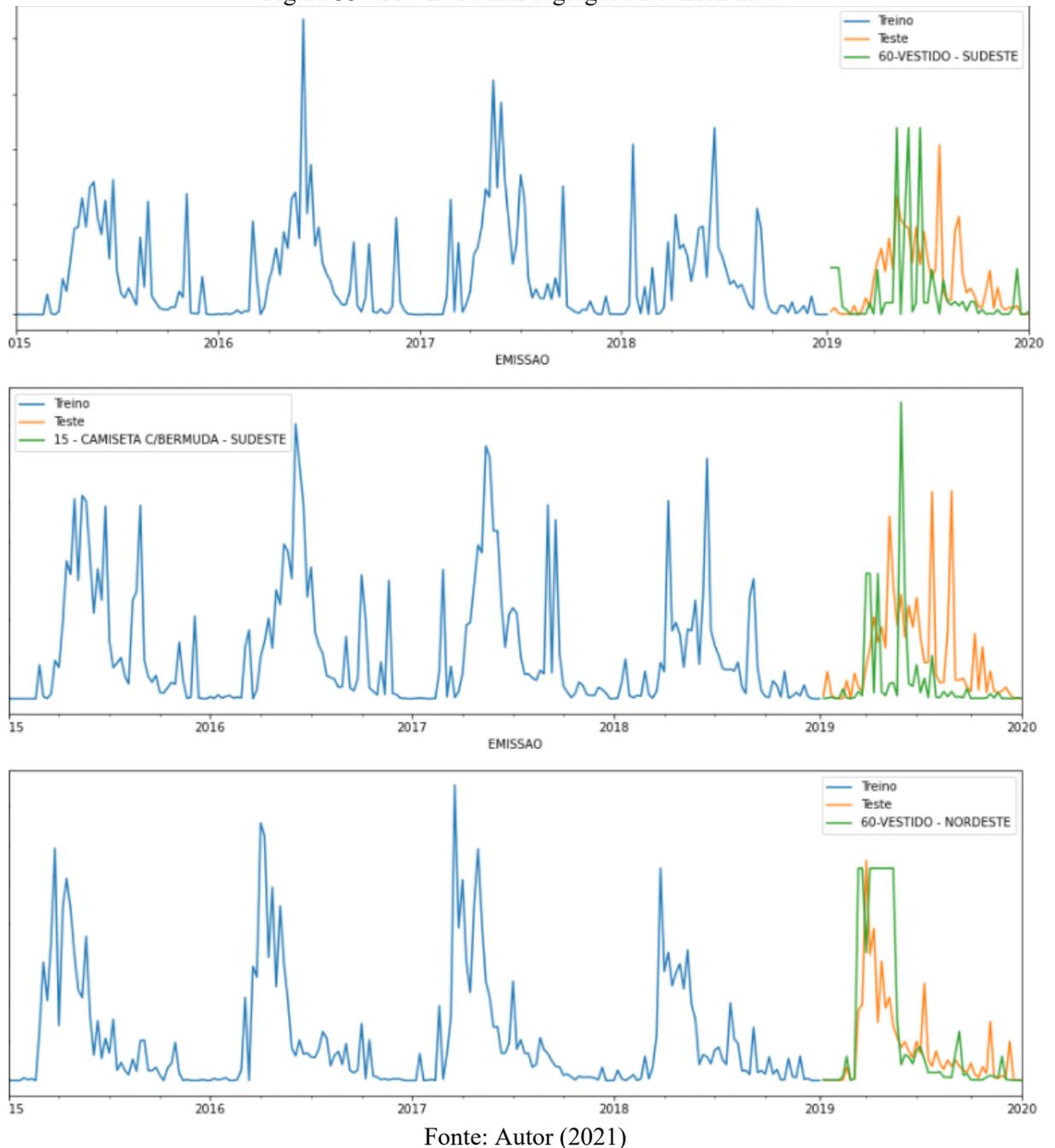
Figura 32 - Previsão têxtil agregada Decision Tree.



Fonte: Autor (2021)

A previsão a nível de par grupo-produto da empresa têxtil foi feita com base em um modelo Decision Tree com todos os parâmetros do modelo com seus valores padrões. Esta previsão apresentou um erro de 776. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 33.

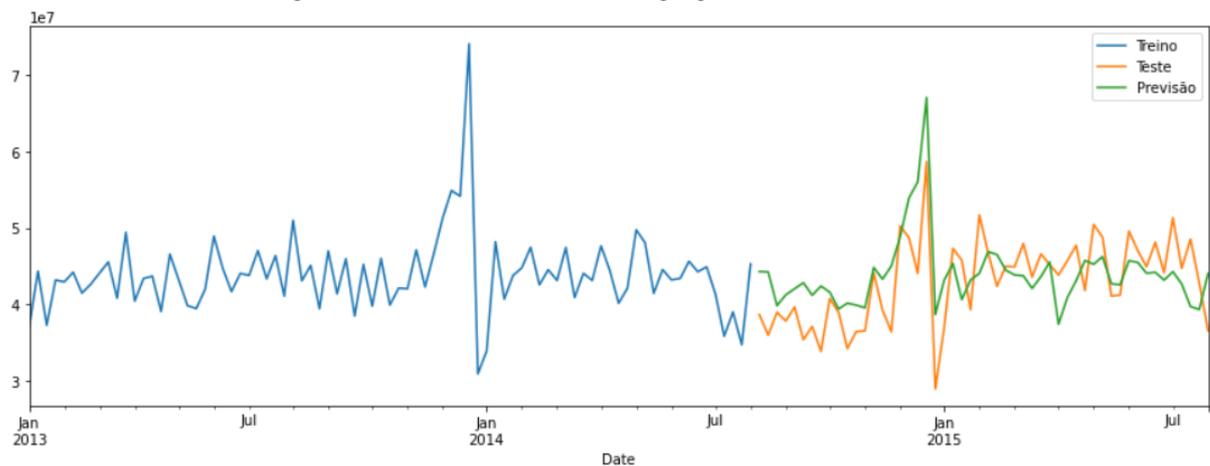
Figura 33 - Previsão têxtil segregada Decision Tree.



4.4.5 Random Forest

O modelo Random Forest, assim como o modelo anterior, foi aplicado por meio do pacote *scikit-learn* desenvolvido por Pedregosa et al. (2011). Este modelo foi aplicado para a série geral da Rossmann com os seguintes parâmetros: “min_samples_split”=0,5; “min_samples_leaf”=1; “max_features”=*None*; “max_depth”=1; “n_estimators”=1000. Esta previsão resultou em um erro de 5.262.006 e pode ser vista na figura 34.

Figura 34 - Previsão Rossmann agregada Random Forest.



Fonte: Autor (2021)

A previsão das séries específicas por lojas foi realizada utilizando os mesmos parâmetros que a previsão agregada para a série da Rossmann. Esta previsão apresentou um erro de 9.562. O erro das 3 lojas com maior participação nas vendas pode ser visto na tabela 4 e sua previsão pode ser vista na figura 35.

Tabela 4 - Erro da previsão Rossmann segregada Random Forest.

Loja	Total de vendas	Erro da previsão
262	19.516.842	11.257
817	17.057.867	9.869
562	16.927.322	9.242

Fonte: Autor (2021)

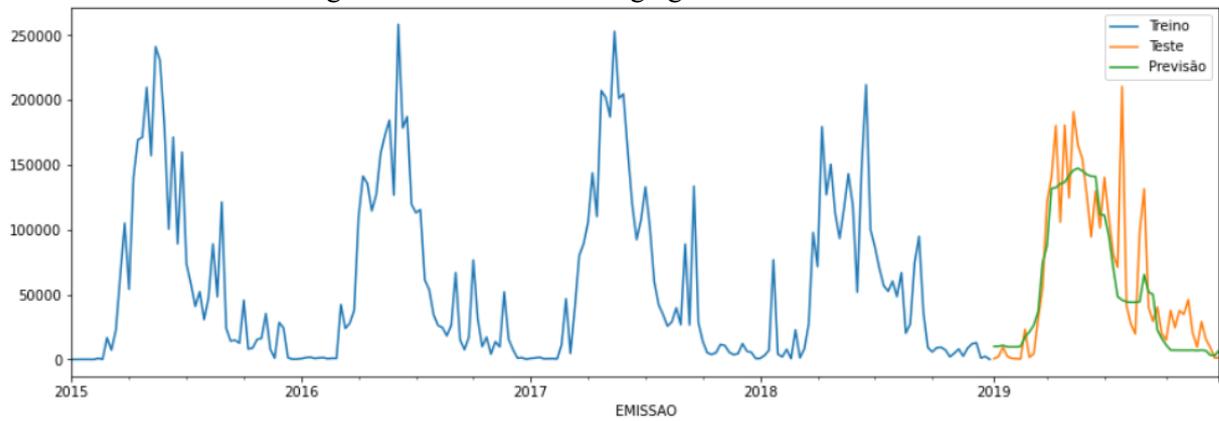
Figura 35 - Previsão Rossmann segregada Random Forest.



Fonte: Autor (2021)

A previsão para a série agregada da empresa têxtil utilizou os seguintes parâmetros: “min_samples_split”=20; “min_samples_leaf”=1; “max_features”=None; “max_depth”=30; “n_estimators”=1000. Esta previsão retornou um RMSE de 32.966 e pode ser vista na figura 36.

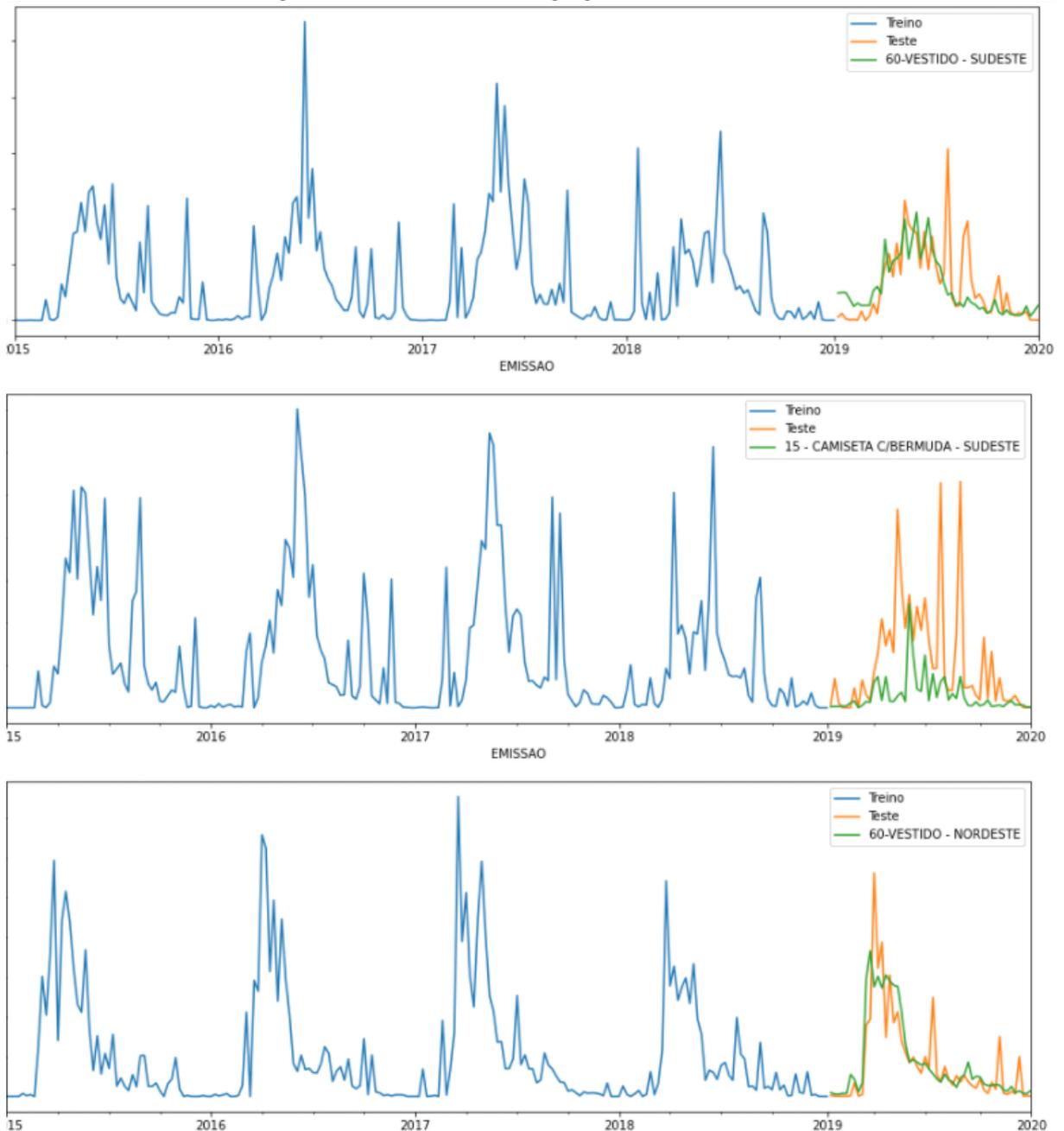
Figura 36 - Previsão têxtil agregada Random Forest.



Fonte: Autor (2021)

A previsão a nível segregado utilizou todos os parâmetros do modelo em seus valores padrões, e apresentou um erro de 542. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 37.

Figura 37 - Previsão têxtil segregada Random Forest.



Fonte: Autor (2021)

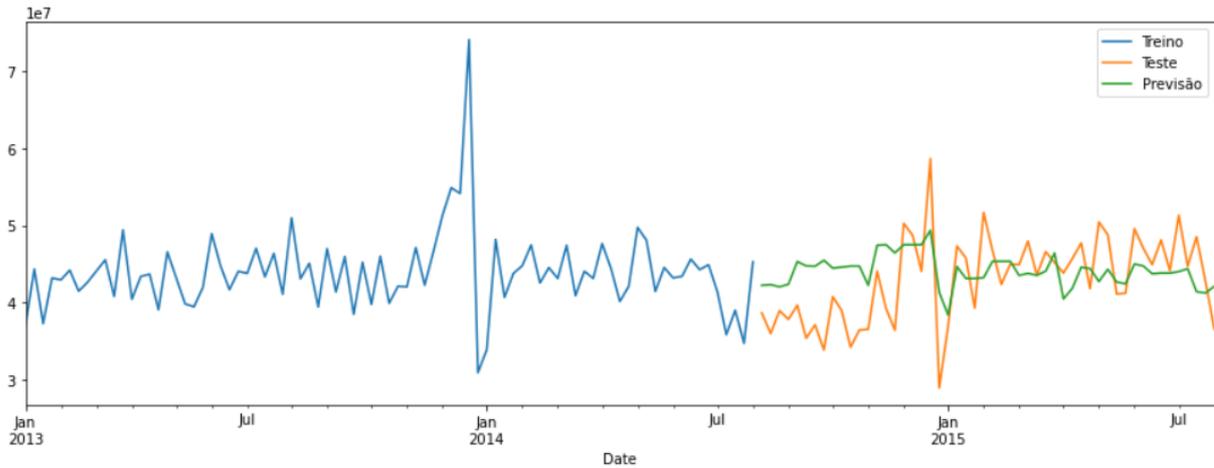
4.4.6 XGBoost

Para as previsões que utilizaram o modelo XGBoost foi utilizado o pacote XGBoost para Python, descrito por Chen e Guestrin (2016).

A previsão para a linha geral da Rossmann por este modelo utilizou os seguintes parâmetros: “subsample”=0,6; “reg_lambda”=0,45; “reg_alpha”=0,75; “n_estimators”=100; “min_child_weight”=10; “max_depth”=5; “learning_rate”=0,07; “gamma”=0;

“colsample_bytree”=0,4. Esta previsão apresentou um erro de 5.454.413 e pode ser vista na figura 38.

Figura 38 - Previsão Rossmann agregada XGBoost.



Fonte: Autor (2021)

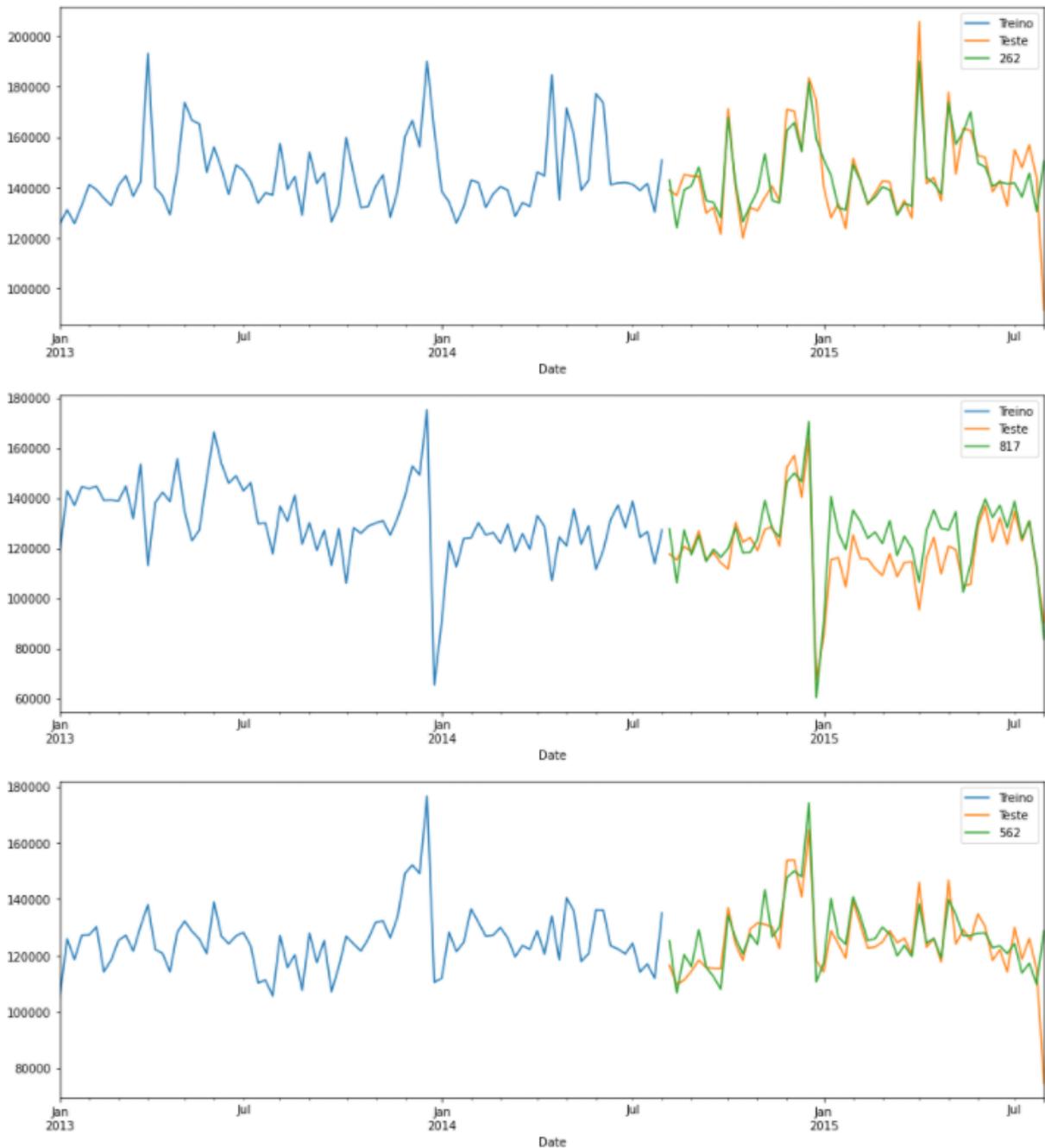
A previsão a nível segregado para a Rossmann foi feita com base em um modelo com os mesmos parâmetros que a linha geral, e apresentou um erro de 9.213. O erro das 3 lojas com maior participação nas vendas pode ser visto na tabela 5 e sua previsão pode ser vista na figura 39.

Tabela 5 - Erro da previsão Rossmann segregada XGBoost.

Loja	Total de vendas	Erro da previsão
262	19.516.842	10.985
817	17.057.867	8.945
562	16.927.322	9.408

Fonte: Autor (2021)

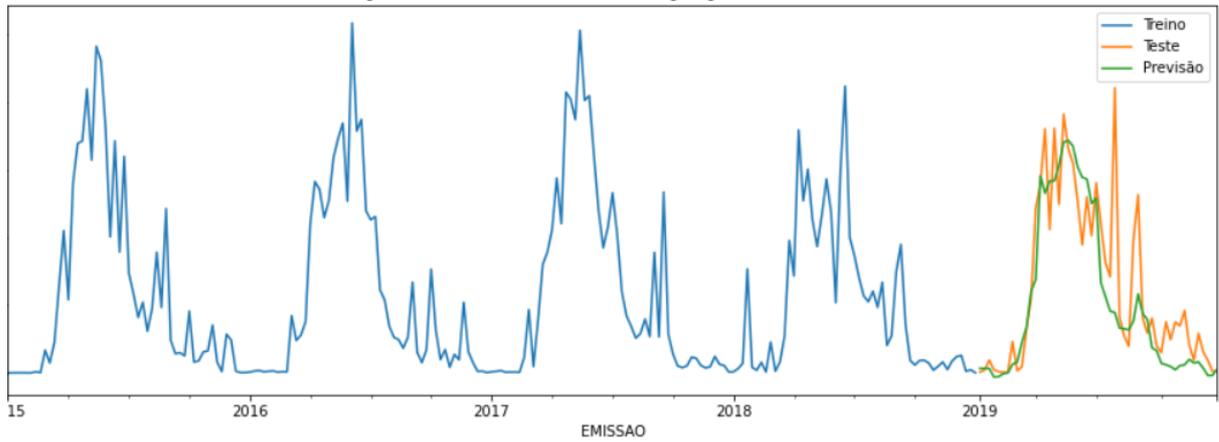
Figura 39 - Previsão Rossmann segregada XGBoost.



Fonte: Autor (2021)

A previsão para a linha geral da empresa têxtil foi feita com base em um modelo com os seguintes parâmetros: “subsample”=0,95; “reg_lambda”=0,45; “reg_alpha”=0,01; “n_estimators”=100; “min_child_weight”=10; “max_depth”=5; “learning_rate”=0,07; “gamma”=0; “colsample_bytree”=0,4; apresentou um erro de 34.104 e pode ser vista na figura 40.

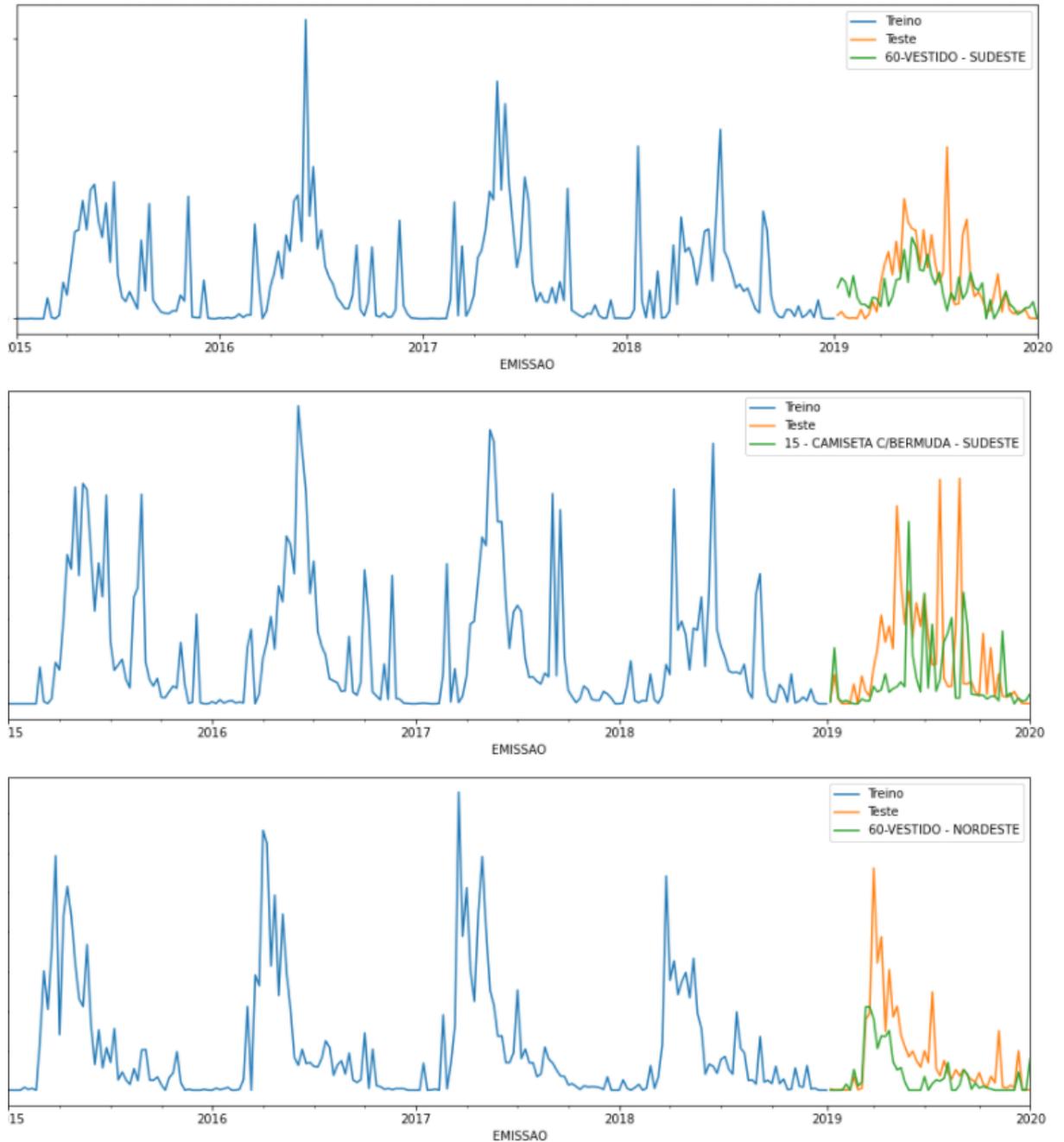
Figura 40 - Previsão têxtil agregada XGBoost.



Fonte: Autor (2021)

A previsão a nível de par grupo-produto da empresa têxtil foi feita com base em um modelo com todos os parâmetros do modelo em seus valores padrões, que apresentou um erro de 635. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 41.

Figura 41 - Previsão têxtil segregada XGBoost.



Fonte: Autor (2021)

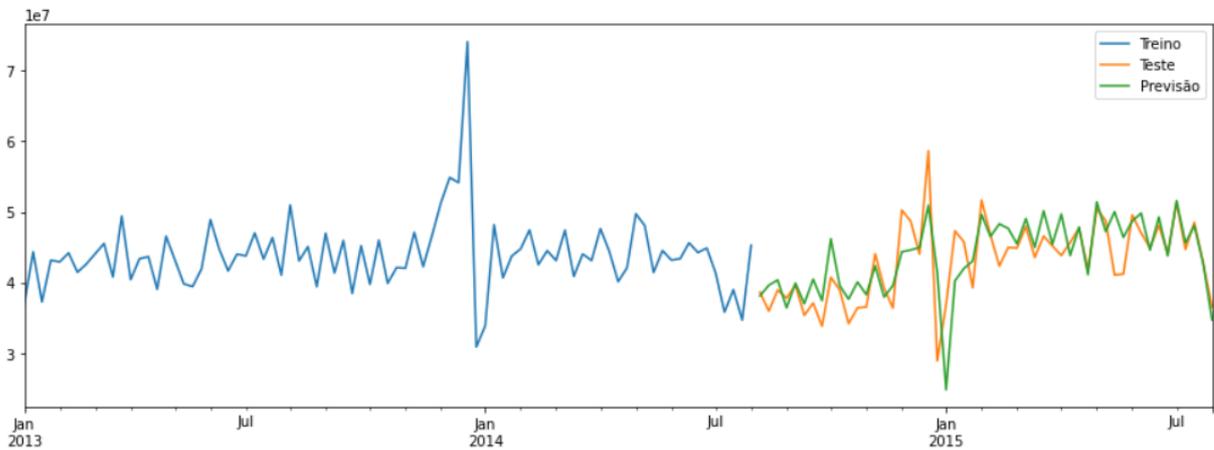
4.4.7 Rede Neural

Para cada cenário de previsão foram analisadas diversas estruturas e princípios de elaboração para as redes neurais utilizadas, incluindo redes neurais artificiais (ANN), redes neurais recorrentes (RNN), redes neurais convolucionais (CNN) e redes neurais de memória de curto prazo longo (LSTM) com diferentes arquiteturas. Para todos os cenários, entretanto, a rede mais eficiente foi uma rede neural convolucional (CNN) com apenas uma convolução e

kernel size=2, diferenciando entre as aplicações apenas no número de filtros da rede, que foi aplicada através da biblioteca Keras, desenvolvido por Gulli e Pal (2017).

A previsão da linha geral da Rossmann por redes neurais foi executada por meio de um CNN com 32 filtros que resultou em um RMSE de 4.016.259 e é vista na figura 42.

Figura 42 - Previsão Rossmann agregada Rede Neural.



Fonte: Autor (2021)

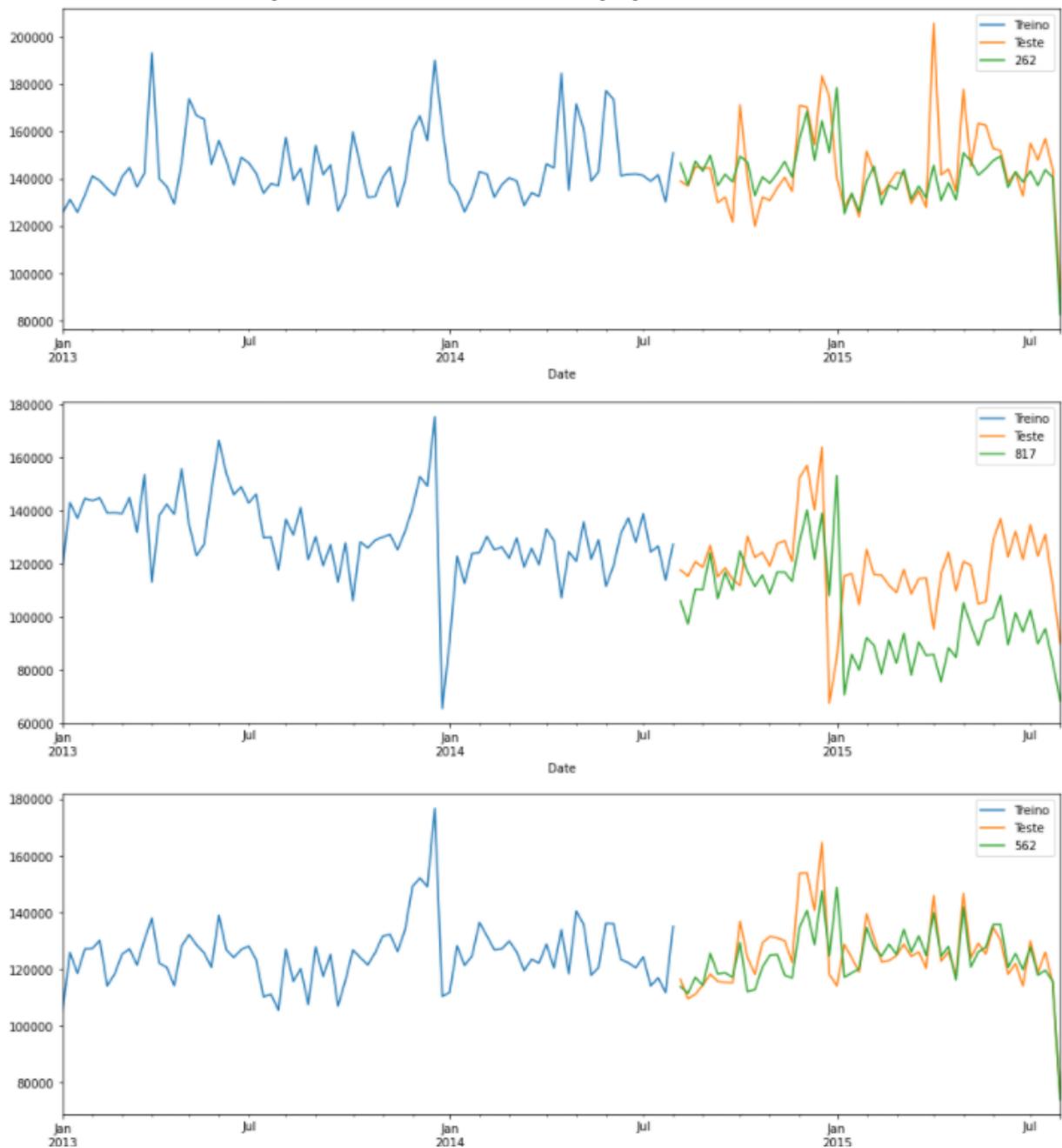
A previsão a nível segregado para a Rossmann foi feita com um CNN com 64 filtros que apresentou um erro de 10.682. O erro das 3 lojas com maior participação nas vendas pode ser visto na tabela 6 e sua previsão pode ser vista na figura 43.

Tabela 6 - Erro da previsão Rossmann segregada Rede Neural.

Loja	Total de vendas	Erro da previsão
262	19.516.842	14.020
817	17.057.867	8.132
562	16.927.322	25.776

Fonte: Autor (2021)

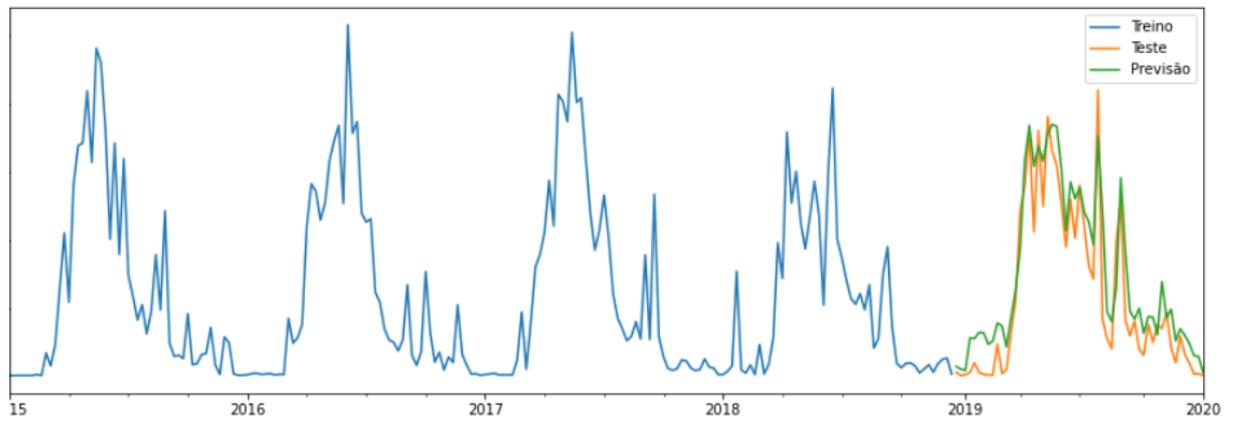
Figura 43 - Previsão Rossmann segregada Rede Neural.



Fonte: Autor (2021)

A previsão para a linha geral da empresa têxtil foi feita por meio de um CNN com 32 filtros que resultou em um RMSE de 24.307 que pode ser vista na figura 44.

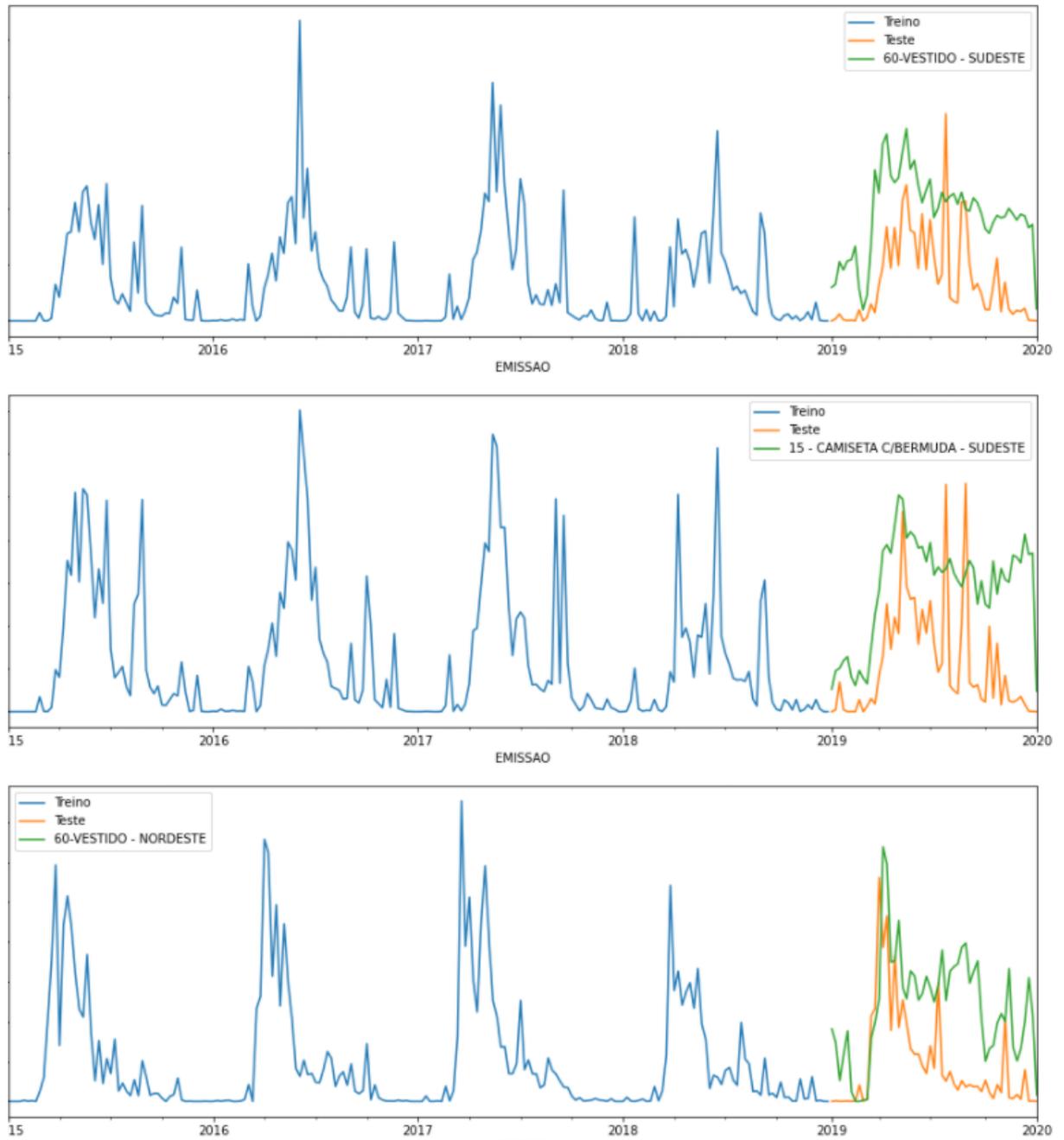
Figura 44 - Previsão têxtil agregada Rede Neural.



Fonte: Autor (2021)

A previsão segregada a nível de par grupo-produto da empresa têxtil foi realizada por meio de um CNN com 32 filtros que resultou em um RMSE de 4.056. A previsão dos 3 pares com maior participação nas vendas pode ser vista na figura 45.

Figura 45 - Previsão têxtil segregada Rede Neural.



Fonte: Autor (2021)

5. RESULTADOS

Com todas as previsões realizadas e seus erros conhecidos é possível concluir as análises inicialmente propostas. Para isso, o quadro 2 resume todos os valores de RMSE calculados para todas as previsões.

Quadro 2 - RMSE de cada modelo.

	ARIMA	Prophet	Holt-Winters	Decision Tree	Random Forest	XGBoost	Rede Neural
Rossmann (Agregado)	5.822.762	5.848.904	-	5.877.981	5.262.006	5.454.413	4.016.259
Têxtil (Agregado)	33.836	33.233	34.437	41.076	32.966	34.104	24.307
Rossmann (Segregado)	10.076	11.075	-	10.892	9.562	9.213	10.682
Têxtil (Segregado)	2.665	2.880	2.615	776	542	635	4065

Fonte: Autor (2021)

Como o RMSE tem seu valor proporcional à magnitude dos dados, não é possível comparar estes valores entre previsões diferentes, o que significa que todas as análises do presente trabalho devem ser realizadas separadamente para cada cenário de previsão. Na perspectiva da tabela apresentada, só faz sentido comparar os valores em uma mesma linha, pois esses erros foram calculados a partir do mesmo cenário de previsão.

Para facilitar a visualização e entendimento dos erros, o quadro 3 mostra a diferença percentual entre o RMSE de cada modelo em comparação com o RMSE do melhor modelo para cada cenário de previsão analisado. Um valor de 50% nesta tabela significa que o RMSE do método foi 50% maior que o melhor RMSE do cenário analisado. Vale ressaltar que mesmo nesta perspectiva a comparação entre cenários diferentes ainda não é válida, pois os percentuais ainda são relativos e proporcionais à magnitude dos dados.

Quadro 3 - Diferenças relativas percentuais entre RMSE.

	ARIMA	Prophet	Holt-Winters	Decision Tree	Random Forest	XGBoost	Rede Neural
Rossmann (Agregado)	45,0%	45,6%	-	46,4%	31,0%	35,8%	0,0%
Têxtil (Agregado)	39,2%	36,7%	41,7%	69,0%	35,6%	40,3%	0,0%
Rossmann (Segregado)	9,4%	20,2%	-	18,2%	3,8%	0,0%	15,9%
Têxtil (Segregado)	391,0%	430,6%	381,8%	43,1%	0,0%	17,0%	647,2%

Fonte: Autor (2021)

Por meio da tabela anterior é fácil perceber o melhor modelo para cada cenário, destacado pela cor verde e pela diferença de 0%, por se tratar do referencial para o cálculo da variação no erro.

O modelo de rede neural foi o melhor para ambas as previsões agregadas, o que vai ao encontro da maioria dos estudos e trabalhos recentes que abordam o tema de previsão de séries temporais, que afirmam que redes neurais tendem a superar a maioria dos métodos tradicionais e de *machine learning* para diversos cenários.

Para ambas as previsões segregadas, que consistem em múltiplas séries temporais paralelas, o melhor modelo foi um modelo de *machine learning*, sendo o modelo Random Forest o melhor para o caso da Rossmann e o modelo XGBoost o melhor para o caso da empresa têxtil. Este resultado também vai ao encontro das expectativas iniciais e da maioria dos estudos sobre o tema, pois estes modelos são capazes de realizar a previsão conjunta das séries e se beneficiando disso.

Um resultado interessante que ainda pode ser analisado por meio desta tabela é quanto ao desempenho das redes neurais para as previsões de séries paralelas, que foi muito abaixo do seu desempenho para as previsões agregadas. Enquanto as redes neurais foram superiores por uma margem considerável nas previsões agregadas, elas foram inferiores também por uma margem considerável nas previsões segregadas, em especial na previsão para a empresa têxtil, em que a rede neural apresentou o pior erro, excedendo o erro do método Random Forest em aproximadamente 6,5 vezes seu valor.

A primeira hipótese para explicar este desempenho baixo é de que a estrutura de rede neural utilizada não foi capaz de assimilar as diferentes séries temporais previstas, e realizou as previsões independentemente, ou generalizando tendências que não eram reais.

Uma segunda hipótese para este desempenho está ligada a variação e quantidade dos dados fornecidos. Enquanto que para as séries agregadas existiam poucas descontinuidades nos dados, isto era muito mais comum para as séries segregadas, em especial para a empresa têxtil, o que explicaria um erro tão grande neste cenário.

O quadro 4 facilita a percepção do desempenho relativo entre os métodos para cada cenário de previsão ao retratar o ranqueamento entre esses métodos, em que a posição 1 é ocupada pelo melhor método, que apresenta o menor RMSE, e a posição 7 é ocupada pelo pior método.

Quadro 4 - Ranking de modelos para cada cenário.

	1	2	3	4	5	6	7
Rossmann (Agregado)	Rede Neural	Random Forest	XGBoost	ARIMA	Prophet	Decision Tree	-
Têxtil (Agregado)	Rede Neural	Random Forest	Prophet	ARIMA	XGBoost	Holt-Winters	Decision Tree
Rossmann (Segregado)	XGBoost	Random Forest	ARIMA	Rede Neural	Decision Tree	Prophet	-
Têxtil (Segregado)	Random Forest	XGBoost	Decision Tree	Holt-Winters	ARIMA	Prophet	Rede Neural

Fonte: Autor (2021)

Por meio da análise desta tabela, em conjunto com as duas apresentadas anteriormente, é possível chegar a novas conclusões interessantes que respondem os objetivos do trabalho.

O modelo Decision Tree foi o pior modelo dentre os 3 de *machine learning* para todos os casos, o que não é uma surpresa, visto que este é o modelo mais básico dos 3. O modelo Random Forest, por outro lado, apresentou resultados satisfatórios e constantes, apresentando o segundo menor erro para 3 dos cenários e o melhor erro para o último, fazendo deste método o mais confiável dentro do cenário analisado.

Os 3 modelos tradicionais apresentaram erros semelhantes entre si, e cada um foi superior aos demais em pelo menos um dos cenários analisados, o que indica que estes métodos estão no mesmo nível no geral, e o método mais indicado depende da situação analisada. O Prophet, por exemplo, é mais indicado para percepção de séries únicas com sazonalidade forte, como a série agregada da empresa têxtil, e o ARIMA tende a apresentar resultados mais estáveis para diferentes cenários.

Os modelos tradicionais, quando analisados como um grupo, não tiveram o melhor desempenho em nenhum dos casos analisados, chegando a apresentar no máximo o 3º menor

erro. Em todos os cenários o modelo Random Forest foi melhor que todos os tradicionais e o modelo XGBoost foi pior apenas que o Prophet e o ARIMA para a previsão agregada da empresa têxtil, que constituiu o cenário ideal para os modelos tradicionais. por ser uma série temporal com forte sazonalidade e variações aleatórias relativamente pequenas.

Por fim, analisando o desempenho dos métodos em relação a previsão de séries multivariadas é possível perceber que os modelos de *machine learning* apresentaram um resultado superior aos métodos tradicionais, em especial os modelos XGBoost e Random Forest, e que a estrutura de CNN com uma convolução apresentou resultados piores que os modelos estatísticos tradicionais.

6. CONCLUSÕES

O trabalho ocorreu de acordo com as etapas descritas nos procedimentos metodológicos, que podem ser observadas na figura 7, o que garantiu que todos os objetivos do trabalho fossem atingidos. O primeiro objetivo específico, que foi definir o experimento de comparação em relação à periodicidade, agregação e métricas de erro foi garantido na etapa de definição da base para comparação, descrita no tópico 4.3. A base de comparação definida especificou uma periodicidade de previsão semanal, para um horizonte de um ano, especificou dois níveis de agregação distintos para cada conjunto de dados e definiu a métrica base para comparação como o RMSE.

O segundo objetivo específico foi revisar a literatura acerca de métodos estatísticos tradicionais e de métodos de inteligência artificial, selecionar e implementar métodos, que foi alcançada por meio das etapas de seleção do modelo e a aplicação de cada modelo e armazenamento dos resultados, descritas nos tópicos 4.1 4.4, respectivamente.

A seleção dos modelos de previsão foi realizada através do estudo de trabalhos e pesquisas divulgados, de maneira que foram selecionados métodos de diferentes princípios, contando com métodos novos e antigos. A aplicação dos modelos foi feita por meio da linguagem de programação python, por meio da qual os parâmetros de cada modelo foram otimizados dentro de um intervalo de valores possíveis, e estão descritos no tópico referente do desenvolvimento.

O terceiro e último objetivo específico foi interpretar resultados obtidos evidenciando a situação de previsão de múltiplos itens possivelmente correlacionados, realizado na etapa de análise dos resultados e descrito no tópico 5 do trabalho. De maneira geral, o resultado obtido foi de que os métodos que obtiveram os melhores resultados são os mais novos e complexos, que são as redes neurais, XGBoost e Random Forest. Analisando especialmente a previsão de múltiplas séries paralelas possivelmente correlacionadas foi observado que a rede neural passou a apresentar resultados ruins, enquanto XGBoost e Random Forest se mantiveram como os melhores métodos.

Por meio do cumprimento de todos os objetivos específicos foi garantido o alcance do objetivo geral do trabalho de comparar o desempenho de métodos tradicionais estatísticos com métodos de inteligência artificial, especificamente em um contexto de previsão de

múltiplos itens possivelmente correlacionados. Como resultado do trabalho é possível afirmar que para os dados utilizados os métodos de inteligência artificial foram superiores aos métodos tradicionais estatísticos, com a exceção da rede neural na previsão de múltiplos itens.

Para trabalhos futuros, sugere-se estudos semelhantes em problemas distintos, com novos conjuntos de dados com características diferentes e possivelmente com outros modelos de previsão. Sugere-se também estudos apenas com os métodos que apresentaram melhores resultados, de uma maneira que seja possível restringir o escopo a fim de aumentar o foco do trabalho na otimização destes métodos, com uma tunagem mais fina e explorando mais possibilidades de estruturação de redes neurais.

REFERÊNCIAS

- ABEPRO. ABEPRO - Associação Brasileira de Engenharia de Produção | **A Profissão**. Disponível em: < <http://portal.abepro.org.br/a-profissao/>>. Acesso em: 20 de maio de 2021.
- AYOTTE, B. et al. **Group Leakage Overestimates Performance: A Case Study in Keystroke Dynamics**. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), p.1410-1417, 2021.
- 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.
- BAYUS, B. L.; PUTSIS, W. P. J. **Product Proliferation: An Empirical Analysis of Product Line Determinants and Market Outcomes**. Marketing Science, v. 18, n. 2, p. 137-153, 1999.
- BERGSTRA, J. et al. **Algorithms for Hyper-Parameter Optimization**. Advances in Neural Information Processing Systems, n. 24, 2011.
- BORGES, L. E. **Python para desenvolvedores**. São Paulo: Novatec, 2014.
- BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. São Francisco: Holden-Day, 1970.
- BREIMAN, L. **Random Forests**. Machine Learning, n. 45, p. 5-32, 2001.
- CHAI, T.; DRAXLER, R. R. **Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature**. Geosci. Model Dev., n. 7, p. 1247-1250, 2014.
- CHATFIELD, C. **The Holt-Winters Forecasting Procedure**. Journal of the Royal Statistical Society. Series C (Applied Statistics), v. 27, n. 3, p. 264-279, 1978.
- CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 785-794, 2016.
- DAWOOD, E. G. **Geo-locating UEs Using Multi-output Decision Tree Regressor**. Florida Institute of Technology, 2019.
- DOBRA, A. **Classification and Regression Tree Construction**. Cornell University, 2002.
- FALK, M. et al. **A First Course on Time Series Analysis-Examples with SAS**. University

of Wurzburg, 2011.

FRIEDMAN, J. H. **Stochastic gradient boosting**. Computational Statistics & Data Analysis, n. 38, p. 367-378, 2002

GRAVES, A. **Generating Sequences With Recurrent Neural Networks**. University of Toronto, 2014.

GUERRINI, F.M.; BELHOT, R.V.; AZZOLINI JR., W. **Planejamento e controle da produção: Projeto e operação de sistemas**. Rio de Janeiro: Elsevier, 2014.

GULLI, A.; PAL, S. **Deep learning with Keras**. Packt Publishing Ltd, 2017.

HYNDMAN, R. J.; KHANDAKAR, Y. **Automatic Time Series Forecasting: The forecast Package for R**. Journal of Statistical Software, 2008.

KALEKAR, P. S. **Time series Forecasting using Holt-Winters Exponential Smoothing**. Kanwal Rekhi School of Information Technology, 2004.

KAUFMAN, S et al. **Leakage in data mining: Formulation, detection, and avoidance**. ACM Transactions on Knowledge Discovery from Data (TKDD), n. 6, p.1–21, 2012

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. **Statistical and Machine Learning forecasting methods: Concerns and ways forward**. PLOS ONE, México, 2018.

MAKRIDAKIS, S. **The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms**. Futures, v. 90, p. 46-60, 2017.

MAURYA, P. K. **Crop Value Forecasting using Decision Tree Regressor and Model s**. European Journal of Molecular & Clinical Medicine, v. 7, 2020.

PAPACHARALAMPOUS, G. A.; TYRALIS, H. **Evaluation of random forests and Prophet for daily streamflow forecasting**. Advances in Geosciences, n. 45, p.201-208, 2018.

PEDREGOSA, et al. **Machine Learning in Python**. Scikit-learn, p. 2825-2830, 2011

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.

PODOBNIK, B; STANLEY, H. E. **Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series**. Physical Review Letters, v. 100, 2008.

SAMAL K. K. R. et al. **Time Series based Air Pollution Forecasting using SARIMA and Prophet Model**. International Conference on Information Technology and Computer Communications, p-80-85, 2019.

SATRIO, C. B. A. et al. **Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET**. 5th International Conference on Computer Science and Computational Intelligence, 2020.

SEABOLD, S.; PERKTOLD, J. **statsmodels: Econometric and statistical modeling with python**. 9th Python in Science Conference, 2010.

SILVEIRA, I. A. **Modelo de previsão de demanda com o uso de aprendizado supervisionado de máquina: um estudo de caso em uma empresa de varejo**. Florianópolis, 2019.

STEVENSON, M.; HENDRY, L. C.; KINGSMAN, B. G. **A review of production planning and control: the applicability of key concepts to the make-to-order industry**. Lancaster, 2004.

TAYLOR, S. J.; LETHAM, B. **Forecasting at Scale**. American Statistician, 2018.

TUBINO, D. F. **Planejamento e controle da produção teoria e prática**. São Paulo: Atlas, 2007.

TULJAPURKAR, S.; HARIDAS, C. V. **Temporal autocorrelation and stochastic population growth**. Ecology Letters, n. 9, p. 327-337, 2006.

WANG, W.; LU, Y. **Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model**. IOP Conf. Series: Materials Science and Engineering, n. 324, 2018

WEN, R.; TORKKOLA, K.; NARAYANASWAMY, B.; MADEKA, D. **A Multi-Horizon Quantile Recurrent Forecaster**. Cornell University, 2017.

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.

YING, X. **An Overview of Overfitting and its Solutions**. Phys.: Conf. Ser. 1168 022022, 2018.

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.

ZHANG, G. P. **Avoiding Pitfalls in Neural Network Research**. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, v. 37, n. 1, 2007.

ZHANG, G. P. **Time series forecasting using a hybrid ARIMA and neural network model.** Neurocomputing, n.50, p. 159-175, 2003.