



Pós-Graduação em Ciência da Computação

Antonio Jorge Ferreira Delgado Filho

Análise de Métodos de Regressão para Previsão de Demanda de Curto Prazo



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
<http://cin.ufpe.br/~posgraduacao>

Recife
2020

Antonio Jorge Ferreira Delgado Filho

Análise de Métodos de Regressão para Previsão de Demanda de Curto Prazo

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Inteligência computacional

Orientador: Prof. Dr. Germano Crispim Vasconcelos

Recife
2020

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

D352a Delgado Filho, Antonio Jorge Ferreira
Análise de métodos de regressão para previsão de demanda de curto prazo
/ Antonio Jorge Ferreira Delgado Filho. – 2020.
93 f.: il., fig., tab.

Orientador: Germano Crispim Vasconcelos.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn,
Ciência da Computação, Recife, 2020.
Inclui referências.

1. Inteligência computacional. 2. Séries temporais. I. Vasconcelos, Germano
Crispim (orientador). II. Título.

006.31

CDD (23. ed.)

UFPE - CCEN 2020 - 97

Antonio Jorge Ferreira Delgado Filho

“Análise de Métodos de Regressão para Previsão de Demanda de Curto Prazo”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação

Aprovado em: 12/03/2020.

BANCA EXAMINADORA

Profa. Dra. Patricia Cabral de Azevedo Resteli Tedesco
Centro de Informática / UFPE

Prof. Dr. Cleviton Vinicius Fonsêca Monteiro
Departamento de Estatística e Informática/UFRPE

Prof. Dr. Germano Crispim Vasconcelos
Centro de Informática / UFPE
(Orientador)

Dedico este trabalho a Deus, minha família e meus amigos, essenciais para o meu sucesso.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a Deus, pois sem Ele não somos nada, por mais essa oportunidade que estou tendo em minha vida: poder realizar uma Pós-graduação, em uma instituição como a UFPE e num centro como o CIn. Também gostaria de realizar um agradecimento especial para a minha família, por todo o apoio dado nesse novo desafio. Muito obrigado por ser essa minha base onde sei que posso encontrar forças e apoio sempre, para qualquer desafio que eu tiver pela frente. Ao meu pai por todas as conversas, conselhos, força e carinho que sempre me dá. À minha mãe por ser essa pessoa tão especial em minha vida com toda a sua atenção e dedicação, me abastecendo de energia, amor e carinho. À minha irmã por ser uma amiga fora de série, sempre disposta a me ouvir, pelas nossas conversas, por me ajudar em todos os momentos em que precisei, por sempre ter um abraço e uma palavra pra me encher de forças. Aos meus primos, agradeço por toda a força e energia positiva, por toda compreensão nos momentos em que estive ausente para a escrita da dissertação e além de todos os momentos de descontração e alegria. Aos demais familiares, meu muito obrigado por todo suporte dado.

Também não posso deixar de agradecer aos novos e velhos amigos que estiveram comigo no decorrer desses 2 anos. Mais uma vez, como foi na graduação, gostaria de agradecer ao meu grande amigo Edvan Soares, por mais uma vez ter ombreado um novo desafio, sempre disposto a ajudar a transpor qualquer obstáculo. Aos meus amigos que trabalharam comigo no NUTES-UFPE, que sempre estavam prontos para transmitir toda energia positiva e auxílio para a escrita desse trabalho. Aos meus amigos da UFRPE, que também estiveram presentes apoiando e aconselhando. Aos meus amigos da "Quinta Série" por toda força dada em todos os momentos, com uma palavra amiga, de incentivo e pelos momentos de diversão e risadas, que me reabasteciam de energia para a conclusão desse trabalho. Muito obrigado meus amigos, pela amizade de vocês e todo o suporte dado nessa trajetória.

Por fim, aos professores do CIn-UFPE, por toda contribuição para meu processo de maturidade como pesquisador. Agradeço meu orientador, Germano Vasconcelos, pela oportunidade, por todas orientações e ensinamentos passados durante o processo de desenvolvimento desse trabalho.

RESUMO

A Previsão de Demandas é uma das tarefas e ferramentas mais importantes para uma boa gestão e planejamento organizacional de uma empresa, auxiliando gestores no gerenciamento e nas tomadas de decisões que podem ajudar a transpor obstáculos necessários para o sucesso do negócio. Diante disso, este trabalho tem o objetivo de realizar uma investigação ampla, utilizando um conjunto de algoritmos de Aprendizagem de Máquinas para a realização da previsão de demanda de curto prazo. Na investigação serão empregados dados de uma grande empresa de distribuição de bebidas no mercado brasileiro. Para atingir esse objetivo serão utilizados Métodos Supervisionados, mais especificamente algoritmos de regressão, para realizar a predição da produção semanal dessa empresa, através da análise de Séries Temporais. Dentre esses algoritmos estão: Linear Regression (LR), Support Vector Regression (SVR), Stochastic Gradient Descent (SGD), Decision Tree, Multilayer Perceptron (MLP), Random Forest (RF) e XGBoost. Durante a investigação experimental foram testadas as técnicas de regressão listadas anteriormente, com um total de 24 configurações analisadas para cada um dos 5 produtos da empresa, sendo essas ajustadas com o auxílio da técnica de Random Search e tratamentos dos dados das séries temporais, como a Normalização e Dessazonalização. Todas essas análises serão realizadas com o apoio da linguagem de programação *Python* e suas bibliotecas, como o *Scikit-learn*. Além da realização dessa predição, o trabalho também propõe a comparação dos resultados alcançados pelos modelos de regressão desenvolvidos nessa pesquisa, com modelos desenvolvidos anteriormente por outras empresa, onde uma delas utilizou o método ARIMA. Juntamente a análise anterior, a utilização de variáveis externas temporais, como dados econômicos e climáticas, nos dados de entrada do modelo a fim de analisar a influência que dados externos podem trazer às previsões. Os resultados alcançados e avaliados com o auxílio das métricas MAPE, MAE, RMSE e RMSLE demonstram que os métodos de regressão são um opção muito eficiente para a realização da previsão de séries temporais, alcançados resultados superiores na maioria dos casos, quando comparados aos já existentes.

Palavras-chaves: Aprendizagem de Máquinas. Métodos de Regressão. Séries Temporais. ARIMA. Previsão de Demanda.

ABSTRACT

Demand Forecasting is one of the most important tasks and tools for the good management and organizational planning of a company, assisting managers in the managing and making decisions that can help overcome obstacles to business success. Therefore, this work has the objective of carrying out a broad investigation, using a set of Machine Learning algorithms for the realization of short-term demand forecasting. The investigation will use data from a large beverage distribution company in the Brazilian market. To achieve this objective, Supervised Methods, more specifically regression algorithms, will be used to predict the weekly production of this company, through the analysis of Time Series. Among these algorithms are: Linear Regression (LR), Support Vector Regression (SVR), Stochastic Gradient Descent (SGD), Decision Tree, Multilayer Perceptron (MLP), Random Forest (RF) and XGBoost. During the experimental investigation, the regression techniques listed above were tested, with a total of 24 configurations analyzed for each of the company's 5 products, which were adjusted with the aid of the Random Search technique and data series treatments, such as Standardization and removing Seasonality. All of these analyzes will be realized with the support of the Python programming language and its libraries, such as Scikit-learn. In addition to making this prediction, the work also proposes comparing the results achieved by the regression models developed in this research, with models previously developed by other companies, where one of them used the ARIMA method. Together with the previous analysis, the use of external temporal variables, such as economic and climatic data, in the input data of the model in order to analyze the influence that external data can bring to the forecasts. The results achieved and evaluated with the aid of the metrics MAPE, MAE, RMSE and RMSLE demonstrate that the regression methods are a very efficient option for the realization of the time series forecast, achieving superior results in most cases, when compared to the existing ones .

Keywords: Machine Learning. Regression Methods. Time Series. ARIMA. Demand Forecasting.

LISTA DE FIGURAS

Figura 1 – Exemplos de métodos de previsão de demanda	19
Figura 2 – Processo de Previsão, adaptado de Kress e Snyder (1994)	20
Figura 3 – Processo de Previsão, adaptado de Scott (2001)	20
Figura 4 – Processo de Previsão, adaptado de Werner e Ribeiro (2006)	20
Figura 5 – Exemplos de Séries Temporais	21
Figura 6 – Componentes de Séries Temporais	22
Figura 7 – Exemplos estacionariedade de Series Temporais	23
Figura 8 – Exemplo modelo regressão linear ajustado	26
Figura 9 – Exemplo arquitetura Rede Neural	27
Figura 10 – Exemplos de funções de ativação. (a) Função Sigmóide Logística (b) Tangente Hiperbólica	28
Figura 11 – Exemplos de aplicação função não linear (kernel) e hiperplano	29
Figura 12 – Exemplos de aplicação função não linear (kernel) e hiperplano	29
Figura 13 – Gráfico com aplicação do Gradient Descent (GD)	30
Figura 14 – Estrutura de uma Árvore de Decisão	31
Figura 15 – Estrutura Random Forest	33
Figura 16 – Fases do CRISP-DM	44
Figura 17 – Amostra do Dataset	45
Figura 18 – Decomposição da Série Temporal - Produto 1	46
Figura 19 – Decomposição da Série Temporal - Produto 2	46
Figura 20 – Decomposição da Série Temporal - Produto 3	46
Figura 21 – Decomposição da Série Temporal - Produto 4	46
Figura 22 – Decomposição da Série Temporal - Produto 5	47
Figura 23 – Boxplot - Produto 1	48
Figura 24 – Boxplot - Produto 2	48
Figura 25 – Boxplot - Produto 3	48
Figura 26 – Boxplot - Produto 4	48
Figura 27 – Boxplot - Produto 5	49
Figura 28 – Exemplo da transformação do campo "Semana" em "Data"	50
Figura 29 – Exemplo da transformação One Hot Encode	51
Figura 30 – Previsões - Produto 1	60
Figura 31 – Previsões - Produto 2	60
Figura 32 – Previsões - Produto 3	60
Figura 33 – Previsões - Produto 4	60
Figura 34 – Previsões - Produto 5	60
Figura 35 – Previsão Modelo 4 Inicial (XGBoost) - Produto 1	67

Figura 36 – Previsão Modelo 8 Inicial (Decision Tree) - Produto 1	67
Figura 37 – Previsão Modelo 7 Normalização (MLP) - Produto 1	68
Figura 38 – Previsão Modelo 1 Normalização (MLP) - Produto 1	68
Figura 39 – Previsão Modelo 7 Normalização (SGD) - Produto 1	68
Figura 40 – Previsão Modelo 2 Inicial (MLP) - Produto 1	68
Figura 41 – Previsão Modelo 5 Normalização (XGBoost) - Produto 1	68
Figura 42 – Previsão Modelo 1 Inicial (XGBoost) - Produto 1	68
Figura 43 – Previsão Modelo 5 Inicial (XGBoost) - Produto 1	69
Figura 44 – Previsão Modelo 2 Inicial (XGBoost) - Produto 1	69
Figura 45 – Previsão Modelo 5 Dessazonalização (Decision Tree) - Produto 2	70
Figura 46 – Previsão Modelo 8 Normalização (Random Forest) - Produto 2	70
Figura 47 – Previsão Modelo 5 Inicial (MLP) - Produto 2	71
Figura 48 – Previsão Modelo 2 Normalização (SGD) - Produto 2	71
Figura 49 – Previsão Modelo 4 Normalização (XGBoost) - Produto 2	71
Figura 50 – Previsão Modelo 8 Normalização (MLP) - Produto 2	71
Figura 51 – Previsão Modelo 8 Normalização (XGBoost) - Produto 2	71
Figura 52 – Previsão Modelo 8 Inicial (Random Forest) - Produto 2	71
Figura 53 – Previsão Modelo 8 Normalização (SVR) - Produto 2	72
Figura 54 – Previsão Modelo 8 Inicial (Decision Tree) - Produto 2	72
Figura 55 – Previsão Modelo 1 Normalização (MLP) - Produto 3	73
Figura 56 – Previsão Modelo 6 Normalização (SGD) - Produto 3	73
Figura 57 – Previsão Modelo 8 Inicial (MLP) - Produto 3	74
Figura 58 – Previsão Modelo 2 Inicial (MLP) - Produto 3	74
Figura 59 – Previsão Modelo 6 Inicial (MLP) - Produto 3	74
Figura 60 – Previsão Modelo 5 Inicial (MLP) - Produto 3	74
Figura 61 – Previsão Modelo 3 Inicial (MLP) - Produto 3	74
Figura 62 – Previsão Modelo 3 Normalização (SGD) - Produto 3	74
Figura 63 – Previsão Modelo 8 Inicial (SGD) - Produto 3	75
Figura 64 – Previsão Modelo 4 Normalização (SGD) - Produto 3	75
Figura 65 – Previsão Modelo 7 Inicial (SGD) - Produto 4	76
Figura 66 – Previsão Modelo 5 Inicial (SGD) - Produto 4	76
Figura 67 – Previsão Modelo 7 Inicial (MLP) - Produto 4	77
Figura 68 – Previsão Modelo 8 Normalização (SGD) - Produto 4	77
Figura 69 – Previsão Modelo 7 Inicial (SVR) - Produto 4	77
Figura 70 – Previsão Modelo 5 Inicial (SVR) - Produto 4	77
Figura 71 – Previsão Modelo 6 Inicial (SGD) - Produto 4	77
Figura 72 – Previsão Modelo 1 Inicial (MLP) - Produto 4	77
Figura 73 – Previsão Modelo 3 Inicial (MLP) - Produto 4	78
Figura 74 – Previsão Modelo 2 Inicial (SVR) - Produto 4	78

Figura 75 – Previsão Modelo 5 Inicial (SVR) - Produto 5	79
Figura 76 – Previsão Modelo 8 Dessazonalização (Decision Tree) - Produto 5	79
Figura 77 – Previsão Modelo 5 Normalização (Decision Tree) - Produto 5	80
Figura 78 – Previsão Modelo 5 Inicial (Decision Tree) - Produto 5	80
Figura 79 – Previsão Modelo 7 Normalização (SVR) - Produto 5	80
Figura 80 – Previsão Modelo 3 Inicial (SGD) - Produto 5	80
Figura 81 – Previsão Modelo 7 Dessazonalização (SGD) - Produto 5	80
Figura 82 – Previsão Modelo 6 Inicial (SVR) - Produto 5	80
Figura 83 – Previsão Modelo 7 Dessazonalização (Decision Tree) - Produto 5	81
Figura 84 – Previsão Modelo 3 Normalização (SVR) - Produto 5	81

LISTA DE TABELAS

Tabela 1 – Atributos do Dataset	45
Tabela 2 – Resultados Augmented Dickey-Fuller (ADF) test	47
Tabela 3 – Parâmetros ajustados - Linear Regression	53
Tabela 4 – Parâmetros ajustados - SVR	54
Tabela 5 – Parâmetros ajustados - SGD	54
Tabela 6 – Parâmetros ajustados - Decision Tree	54
Tabela 7 – Parâmetros ajustados - MLP	55
Tabela 8 – Parâmetros ajustados - Random Forest	55
Tabela 9 – Parâmetros ajustados - XGBoost	56
Tabela 10 – Modelos e atributos	57
Tabela 11 – Resultados Empresa A	59
Tabela 12 – Resultados Empresa B	59
Tabela 13 – Resultados Produto 1 - Abordagem Inicial	61
Tabela 14 – Resultados Produto 2 - Abordagem Inicial	61
Tabela 15 – Resultados Produto 3 - Abordagem Inicial	62
Tabela 16 – Resultados Produto 4 - Abordagem Inicial	62
Tabela 17 – Resultados Produto 5 - Abordagem Inicial	63
Tabela 18 – Resultados Produto 1 - Transformação dos Dados	64
Tabela 19 – Resultados Produto 2 - Transformação dos Dados	64
Tabela 20 – Resultados Produto 3 - Transformação dos Dados	65
Tabela 21 – Resultados Produto 4 - Transformação dos Dados	65
Tabela 22 – Resultados Produto 5 - Transformação dos Dados	66
Tabela 23 – Melhores Resultados Produto 1	67
Tabela 24 – Melhores Resultados Produto 2	70
Tabela 25 – Melhores Resultados Produto 3	73
Tabela 26 – Melhores Resultados Produto 4	76
Tabela 27 – Melhores Resultados Produto 5	79
Tabela 28 – Resultados Comparação Produto 1	82
Tabela 29 – Resultados Comparação Produto 2	82
Tabela 30 – Resultados Comparação Produto 3	83
Tabela 31 – Resultados Comparação Produto 4	83
Tabela 32 – Resultados Comparação Produto 5	83

LISTA DE ABREVIATURAS E SIGLAS

ADF	Augmented Dickey-Fuller
ANN	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
CNN-LSTM	Convolutional Neural Network - Long short-term memory
CRISP-DM	CRoss Industry Standard for Data Mining
GD	Gradient Descent
INMET	Instituto Nacional de Meteorologia
IPCA	Índice Nacional de Preços ao Consumidor Amplo
IPEA	Instituto de Pesquisa Econômica Aplicada
LR	Linear Regression
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Spline
MLP	Multilayer Perceptron
PSO	Particle Swarm Optimization
RF	Random Forest
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Logarithmic Error
SEBRAE	Serviço Brasileiro de Apoio às Micro e Pequenas Empresas
SGD	Stochastic Gradient Descent
SVM	Support Vector Machines
SVR	Support Vector Regression

SUMÁRIO

1	INTRODUÇÃO	15
1.1	APRESENTAÇÃO E MOTIVAÇÃO	15
1.2	OBJETIVOS	16
1.3	ESTRUTURA DA DISSERTAÇÃO	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	PREVISÃO DE DEMANDA E A GESTÃO DAS EMPRESAS	18
2.2	SÉRIES TEMPORAIS	21
2.3	APRENDIZAGEM DE MÁQUINA	23
2.3.1	Métodos de Regressão	25
2.3.1.1	Linear Regression	25
2.3.1.2	Multilayer Perceptron	26
2.3.1.3	Support Vector Regression	28
2.3.1.4	Stochastic Gradient Descent	30
2.3.1.5	Decision Tree	31
2.3.1.6	Random Forest	32
2.3.1.7	XGBoost Regressor	33
3	TRABALHOS RELACIONADOS	35
3.1	METODOLOGIA	35
3.2	ALGORITMOS ESTUDADOS	35
3.2.1	ARIMA	35
3.2.2	Linear Regression	36
3.2.3	Multilayer Perceptron	37
3.2.4	SVR	38
3.2.5	SGD	39
3.2.6	Decision Tree	39
3.2.7	Random Forest	40
3.2.8	XGBoost	41
3.3	CONCLUSÃO	42
4	METODOLOGIA DE EXPERIMENTAÇÃO	43
4.1	ENTENDIMENTO DO NEGÓCIO E ENTENDIMENTO DOS DADOS	44
4.2	PREPARAÇÃO DOS DADOS	49
4.3	MODELAGEM DOS DADOS	52
4.4	AValiação	57

5	EXPERIMENTOS E ANÁLISES DE RESULTADOS	59
5.1	ABORDAGEM INICIAL	61
5.2	ABORDAGEM TRANSFORMAÇÃO DOS DADOS	63
5.3	ADIÇÃO DAS VARIÁVEIS EXTERNAS E RESULTADOS FINAIS	66
5.4	CONCLUSÕES	81
6	CONCLUSÃO	84
6.1	CONSIDERAÇÕES FINAIS	84
6.2	TRABALHOS FUTUROS	86
	REFERÊNCIAS	87

1 INTRODUÇÃO

Este Capítulo tem como objetivo apresentar uma visão geral sobre o trabalho desenvolvido nesta pesquisa, salientando a motivação para o seu desenvolvimento. Além disso serão apresentados os objetivos e como a pesquisa está estruturada.

1.1 APRESENTAÇÃO E MOTIVAÇÃO

Na atualidade as pequenas e médias empresas no Brasil têm enfrentado dificuldades para sobreviverem por um longo período. Segundo dados do Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE) (SEBRAE, 2016), cerca de 24% das empresas não se mantêm ativas por mais de dois anos de atividade. Nessa mesma análise foi apurado que um dos principais fatores responsável por esse fato é a falta de planejamento do negócio. Analisando o mesmo contexto em empresas internacionais, de acordo com os números disponibilizados pela *Bureau of Labor Statistics*, que é uma unidade do Departamento do Trabalho dos Estados Unidos, 20% das pequenas empresas falham no seu primeiro ano. Além disso, destacam que 23% falham por não ter uma equipe especializada administrando o negócio, ou seja a falta de planejamento de negócio contribui para o seu fracasso (FUNDERA, 2020). Dentre as tarefas importantes para um bom planejamento organizacional encontra-se a previsão de demanda, que se trata de uma estimativa do que vai acontecer, em um determinado espaço de tempo, com as vendas de produtos ou serviços de uma determinada empresa. Essa previsão não necessariamente é exata, fornecendo uma ideia aproximada possível da realidade (SEBRAE, 2017). Para o planejamento estratégico da produção, vendas e finanças de uma empresa, a previsão de demanda é a base, pois com ela os gestores podem antever o futuro, planejando adequadamente suas ações, servindo de ponto de partida para o planejamento de fluxo de caixa, produção e vendas, controle de estoque e etc (MIRANDA et al., 2011). Devido a essas características, a previsão de demanda se torna uma importante ferramenta de apoio à decisão, auxiliando gestores a transpor as barreiras necessárias para o sucesso da empresa.

Para realizar a previsão de demanda existem duas abordagens, a qualitativa e a quantitativa. A análise qualitativa é utilizada quando a quantidade de dados históricos não é suficiente. Nessa abordagem a experiência e intuição do administrador e sua equipe são essenciais. Além disso, podem ser realizadas pesquisas com os consumidores para levantamento de informações. Na análise quantitativa, são utilizados os dados históricos da empresa, que se encontram estruturados em um certo período de tempo sejam mensais, semanais ou anuais. Nesses casos são utilizados métodos estatísticos para realizar as previsões. (SEBRAE, 2017). Nessa última abordagem, os dados históricos analisados são chamados de Séries Temporais. As Séries Temporais são sequências ordenadas de valores

registrados em um determinado intervalo de tempo (diário, semanal, mensal, anual, etc). Diante dessa estrutura nos dados observados é possível a identificação de padrões, que proporcionam a possibilidade de previsões futuras (BROCKWELL; DAVIS, 2016).

Diante desse problema, vê-se a oportunidade da realização de estudos sobre técnicas de previsão de demanda de curto prazo com a utilização de algoritmos preditivos. Seguindo esse objetivo, algumas das abordagens possíveis para a solução do problema é a aplicação de Métodos de Regressão e Modelos de Redes Recorrentes. Neste trabalho será aplicada a primeira abordagem, pelos métodos de regressão serem amplamente difundidos como capazes de estimar valores futuros com base em valores do passado, já tendo alguns deles sido aplicados em outras aplicações de previsão de séries temporais. Esta pesquisa, em particular, se destaca pela amplitude de modelos e configurações investigados, pela tarefas de processamento dos dados aplicadas nas séries temporais e da avaliação do efeito de variáveis externas na previsão. A fim de atingir tal finalidade, será empregada a base de dados de um grande fabricante e distribuidor de bebidas para teste dos modelos, realizando a previsão da quantidade, em litros, que o produto deverá ser produzido. Os dados a serem estudados encontram-se organizados em períodos semanais, entre Janeiro de 2015 e Setembro de 2017, para diferentes produtos, marcas e unidades distribuídas geograficamente. Esses conjuntos de dados foram divididos em dois, um para treinamento e outro para testes, em que o primeiro encontra-se entre Janeiro de 2015 a Julho de 2017 e o último de Agosto de 2017 a Setembro de 2017, sendo as últimas seis semanas do conjunto de dados.

Com o intuito de realizar a predição dos dados serão utilizados os seguintes algoritmos de regressão: Linear Regression (LR), Support Vector Regression (SVR), Stochastic Gradient Descent (SGD), Decision Tree, Multilayer Perceptron (MLP), Random Forest e XGBoost. Todos esses algoritmos encontram-se disponíveis no módulo Scikit-learn, que apresenta um conjunto de algoritmos de aprendizado de máquinas, na linguagem de programação Python (PEDREGOSA et al., 2011). Além da análise com o conjunto de algoritmos apresentados anteriormente, outras oportunidades observadas são a análise e comparações entre modelos que serão desenvolvidos na pesquisa e modelos já existentes que utilizam Autoregressive Integrated Moving Average (ARIMA), com o propósito de investigar e comparar o desempenho dos algoritmos de regressão entre si e usar como *baseline* modelos ARIMA como métodos estatísticos clássicos para previsão de séries temporais. Para a avaliação e comparação dos resultados serão utilizadas as métricas Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) e Root Mean Squared Logarithmic Error (RMSLE).

1.2 OBJETIVOS

Este trabalho tem como objetivo principal desenvolver modelos preditivos utilizando métodos de Aprendizagem de Máquinas Supervisionados, mais especificamente Métodos de

Regressão, para previsão de Séries Temporais, comparando-os com modelos desenvolvidos previamente a essa pesquisa, onde um deles é o ARIMA. Para que esse objetivo seja alcançado, serão executados os seguintes objetivos específicos:

- Investigar de forma ampla, um conjunto de algoritmos de regressão e suas configurações, como solução para o problema apresentado;
- Investigar o impacto de variáveis externas no desempenho dos modelos de regressão;
- Análise comparativa entre modelos de regressão e os modelos já existentes;
- Avaliação experimental com conjunto de dados, que contém produção semanal de uma empresa de bebidas

1.3 ESTRUTURA DA DISSERTAÇÃO

O trabalho desenvolvido nessa pesquisa encontra-se estruturado em 6 capítulos. O Capítulo 2 apresenta o referencial teórico utilizado para a elaboração do trabalho. No Capítulo 3 é apresentado os trabalhos relacionados, que utilizaram abordagens similares a proposta dessa pesquisa, segmentado pelos métodos de regressão que serão utilizados no decorrer dos experimentos. No 4º Capítulo é descrita a investigação experimental desenvolvida na dissertação, apresentando cada etapa seguida. Já no Capítulo 5 são apresentados os experimentos e resultados alcançados pelos modelos de regressão desenvolvidos, realizando comparações com os modelos previamente desenvolvidos por outra equipe. Por fim, o Capítulo 6 apresenta as conclusões do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

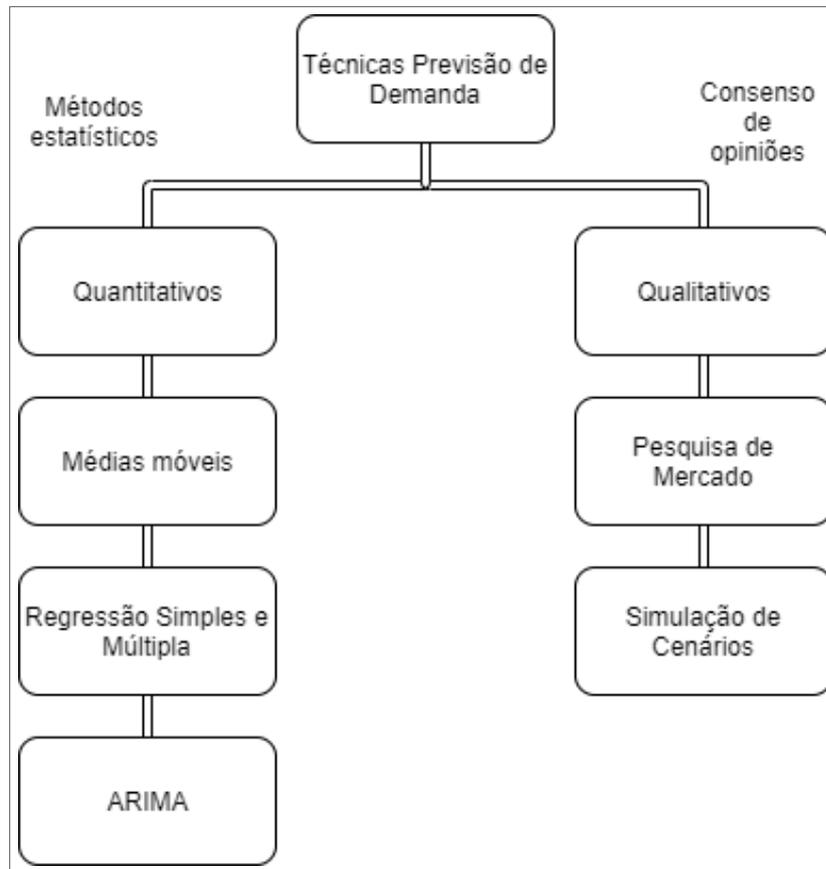
2.1 PREVISÃO DE DEMANDA E A GESTÃO DAS EMPRESAS

Pesquisas têm demonstrado que empresas enfrentam dificuldades para se manter ativas por mais de 2 anos, desde a sua fundação. Tal informação pode ser confirmada nos dados disponibilizados pelo SEBRAE no ano de 2016, onde é informado que no ano de 2012 foi alcançado uma taxa de sobrevivência de 76,6%, a maior alcançada entre os anos de 2008 e 2012, ou seja 23,4% das empresas não sobrevivem por mais de 2 anos de atividade (SEBRAE, 2016).

Na mesma pesquisa é evidenciado dentre os fatores que contribuem para essa taxa de mortalidade a falta de Planejamento do Negócio e Gestão do Negócio. Em trabalhos como o desenvolvido por BARRETO e ANTONOVZ (2016) é reafirmado que erros de gestão e planejamento podem custar a sobrevivência da organização, concluindo que existem uma série de ferramentas gerenciais que podem aprimorar o trabalho de gestão nas empresas, consequentemente aumentando as chances de sobrevivência da empresa. Essas ferramentas são métodos que assistem os gestores no planejamento e desenvolvimento de produtos ou processos. Para a gestão estratégica englobam controle e registro que irão apoiar na tomada de certas decisões (MENDONÇA et al., 2017)(MIRANDA et al., 2016). E como exemplo de uma dessas ferramentas, podemos citar a Previsão de Demanda.

A Previsão de Demanda é um método que serve como ponto de partida para o planejamento de muitas atividades realizadas em uma organização, como por exemplo, o planejamento do fluxo de caixa, planejamento estratégico, planejamento de produção e vendas, controle de estoques etc (MIRANDA et al., 2011). Tal método é definido como um ou mais modelos utilizados pelas empresas com a finalidade de realizar previsões que obtenham resultados mais acurados, sendo definido de acordo com o objetivo da empresa (ROSSETO et al., 2011). Para atingir seu objetivo, existem três formas de elaborar previsões de demanda: utilizando métodos quantitativos, qualitativos ou a combinação de ambos. Os métodos quantitativos são baseados na análise de Séries Temporais, que são conjuntos de dados que denotam a variação de demanda em um certo período de tempo, enquanto que os métodos qualitativos baseiam-se em opiniões e estudos realizados por especialistas (PELLEGRINI; FOGLIATTO, 2001).

Figura 1 – Exemplos de métodos de previsão de demanda



Fonte: (LUSTOSA; MESQUITA; OLIVEIRA, 2008)

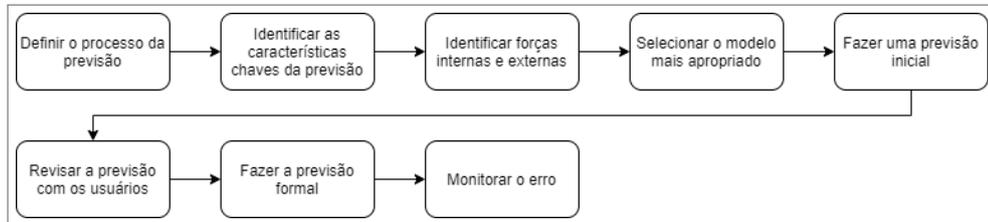
Na figura 1 podemos observar exemplos de métodos qualitativos e quantitativos. Uma das principais características dos métodos qualitativos são o seu grande grau de subjetividade, parecendo menos adequados que os quantitativos. Porém, quando não há dados disponíveis são a única alternativa. Já os quantitativos possuem uma pluralidade de métodos estatísticos, com diferentes características e complexidade (LUSTOSA; MESQUITA; OLIVEIRA, 2008).

Em seu trabalho Junior et al. (2010) ressalta que a Tecnologia da Informação é um recurso estratégico que torna capaz a aplicação de tais métodos, com o auxílio de softwares destinados a esse propósito. Ainda sobre a importância da utilização da previsão de demanda, Popadiuk e Santos (2010) evidencia que os riscos do futuro desconhecido podem ser minimizados ao estimar situações futuras. Como exemplo cita que quando aplicado no campo do marketing, o conjunto de conhecimentos gerado pela previsão ganha destaque em ambientes de competitividade. Além disso, afirma que os resultados gerados na abordagem quantitativa preenchem as lacunas geradas pela qualitativa, já que cria passos sistemáticos, que devem ser seguidos para se formar uma previsão.

Em sua pesquisa, Miranda et al. (2011) apresenta um conjunto de processos de previsão de demanda, compostos por um conjunto de atividades envolvidas no procedimento.

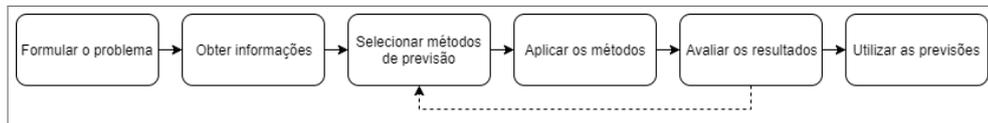
Dentre esses processos destacam-se os propostos por Kress e Snyder (1994), Scott (2001) e Werner e Ribeiro (2006). Nas figuras abaixo, é possível analisarmos as etapas propostas por cada autor.

Figura 2 – Processo de Previsão, adaptado de Kress e Snyder (1994)



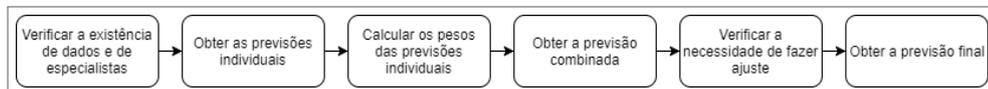
Fonte: (MIRANDA et al., 2011)

Figura 3 – Processo de Previsão, adaptado de Scott (2001)



Fonte: (MIRANDA et al., 2011)

Figura 4 – Processo de Previsão, adaptado de Werner e Ribeiro (2006)



Fonte: (MIRANDA et al., 2011)

Analisando as três figuras, é possível concluir que a etapa inicial de definição e conhecimento do problema é bastante importante para se criar um bom modelo de previsão. Obter informações internas e externas, além de contar com a assistência de um especialista fazem parte dessa primeira etapa. Após conhecer e definir bem o problema, vem a fase da escolha do modelo mais apropriado a situação, fazendo os seus ajustes, avaliando os seus resultados. Por fim, a utilização do modelo, monitorando-o.

ROSSETO et al. (2011) frisa que para realizar boas previsões da demanda existe a necessidade de conhecer bem os produtos e o mercado, envolvendo todos os setores da empresa, pois quanto mais informações disponibilizadas, melhores serão os resultados obtidos. Ou seja, essa etapa é tão importante quanto a seleção de modelos, ferramentas etc. Diante disso, podemos concluir que a previsão de demanda não é uma ciência puramente exata, já que depende de fatores externos como o próprio mercado, de baixa previsibilidade. Com isso é importante contar com variáveis externas, que projetem bem o cenário que a organização está inserida (JUNIOR et al., 2010).

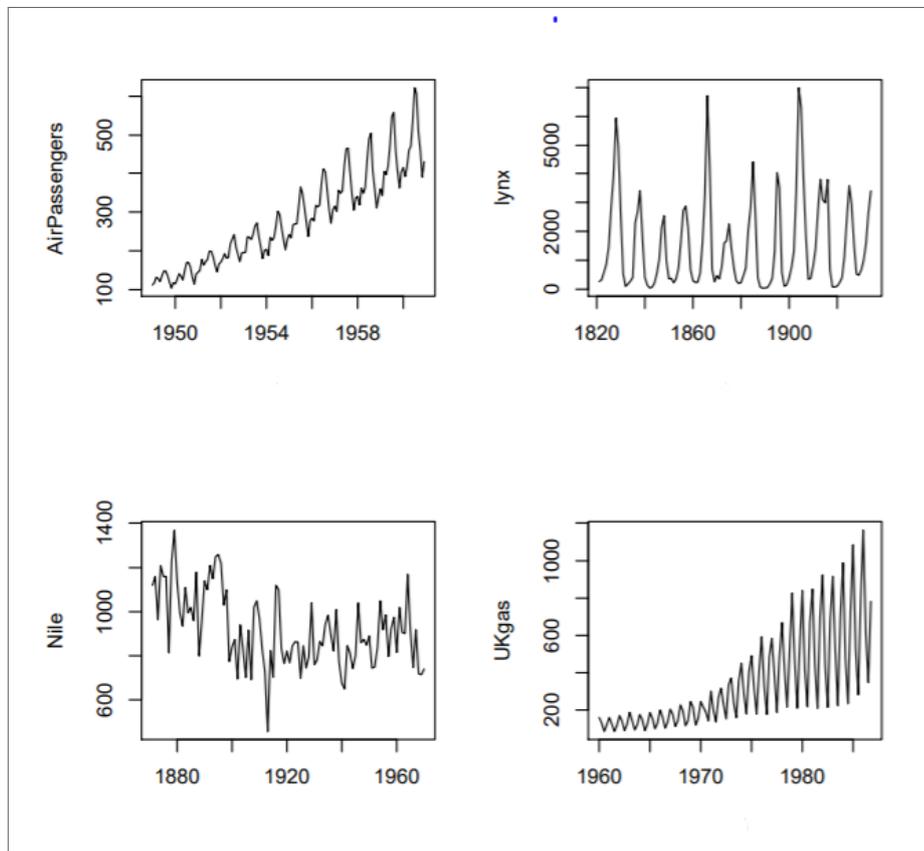
Neste trabalho o foco será dado à abordagem quantitativa, realizando análise de Séries Temporais e desenvolvendo modelos de predição com o uso de modelos de regressão. Após o desenvolvimento, serão realizadas análises comparativas com modelos previamente desenvolvidos por duas empresas, onde um delas utilizou o ARIMA.

2.2 SÉRIES TEMPORAIS

Como foi dito no capítulo anterior, esse trabalho tem como um dos objetivos específicos avaliar um conjunto de dados que contém a informação de produção de um determinado produto, desenvolvendo modelos de predição com o suporte de métodos de regressão. Tais informações são apresentadas através de Séries Temporais.

Séries temporais podem ser definidas como um conjunto de observações realizadas sequencialmente ao longo do tempo. As séries são divididas em contínuas e discretas, onde a primeira é caracterizada por observações feitas continuamente no tempo, enquanto que a segunda quando as observações são feitas em tempos equiespaçados (diariamente, semanalmente, mensalmente, anualmente etc). Além dessas características, a mais importante que define uma série temporal é a dependência existente entre as observações vizinhas. E é justamente essa dependência que se busca modelar (EHLERS, 2007).

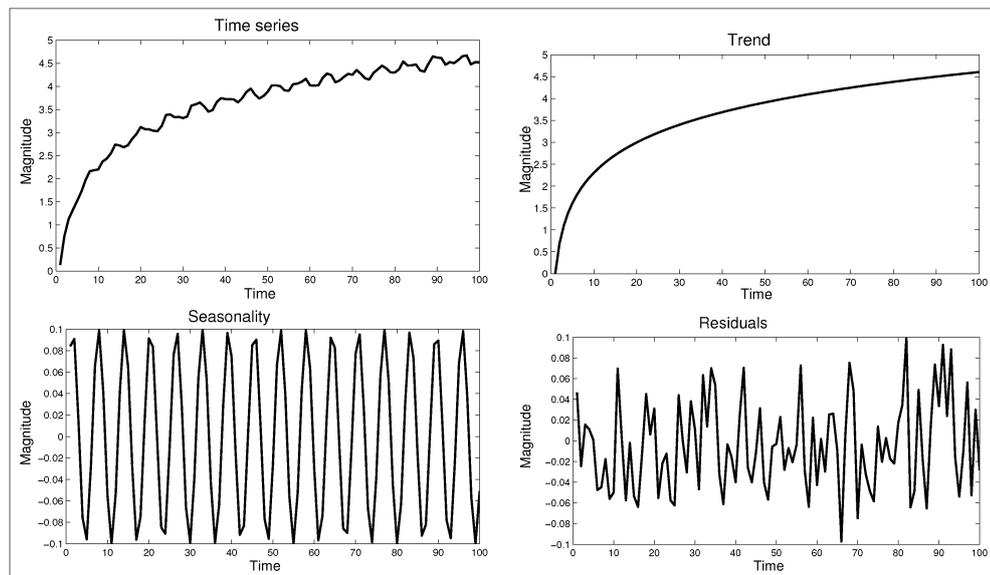
Figura 5 – Exemplos de Séries Temporais



Fonte: (EHLERS, 2007)

Na Figura 5 é possível observar diversos exemplos de séries temporais, com diferentes tipos de comportamentos e periodicidade. Para realizar a análise de uma Série Temporal, uma estratégia comum é decompô-la em seus componentes. Esses componentes são: Tendência, sazonalidade e resíduos(ou ruído). A tendência é o movimento que a variável apresenta durante o período observado. Nesse intervalo de tempo observado não é considerado a sazonalidade e irregularidades. Esse movimento pode apresentar um perfil linear, exponencial ou parabólico. A sazonalidade representa as flutuações periódicas da variável, consistindo em efeitos razoavelmente estáveis. Geralmente surge de fatores como condições climáticas, ciclos econômicos, feriados etc. Por último os resíduos, que são considerados os *outliers* da série. Normalmente são identificados após a tendência e oscilações cíclicas serem calculadas e removidas (MARTÍNEZ-ÁLVAREZ et al., 2015). Na imagem a seguir, é possível observar uma Série Temporal decomposta em seus componentes.

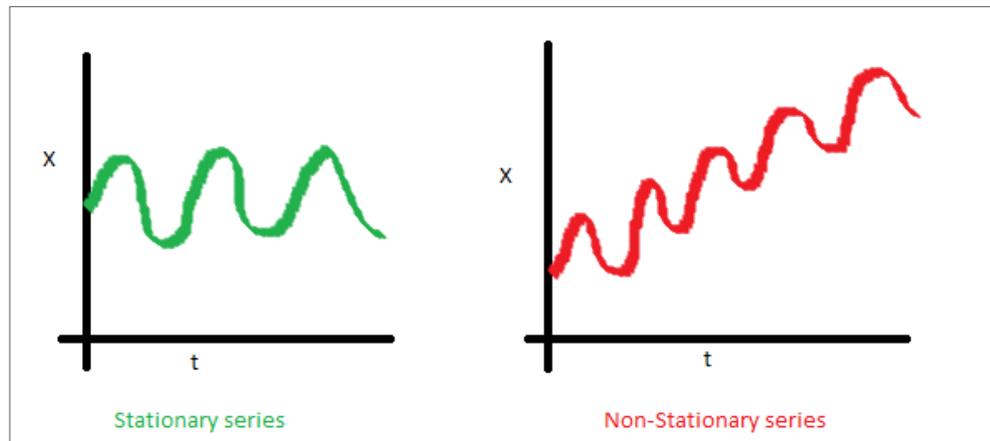
Figura 6 – Componentes de Séries Temporais



Fonte: (MARTÍNEZ-ÁLVAREZ et al., 2015)

Outra análise bastante importante a ser feita é a estacionariedade da série. Uma série temporal é dita estacionária quando apresenta características estatísticas, como média, variância, autocorrelação etc, são constantes ao longo do tempo, desenvolvendo-se aleatoriamente no tempo. Esse desenvolvimento é em torno de uma média constante, apresentando um equilíbrio estatístico estável. Identificar se uma série temporal é estacionária ou não é importante para definir o método de previsão que será empregado. Dentre as técnicas mais comuns estão as transformações para tornar uma série "estacionarizada", trazendo um comportamento mais constante no tempo (MADDALA; LAHIRI, 2009).

Figura 7 – Exemplos estacionariedade de Series Temporais



Fonte: (VIDHYA, 2015)

Diante dessas características, existem diversas formas de desenvolver um modelo para realizar a previsão ou predição de uma série temporal. Os modelos considerados adequados para dados de séries temporais são chamados de processos estocásticos, que pode ser definido como um conjunto de variáveis aleatórias ordenadas no tempo, podendo ser discretos ou contínuos (EHLERS, 2007). Tais modelos podem ser baseados em métodos lineares e não-lineares. Os modelos lineares são aqueles que modelam o comportamento da série através de funções lineares. Exemplos de algoritmos que descrevem bem essas ideias são os baseados em Auto-regressão (AR), Média Móvel (MA), Modelo Auto-regressivo de Médias Móveis (ARMA) e Modelo Auto-regressivo Integrado de Médias Móveis (ARIMA). Os modelos ARIMA é um dos mais utilizados e é um resultado da combinação dos processos AR e MA, sendo caracterizado como $ARIMA(p, d, q)$, onde p é o número de termos autoregressivos, d é o número de diferenças não sazonais e q o de erros de previsão na equação. Já os modelos não-lineares são aqueles que modelam com base em funções não lineares (MARTÍNEZ-ÁLVAREZ et al., 2015).

Nesse trabalho será realizada a comparação entre modelos ARIMA já existentes e desenvolvidos em outras empresas, com os modelos de regressão não lineares desenvolvidos durante a pesquisa.

2.3 APRENDIZAGEM DE MÁQUINA

Atualmente, algoritmos de Aprendizagem de Máquina têm sido cada vez mais usados para resolver desde problemas de nosso dia-a-dia aos mais complexos. Dentro desse contexto encontram-se as empresas que vêm utilizando tais métodos para melhorar as tomadas de decisões de negócios, aumentar produtividade, melhorar a competitividade no mercado e muito mais. Tudo isso tem se tornado possível graças ao crescimento exponencial da tecnologia nos últimos anos (HARRINGTON, 2012).

Os algoritmos de Aprendizagem de Máquinas consistem em métodos que têm o objetivo de realizar atividades preditivas. Através de métodos estatísticos, os algoritmos podem expressar os dados analisados como uma função matemática, que é definida durante uma fase conhecida como treinamento. Nessa etapa, dados históricos são submetidos aos algoritmos, que identificam seu comportamento (BISHOP, 2006).

Em seu livro Harrington (2012) cita os seguintes passos para desenvolver uma aplicação com Aprendizagem de Máquinas:

1. Coletar os dados: consiste em reunir dados do problema que se tem o objetivo de solucionar.
2. Preparar os dados de entrada: após juntar os dados, verificar se estão em um formato utilizável. Nessa etapa deve ser observado se os dados encontram-se no formato ideal para o algoritmo que será utilizado.
3. Analisar os dados de entrada: após a preparação da fase anterior, deve ser estudado os dados que serão submetido aos métodos. Nessa etapa é comum observar quais dados são numéricos ou categóricos, construir gráficos a fim de conhecer melhor as informações contidas na base de dados. Além disso, podem ser utilizados métodos mais complexos para seleção de atributos, transformações de dados etc.
4. Treinar o algoritmo: essa é a etapa onde a aprendizagem de máquina começa a ser aplicada ativamente. É nesse momento onde os algoritmos começam a identificar padrões através de métodos estatísticos, construindo a função matemática que fará a predição dos dados.
5. Testar o algoritmo: após o treinamento, os modelos desenvolvidos são analisados. Esse teste é realizado com a submissão de um conjunto de dados, que não foi submetido na fase anterior e que já possui os valores reais, a fim de observar se o modelo realiza as predições de forma correta.
6. Aplicar o algoritmo: por último, caso na fase anterior o modelo esteja obtendo um bom desempenho, aplicar o algoritmo para resolver o problema estudado. Nessa etapa é importante manter o modelo sendo avaliado, para identificar o momento que passe a ter uma diminuição no desempenho e novos dados sejam coletados e o processo reiniciado.

Diante das características e etapas citadas anteriormente é importante frisar que os algoritmos de aprendizagem são divididos em Algoritmos supervisionado e Não Supervisionados. O primeiro tipo é aplicado quando os dados que serão utilizados no treinamento do modelo são conhecidos, ou seja, os dados encontram-se rotulados com uma categoria ou informação contínua. Nesse caso quando os dados a serem preditos possuem um conjunto finito de categorias, são chamados problemas de classificação e quando esse conjunto

consiste em valores contínuos, são chamados problemas de regressão. No segundo tipo, a aplicação é realizada quando se deseja separar e agrupar dados de acordo com os padrões identificados e os seus rótulos não são conhecidos.

Nesse trabalho foi utilizado Métodos Supervisionados para a predição de valores contínuos, que são a quantidade em litros que uma empresa deverá produzir em uma determinada semana, prevendo a demanda semanal de um determinado produto.

2.3.1 Métodos de Regressão

Como foi dito na seção anterior, os métodos de regressão são utilizados quando o objetivo é prever variáveis numéricas e contínuas (HARRINGTON, 2012). Nesta seção serão apresentados os algoritmos utilizados no desenvolvimento dessa pesquisa. São eles: Linear Regression, Multilayer Perceptron, Support Vector Regression, Stochastic Gradient Descent, Decision Tree, Random Forest e XGBoost Regressor.

2.3.1.1 Linear Regression

O regressão linear é caracterizada por ser uma das formas mais simples de modelar um conjunto de dados. Essa simplicidade se deve ao fato de que os modelos são baseados em funções lineares polinomiais. Neles o modelo mais simples consiste em uma combinação linear das variáveis de entrada e seus respectivos pesos. A fórmula a seguir é um exemplo da representação dessa base linear (BISHOP, 2006).

$$y(X, W) = w_0 + w_1x_1 + \dots + w_Dx_D \quad (2.1)$$

Na equação 2.1 X é um vetor de variáveis $(x_1, x_2, \dots, x_D)^T$, enquanto W é o vetor com os pesos ou coeficientes de cada variável de entrada. Partindo dessa base linear, um modelo simples de regressão linear consiste na relação linear entre um regressor x e uma resposta y que é uma linha reta (MONTGOMERY; PECK; VINING, 2012).

$$y = \beta_0 + \beta_1 + \epsilon \quad (2.2)$$

Na equação 2.2, β_0 e β_1 são constantes desconhecidas e ϵ é um componente de erro aleatório, que se assume ter uma média zero e variância desconhecida. As variáveis de erros são descorrelacionados, ou seja, não dependem de qualquer outro valor de erro. Diante dessa equação é conveniente concluir que o regressor x é um fator controlado e medido com um erro desprezível, enquanto que o resultado y é uma variável aleatória. Os parâmetros β_0 e β_1 são chamados de coeficientes de regressão, onde β_1 é a inclinação representada pela mudança na média de distribuição y , resultante de uma mudança unitária em x (MONTGOMERY; PECK; VINING, 2012).

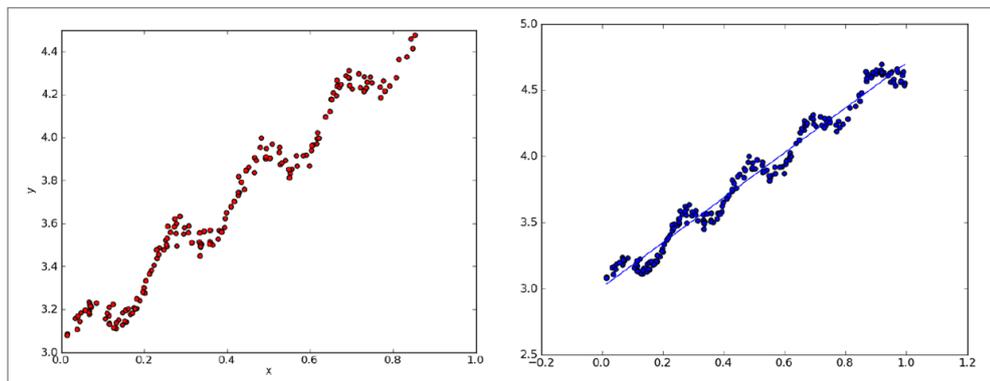
Para estimar os parâmetros dessa função, geralmente é utilizado o método de estimativa ordinária de mínimos quadrados. Tal função tem o objetivo de selecionar os

parâmetros que minimizarão a soma residual dos quadrados entre os valores observados no conjunto de dados e os valores previstos pela aproximação linear (WEISBERG, 2005). A seguir é ilustrada a equação utilizada para o cálculo dos mínimos quadrados ordinários na ferramenta *Scikit-learn*, utilizada nesse trabalho (SCIKIT-LEARN, 2019)

$$\min_w \|X_w - y\|_2^2 \quad (2.3)$$

Na figura a seguir, é mostrado um exemplo de um conjunto de dados, seguido pela solução dada para a melhor linha que se ajusta a esses dados.

Figura 8 – Exemplo modelo regressão linear ajustado



Fonte: (HARRINGTON, 2012)

2.3.1.2 Multilayer Perceptron

Os algoritmos de Redes Neurais (Artificial Neural Networks) foram idealizados em 1943 por McCulloch e Pitts, com o objetivo de resolver problemas através de um sistema inspirado no cérebro humano e sua estrutura neural. Nesse sistema suas unidades de processamento simbolizariam os neurônios e suas combinações ocorreriam através de conexões semelhantes ao sistema nervoso central humano (MARTÍNEZ-ÁLVAREZ et al., 2015).

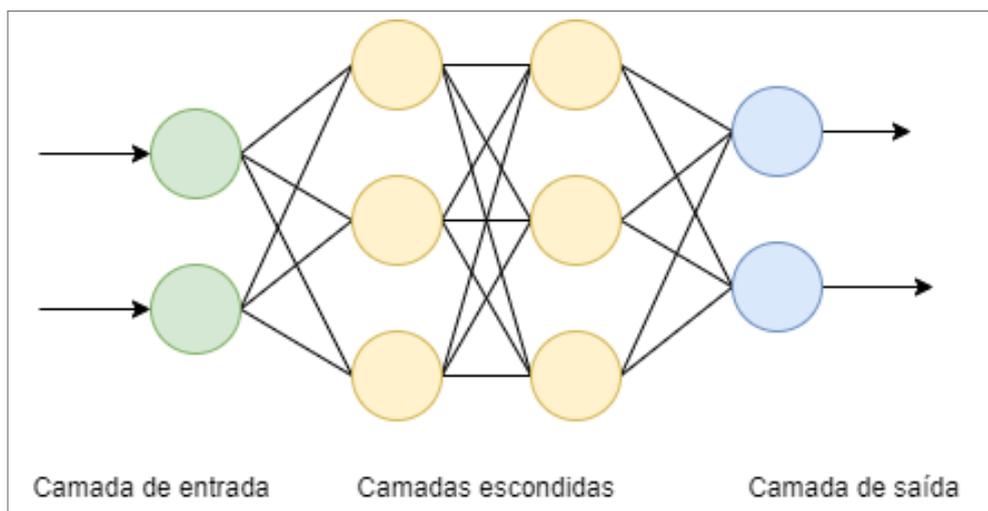
Em seu trabalho, os autores Martínez-Álvarez et al. (2015) citam que existem 3 características principais que descrevem uma Rede Neural:

1. Topologia da Rede Neural: Consiste na organização e posição dos neurônios na rede. Representado pelo número de camadas, número de neurônios em cada camada, tipo de conexões entre os neurônios etc. Dependendo da quantidade de camadas, a Rede Neural pode ser classificada em Monolayer ou Multilayer (MLP).
2. Paradigma de aprendizagem: O processo de aprendizagem de uma Rede Neural consiste na modificação dos pesos de cada conexão pertencente a rede. Esses pesos são ajustados a medida que as informações de entrada são processadas.

3. Representação das informações de entrada e saída: As Redes Neurais também são classificadas de acordo com a representação dos dados de entrada e saída. Essas diferentes formas de representação implicam em diferentes funções de ativação.

Como foi dito nas caracterizações acima, uma rede neural pode ser classificada de acordo com a quantidade de camadas que possui. O algoritmo MLP (Multilayer Perceptron) pertence a uma classe de algoritmos de Redes Neurais, contendo elementos de computação não-linear organizados em uma estrutura de camadas avançadas (BOURLARD; KAMP, 1988). Para o seu treinamento, o MLP utiliza uma técnica de aprendizagem supervisionada conhecida como backpropagation (ROSENBLATT, 1961) (RUMELHART; HINTON; WILLIAMS, 1985). O objetivo dessa técnica é encontrar o gradiente de uma função de erro, passando uma mensagem na rede neural, tanto no sentido posterior, quanto no sentido anterior (BISHOP, 2006). Juntamente com essa técnica, há a aplicação de um procedimento conhecido como Regra Delta, que é um algoritmo que calcula os erros de uma determinada saída e compara com a saída desejada, ajustando os pesos das conexões, a fim de diminuir essa diferença. Uma das vantagens dos algoritmos de Rede Neurais são suas multicamadas que tornam esse modelo poderoso, possuindo propriedades que podem ser aplicadas em diversas situações no mundo real. Apesar dessa aplicabilidade, é nela que também reside um de seus problemas, que é ajustar ou regularizar os parâmetros para cada tipo de problema. Um mal ajuste pode gerar generalizações fracas e um aprendizado não adequado (DUDA; HART; STORK, 2012).

Figura 9 – Exemplo arquitetura Rede Neural

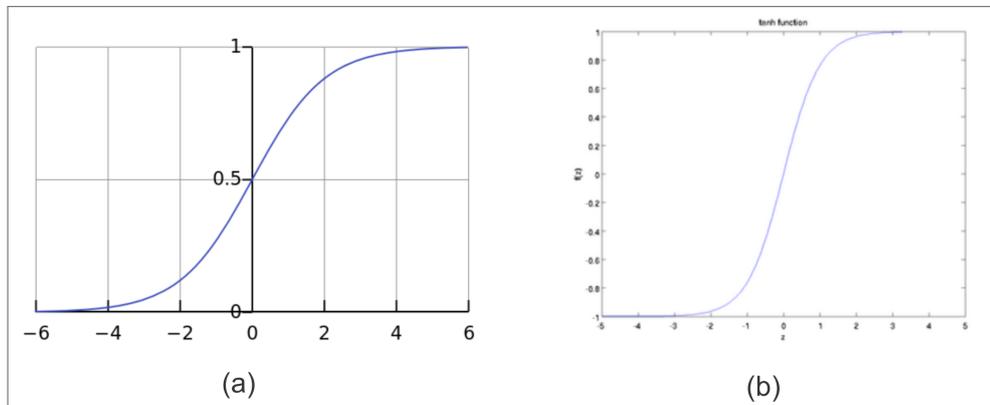


Na figura 9 encontramos um exemplo de arquitetura multicamadas. Cada unidade representada pelos círculos verdes, amarelos e azuis são os neurônios e as arestas são as conexões entre eles. Os neurônios na cor verde compõem a camada de entrada, responsável por receber os dados de entrada do modelo. As unidades em amarelo formam as camadas escondidas que contém os neurônios que processam as informações de entrada. Por último,

a camada em azul contém os neurônios de saída, que são responsáveis por retornarem o resultado do modelo.

Como foi dito anteriormente, as Redes Neurais possuem uma base não linear. Em suas unidades de processamento, o fator responsável por fazer essas transformações não lineares são as funções de ativação. Essa transformação é importante para os neurônios, pois indicam se a informação que está sendo recebida é relevante ou não, sendo essa etapa conhecida como ativação do neurônio. Normalmente as funções escolhidas são as sigmóides, como as sigmóides logísticas e tangente hiperbólica. Essa escolha é determinada de acordo com a distribuição dos dados analisados (BISHOP, 2006).

Figura 10 – Exemplos de funções de ativação. (a) Função Sigmóide Logística (b) Tangente Hiperbólica

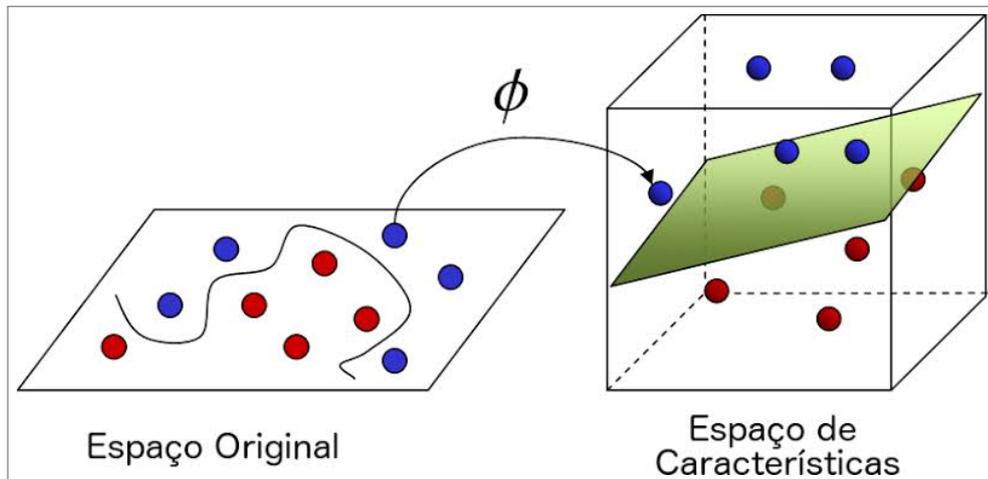


Fonte: (SHARMA, 2017)

2.3.1.3 Support Vector Regression

O algoritmo Support Vector Machines (SVM) é um classificador utilizado em problemas que buscam classificar dois grupos. Para atingir esse objetivo, o SVM implementa um modelo que recebe vetores que não são mapeados linearmente e os projeta em uma dimensão mais alta para que possam ser mapeados linearmente, garantindo uma decisão linear (CORTES; VAPNIK, 1995). Para que esse mapeamento seja realizado é necessário que os dados sejam pré-processados a fim de que seus padrões sejam representados em uma dimensão acima da sua origem. Esse aumento de dimensão é proporcionado pela aplicação de uma função não linear, conhecida como kernel, que realizará o mapeamento dos dados nessa alta dimensão, tornando possível a separação dos dados. As informações são divididas por um hiperplano, que é um espaço vetorial ou um sub-espaço em uma dimensão (DUDA; HART; STORK, 2012).

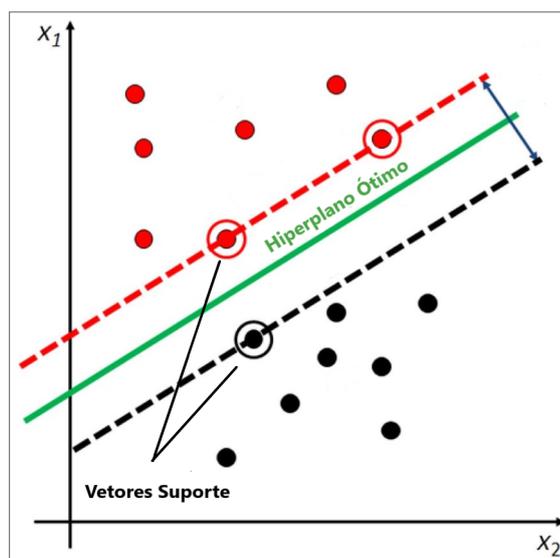
Figura 11 – Exemplos de aplicação função não linear (kernel) e hiperplano



Fonte: (BONIFÁCIO et al., 2017)

Na figura 11 é possível observar a aplicação da função não linear (kernel), realizando a transformação dos dados para uma dimensão mais alta, 2D para 3D, sendo separado perfeitamente por um hiperplano. Existem muitos algoritmos que realizam a divisão de dados com o apoio de um hiperplano. Além disso, nessa separação existem infinitos hiperplanos que podem dividir o conjunto de dados. Logo, o ponto principal do SVM é encontrar uma fronteira de decisão, hiperplano ótimo, que separa os dados, e ao mesmo tempo, maximiza a distância entre os pontos com maior proximidade em relação a uma das classes. Para conseguir isso, é levado em conta apenas os pontos pertencentes às fronteiras da região de decisão, que são os pontos que claramente não pertencem a um classe ou outra, que são chamados de Vetores de Suporte (MARTÍNEZ-ÁLVAREZ et al., 2015).

Figura 12 – Exemplos de aplicação função não linear (kernel) e hiperplano



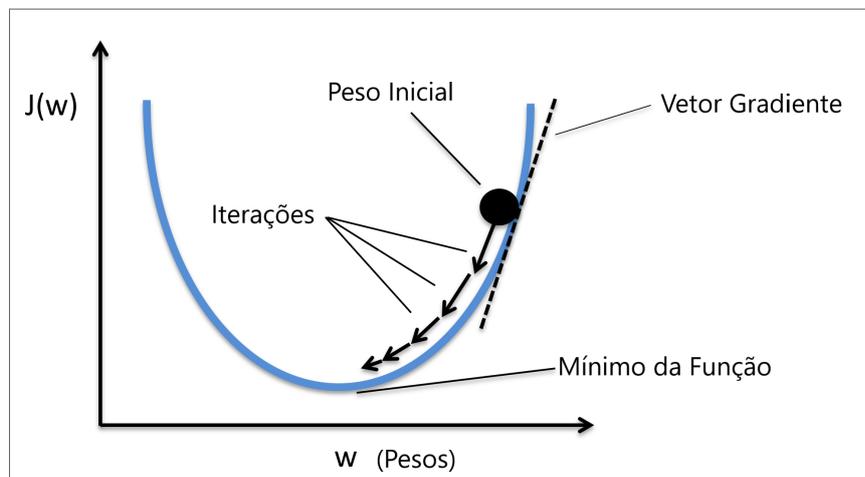
Fonte: (BONIFÁCIO et al., 2017)

Além de resolver problemas de classificação, SVMs também são aptas a desenvolver modelos de regressão. Nesses casos, são conhecidos como SVR, comumente utilizados para a previsão de séries temporais. Nesses casos o objetivo é determinar uma função não-linear que se afaste das saídas medidas por um valor inferior a um termo de erro, minimizando o erro de previsão do conjunto de treinamento. Diante dessa informação é possível concluir que o desempenho geral dos modelos de SVR dependem diretamente da seleção adequada do kernel e seus parâmetros (MARTÍNEZ-ÁLVAREZ et al., 2015) (DEB et al., 2017). Uma das vantagens desse algoritmo é o fato de não pressupor linearidade do modelo, quando são utilizadas funções de kernel não lineares, sem necessitar de normalização dos dados, se adaptando prontamente a dados de alta dimensionalidade (BONIFÁCIO et al., 2017).

2.3.1.4 Stochastic Gradient Descent

O SGD é uma simplificação do algoritmo de GD, que tem o objetivo de realizar o ajuste dos parâmetros (pesos), através de iterações, minimizando a função de erro. A minimização desse erro visa encontrar o mínimo local da função utilizada (BOTTOU, 2010). Nesse caso, o Gradient Descent pode ser interpretado como um vetor que aponta a direção da diminuição da função (HARRINGTON, 2012).

Figura 13 – Gráfico com aplicação do GD



Fonte: Baseado em (MLXTEND, 2019)

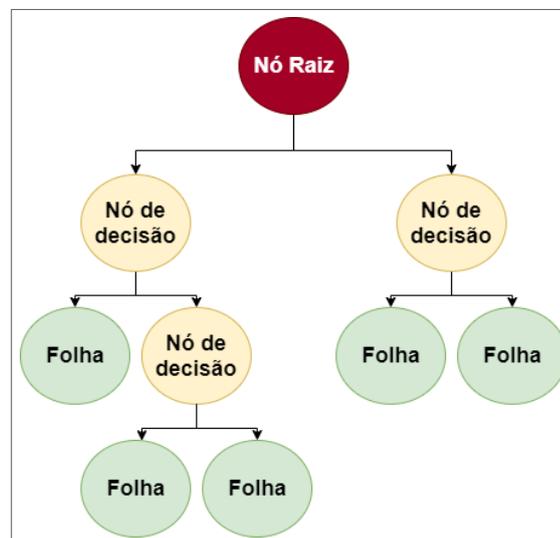
O Stochastic Gradient Descent utiliza gradientes que são estimados a partir de um subconjunto dos dados de treinamento. Essa abordagem o torna mais rápido, pois requer menos tempo de treinamento (TSURUOKA; TSUJII; ANANIADOU, 2009). Nele o processo estocástico depende dos exemplos de parâmetros escolhidos aleatoriamente para cada iteração. Com esse comportamento aleatório, o algoritmo pode processar os exemplos instantaneamente, já que não precisa se lembrar de quais exemplos foram visitados e conseqüentemente a descida do gradiente se torna otimizada (BOTTOU, 2010).

Bottou (2010) demonstra que com essa abordagem, o SGD torna-se bastante eficaz ao lidar com um grande conjunto de dados. Porém, essa eficácia sofre uma diminuição quando se depara com uma grande quantidade de dados ruidosos. A convergência da descida do gradiente estocástico prejudica-se pois os ganhos do método diminuem muito lentamente, tornando a variação dos parâmetros também lenta.

2.3.1.5 Decision Tree

O algoritmo de Decision Tree ou *Árvore de Decisão* é uma das ferramentas simples da aprendizagem de máquinas, porém eficaz para solucionar problemas de classificação e regressão. Ao invés de utilizar um conjunto de recursos para executar a classificação em conjunto, as *Árvores de Decisão* utiliza um esquema de decisão em vários estágios, sendo eles hierárquicos ou em estrutura de árvore. Tal estrutura é composta por um conjunto de nós que simbolizam a tomada de decisão, sendo essa binária que separa um ou mais classes da outra. Os nós recebem um termo de acordo com o seu posicionamento, podendo ser chamados de *nó raiz*, que contém todos os dados, *nós internos* ou *nós de decisão*, que são as divisões e *folhas*, que são os nós terminais (XU et al., 2005).

Figura 14 – Estrutura de uma *Árvore de Decisão*



Fonte: Autor

Magerman (1995) define uma *Árvore de decisão* como um dispositivo de tomada de decisão que atribui probabilidades a cada uma das opções possíveis dentro do contexto da decisão. Essas probabilidades são definidas após uma sequência de perguntas serem respondidas. Esses questionamentos são caracterizados pelos nós citados anteriormente. A probabilidade é calculada através da probabilidade condicional de $P(f|h)$, onde f é o evento futuro e h o histórico dos dados.

Logo, a modelagem das *Árvores de decisão* representa um conjunto de dados segmentados através de uma série de regras simples, gerando um conjunto de regras que tornam

possível a previsão através do processo repetitivo da divisão. Portanto, o objetivo básico visado na modelagem é realizar a divisão de uma decisão complexa em várias decisões mais simples, que podem levar a uma solução mais fácil de interpretar (TSO; YAU, 2007) (XU et al., 2005).

Como foi dito no início dessa seção, as Árvores de Decisão também solucionam problemas que envolvem regressão. Essa solução é construída se baseando na ideia de que os relacionamentos os atributos e valor a ser predito são lineares ou não lineares. Na regressão, as árvores são variantes de um classificador usado para aproximar funções de valor real. Por conseguinte, a construção de uma Árvore de Regressão também se baseia no particionamento binário, onde inicialmente toda a amostra dos dados são usadas para definir a estrutura da árvore e a divisão é selecionada buscando minimizar a soma dos desvios quadrados da média nas partes separadas. Nesse processo de treinamento é importante atentar a possibilidade do modelo sofrer um ajuste excessivo, prejudicando a capacidade de generalização, processo conhecido como *overfitting*. A fim de evitar isso é importante utilizar um conjunto de validação, que auxilia nessa etapa de ajuste da árvore (XU et al., 2005).

2.3.1.6 Random Forest

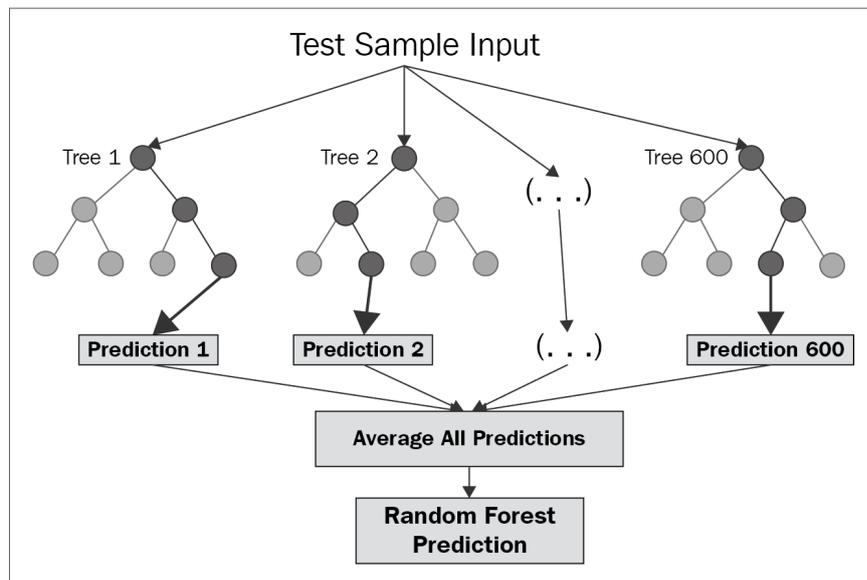
Random Forest (RF) ou Florestas Aleatórias são combinações de árvores de decisão, construído de forma que cada árvore dependa dos valores aleatórios de uma amostra independente e com a mesma distribuição do conjunto de dados analisados (BREIMAN, 2001). Essa técnica que gera a combinação de múltiplos classificadores e agregam os seus resultados é conhecido como "Ensemble Learning". Dessa técnica, as duas abordagens mais conhecidas são *boosting* e *bagging*. Na primeira, árvores sucessivas atribuem pesos extras aos pontos previstos erroneamente pelos classificadores anteriores, gerando no final uma votação ponderada para a previsão final. Já na segunda, as árvores não dependem entre si, sendo construída de forma independente. No fim, também é executada uma votação, mas dessa vez sendo majoritária e simples (LIAW; WIENER et al., 2002).

Breiman (2001) propôs que o Random Forest adicionassem uma camada de aleatoriedade para a técnica de Bagging. Além da construção das árvores com amostras diferentes dos dados estudados, houve uma alteração na forma como as árvores são construídas. As árvores de decisão padrão são construídas com cada nó sendo dividido usando a melhor divisão entre todas as variáveis, enquanto que nas Florestas Aleatórias os nós se dividem usando o melhor entre um subconjunto de preditores escolhidos aleatoriamente (LIAW; WIENER et al., 2002).

Random Forest também soluciona problemas de regressão, seguindo a mesma ideia de se obter um conjunto de diferentes Árvores de Regressão. Nesse contexto cada árvore encarrega-se de realizar um mapeamento não linear, dividindo o problema original em problemas menores, solucionável por um conjunto de modelos simples (PHAN et al., 2014).

Para isso assume-se que o conjunto de treinamento é modelado independentemente dos vetores aleatórios utilizados como entrada do modelo. Além disso, as florestas de regressão precisam de uma baixa correlação entre os resíduos e árvores com taxa baixa de erro, ou seja, a randomização aplicada precisa ter como objetivo uma baixa correlação. Diante dessas características, é possível se concluir que o Random Forest é uma excelente ferramenta para solucionar problemas de regressão e classificação, já que por contar com uma combinação de árvores os modelos não se ajustam demais, causando um problema de *overfitting*, tendo assim um grande poder de generalização (BREIMAN, 2001).

Figura 15 – Estrutura Random Forest



Fonte: (TOWARDS-DATA-SCIENCE, 2019)

2.3.1.7 XGBoost Regressor

XGBoost é uma implementação escalável e eficiente da estrutura *gradient boosting* de Friedman (2001), sendo a abreviação do pacote *eXtreme Gradient Boosting*. Nesse pacote encontra-se um solucionador para modelos lineares e um algoritmo de aprendizado em árvore, podendo resolver problemas de regressão, classificação e raking (CHEN et al., 2015).

O impacto desse algoritmo foi amplamente reconhecido em vários desafios de aprendizagem de máquinas e mineração de dados. Isso se deve ao fato de que o pacote oferece resultados muito bons em diferentes tipos de problemas. Dentre os problemas resolvidos podemos citar: previsão de vendas, classificação de texto na web, previsão de comportamento de cliente, categorização de produtos etc. Tais soluções para esses problemas serão apresentados no próximo Capítulo, com exemplos de pesquisas que utilizaram o XGBoost para essa finalidade. Todo esse sucesso é justificado pela escalabilidade do algoritmo em todos os cenários. Além disso, o XGBoost roda dez vezes mais rápido do que as soluções populares existentes. Isso é possível já que o seu algoritmo consiste na otimização e

combinações de técnicas como Gradiente Boosting, que utiliza o algoritmo de gradiente descendente para a minimização dos erros em conjunto com técnicas de *ensemble*. Ou seja, são consideradas a implementação de um conjunto de árvores de decisão otimizadas por gradiente (CHEN; GUESTRIN, 2016).

Além dessas características, Chen et al. (2015) e Dhaliwal, Nahid e Abbas (2018) citam que existem outras vantagens na utilização do XGBoost como:

- Capacidade para processamento paralelo, em que o algoritmo utiliza todos os núcleos da máquina que o executam tornando-o altamente eficaz para pré-processamento de dados em alto nível
- Aceita vários tipos de dados como entrada, podendo ser uma matriz densa ou esparsa e arquivos locais.
- O algoritmo é customizável, suportando funções de objetivo e avaliação personalizadas.
- Pode ser executado em diversas plataformas, como por exemplo: AWS, Azure, Alibaba etc.
- Suporte para diversas linguagens de programação.
- Capacidade de transformar um aprendiz fraco em forte, por meio da etapa de otimização para cada árvore implementada no sistema.
- Validação cruzada como recurso interno.
- Regularização que evita problemas de *overfitting*
- Bem preparado para detectar e lidar com *missings* ou valores ausentes.

3 TRABALHOS RELACIONADOS

Este capítulo tem como objetivo apresentar um levantamento geral sobre as pesquisas e trabalhos desenvolvidos que possuem um propósito similar ao desta dissertação. Foram consideradas pesquisas similares a este trabalho aqueles que abrangiam assuntos como previsão de demanda, modelos de regressão, abordagens para análise de séries temporais e algoritmos de aprendizagem de máquinas lineares e não lineares. Diante disso, a seguir serão apresentados trabalhos relacionados, em que foram aplicados algoritmos de aprendizagem de máquina, com foco na previsão de demanda através de regressores. Vale evidenciar que não é objetivo desse capítulo apresentar uma revisão sistemática formal, e sim a introdução e apresentação de trabalhos científicos que tiveram como objetivo solucionar problemas de previsão de demanda.

3.1 METODOLOGIA

Para que o objetivo desse capítulo seja alcançado, a revisão bibliográfica teve como foco mapear as fontes de divulgação científica, com o intuito de identificar trabalhos similares na área. Com isso, foi utilizada a plataforma *Google Scholar*, que provém a possibilidade de ser pesquisado trabalhos científicos de diversas áreas e fontes, sendo elas artigos, teses, livros etc, utilizando diversos filtros.

3.2 ALGORITMOS ESTUDADOS

As seções a seguir serão divididas com o intuito de apresentar trabalhos que foram desenvolvidos tendo como base a aplicação do método ARIMA e dos algoritmos de regressão que foram utilizados para a previsão de demanda.

3.2.1 ARIMA

Ainda que existam diversas técnicas com o objetivo de realizar a previsão de Séries Temporais, o algoritmo ARIMA ainda é um dos mais utilizados para atingir esse propósito. Na literatura, podemos encontrar diversos trabalhos que citam sua aplicação na Previsão de Demanda.

Em pesquisa recente, Abolghasemi et al. (2020) aplica o método ARIMA, em uma de suas variações, com o intuito de realizar a previsão de demanda de uma cadeia de suprimentos, investigando diversas séries temporais, com diferentes tipos de comportamentos. Nesse trabalho, o desafio era alta volatilidade contida nas Séries Temporais, causada por promoções durante as vendas dos produtos. Os autores destacam a importância da análise dos componentes das séries Temporais como a Sazonalidade e Tendência. Seguindo

o mesmo objetivo, para a previsão de demanda em um centro de distribuição de produtos, os autores Rizkya et al. (2019) utilizam o ARIMA, procurando o melhor ajuste dos parâmetros, para cada centro de distribuição separadamente. Nesse trabalho os autores destacam a eficiência que o método ARIMA apresenta na previsão de séries temporais em um curto período do tempo e que possuem instabilidades nos dados. Além disso ressaltam a importância de análises como a estacionariedade da série.

Já em sua pesquisa, a fim de prever a demanda de uma agência farmacêutica, o pesquisador Ayele (2019) realizou um estudo comparativo entre os algoritmos Long Short Term Memory (LSTM), MLP e ARIMA, com o objetivo de utilizar o algoritmo que obtivesse o melhor desempenho para prever a demanda dos produtos. Nesse estudo, utilizando a métrica RMSE, concluiu-se que o método Auto-ARIMA oferece o melhor desempenho de previsão.

Também no ano de 2019, Noureen et al. (2019) desenvolveram um trabalho que tinha como objetivo realizar a previsão de séries temporais sazonais utilizando o método ARIMA, utilizando um conjunto de dados em pequena escala. Nesse trabalho os autores relatam que a análise da estacionariedade da série temporal é importante, pois para uma previsão eficiente é necessário que a série apresente um comportamento estacionário, informando também que as variações sazonais tornam uma série temporal não estacionária. Para essa análise, entre uma das técnicas, utilizaram o teste de Dickey Fuller Aumentado (ADF).

3.2.2 Linear Regression

Apesar de ser considerado um método simples para realizar a predição de dados contínuos, a regressão linear é um método que vem sendo analisado e estudado para promover a previsão de demandas.

No ano de 2010, Hong et al. (2010) desenvolveram uma pesquisa com o intuito de modelar a demanda horária de consumo de energia de uma empresa norte-americana, utilizando a regressão linear múltipla. No modelo foi utilizado como variável externa a temperatura, buscando uma correlação entre ela e o consumo de energia elétrica. Os autores apresentam que os modelos desenvolvidos obtiveram uma boa interpretabilidade do problema, auxiliando no gerenciamento do consumo de energia, alcançando um MAPE de 4.55% no seu melhor modelo. Já em 2012, Kazemi, Hosseinzadeh et al. (2012) desenvolveram um trabalho que tinha como objetivo desenvolver um modelo de previsão para prever a demanda de energia anual, no setor industrial no Irã. Nesse trabalho, foram utilizados dados no intervalo de 1994 a 2008 para desenvolver o modelo e de 2011 a 2020 para prever e avaliar o modelo. Na modelagem foram utilizados dados de indicadores socioeconômicos e um modelo de regressão linear difusa (Fuzzy Linear Regression). Seguindo o mesmo propósito, (BIANCO; MANCA; NARDINI, 2013) e (VU; MUTTAQI; AGALGAONKAR, 2015) também propuseram modelos que envolvem o consumo de energia elétrica. No primeiro,

Bianco, Manca e Nardini (2013) desenvolveram modelos de regressão linear simples e múltipla, analisando a influência de variáveis econômicas e demográficas no consumo anual de eletricidade na Itália, obtendo uma taxa de erro $\pm 5\%$. Já no segundo, Vu, Muttaqi e Agalgaonkar (2015) utilizaram um modelo de regressão linear múltipla e métodos de seleção de atributos para a seleção de variáveis climáticas mais apropriadas para auxiliarem na previsão.

Além do contexto de consumo de eletricidade, encontram-se trabalhos que buscam realizar a previsão de consumo de água em cidades. Seguindo dessa ideia Bakker et al. (2014) elaboraram um trabalho em que o objetivo é realizar a previsão de consumo de água, também utilizando dados externos como as condições climáticas. Os autores utilizaram regressão linear múltipla e fizeram uma comparação dos resultados com e sem as variáveis externas. No final concluíram que a utilização de variáveis externas trazem melhorias ao modelo.

Além dessas áreas, Fang e Lahdelma (2016) utilizaram regressão linear múltipla para realizar a previsão de demanda no sistema de aquecimento urbano de uma cidade da Finlândia. Além da análise com variáveis externas, realizaram uma comparação e combinação com modelos SARIMA. Após as análises, concluíram que o modelo de regressão linear apresenta os melhores resultados na prática.

3.2.3 Multilayer Perceptron

Os algoritmos Artificial Neural Networks (ANN) é um dos mais citados na literatura para resolver problemas de regressão, principalmente devido a sua grande capacidade de resolver problemas não lineares (MARTÍNEZ-ÁLVAREZ et al., 2015). Da forma similar à seção anterior, foram encontrados trabalhos que tinham como objetivo realizar a predição de consumo de energia elétrica. Em seu trabalho Martínez-Álvarez et al. (2015) apresentam diversos trabalhos com esse objetivo, demonstrando que a área de consumo elétrico tem utilizado bastante modelos de regressão para previsão de demanda. Porém nessa pesquisa os trabalhos listados não são tão recentes, com publicações anteriores ao ano de 2010. Em uma nova análise foram encontrados trabalhos como o de Vinagre et al. (2016), em que os autores utilizaram MLP como um dos métodos para previsão de consumo de energia elétrica, em Portugal, analisando a correção entre o consumo elétrico e a radiação solar. Na mesma área também é possível encontrar trabalhos como o de (KATSATOS; MOUSTRIS, 2019), que desenvolveram modelos com Redes Neurais a fim de prever o consumo de energia em Atenas, Grécia. Foram desenvolvidos 9 modelos, pra prever um período específico no ano, utilizando dados meteorológicos como variáveis do modelo. Os resultados alcançados foram satisfatórios, conseguindo prever o consumo nas próximas 24h.

Além da área citada anteriormente, encontram-se trabalhos como o desenvolvido por Szoplik (2015), que projetou o treinamento de MLP para realizar a predição do consumo de gás natural na Polônia, por consumidores individuais e pequenas empresas. No modelo

desenvolvido foram considerados os elementos da data (mês, dia, dia da semana, hora) e a temperatura, concluindo-se que a utilização de Redes MLP de alta qualidade trazem resultados que podem ser utilizados para prever o consumo em qualquer dia do ano e em qualquer horário. Altunkaynak e Nigussie (2018) desenvolveram um trabalho em que utilizaram MLP combinado com técnicas que visam o pré-processamento dos dados, para realizar a previsão do consumo de água em Istambul, Turquia. A utilização dessas técnicas no trabalho desenvolvido procuravam a melhoria do desempenho dos modelos de Redes Neurais, na modelagem de Séries Temporais, como o tratamento de séries não estacionárias. Os resultados mostraram que tais técnicas influenciaram positivamente no desempenho dos modelos de previsão.

No contexto de produção de produtos em fábricas, encontram-se trabalhos como o dos autores (SLIMANI; FARISSI; AL-QUALSADI, 2016) e (BIBAUD-ALVES; THOMAS; BRIL, 2019). No primeiro trabalho, os autores utilizaram dados de um supermercado no Marrocos, para modelar um sistema que preveria a demanda nas cadeias de suprimentos dessa organização. No segundo, o objetivo é prever a demanda de vendas e produção para uma empresa fabricante de móveis. Para realizar essa tarefa são relatadas as dificuldades para lidar com a sazonalidade e ao tamanho pequeno do conjunto de dados.

3.2.4 SVR

De forma semelhante às seções anteriores, o algoritmo de SVR vem sendo utilizado para prever a demanda de energia elétrica. Guo et al. (2017) utilizaram esse algoritmo para realizar a previsão de consumo de eletricidade em uma província da China. Nesse trabalho os autores focaram no ajuste de hiper-parâmetros do modelo e utilização de técnicas de pré-processamento para lidar com uma amostra pequena dos dados. A fim de resolver o mesmo problema, Son e Kim (2017) desenvolveram modelos de regressão que previam o consumo um mês a frente, no setor residencial na Coreia do Sul. Para isso utilizaram variáveis que informavam o consumo mensal da eletricidade, variáveis meteorológicas e sociais. Devido a essa grande quantidade de atributos, usaram o algoritmo Particle Swarm Optimization (PSO) para fazer a seleção de características relevantes. Posteriormente, com a finalidade de demonstrar a eficiência do modelo desenvolvido, foram feitas comparações com modelos construídos a partir de ARIMA, Redes Neurais e Regressão Linear, concluindo-se que o SVR trabalhando de forma conjunta com algoritmos de seleção de características conseguem atingir resultados melhores que os anteriores. Já Al-Musaylh et al. (2018) realizaram uma comparação entre o SVR, ARIMA e Multivariate Adaptive Regression Spline (MARS) para prever o produção de energia no Mercado Nacional de Eletricidade, na Austrália. Nesse trabalho os autores concluem que o modelo SVR apresenta os melhores resultados nas previsões a curto prazo, juntamente com o MARS, superando o ARIMA.

No campo de previsão de demanda de água encontra-se o trabalho dos autores Can-

delieri et al. (2019), que utilizaram estratégias para o ajuste dos parâmetros do SVR, realizando tratamentos nas séries temporais estudadas, agrupando-as. Os autores concluem que ao realizar um agrupamento na fase preliminar e um ajuste global paralelizado dos parâmetros promovem um bom desempenho nas previsões a curto prazo.

Além dessas que são comuns às seções anteriores, é possível encontrar também esse algoritmo sendo usado na previsão do fluxo de turismo em uma cidade da China (ZHANG; PU, 2018). Nesse trabalho os autores tem como objetivo realizar o ajuste dos parâmetros do SVR, obtendo um bom resultado na previsão dos próximos 12 meses. Wan et al. (2016) utilizaram uma combinação de SVM com a técnica de Mínimos Quadrados para a previsão da cadeia de fornecimento de produtos eletrônicos no mercado chinês. A combinação citada é uma das técnicas utilizadas para lidar com um número limitado de amostras para a modelagem.

3.2.5 SGD

O SGD é um algoritmo que muitas vezes é utilizado combinado com outro método de regressão a fim de obter melhorias em seu desempenho. Porém também são encontrados trabalhos em que ele é utilizado separadamente para realizar previsão de demandas. Na pesquisa desenvolvida por Saputra et al. (2019), o SGD Regressor foi um dos métodos escolhidos para realizar a predição de energia para uma rede de veículos elétricos, alcançando resultados tão bons quanto os alcançados por MLP, SVR, entre outros. Rai et al. (2019) seguem uma abordagem similar, fazendo comparações entre os algoritmos mais utilizados na literatura de predição de demanda, e entre eles está o SGD. Nessa análise o objetivo é realizar a previsão de demanda para anúncios de comércio eletrônico.

Seguindo o contexto citado anteriormente, os demais exemplos encontrados, utilizam o SGD como um método de otimização dos resultados. No trabalho dos autores Marino, Amarasinghe e Manic (2016) o objetivo era prever o consumo e demanda de energia elétrica utilizando Redes Neurais Profundas combinadas com o SGD aumentando a velocidade de convergência nas taxas de erro. Aplicando esse método da mesma forma, Chang e Tsai (2017) utilizaram Redes Neurais Profundas com SGD como otimizador a fim de prever a demanda de turismo.

3.2.6 Decision Tree

O algoritmo de Árvore de Decisão também é utilizado em diversos contextos de previsão de demanda. Na área de produção de energia elétrica também são encontradas pesquisas utilizando Árvore de Decisão para a previsão de consumo. Hambali, Akinyemi e JYusuf (2016) propuseram a utilização de diversas variações de Árvores de Decisão para a previsão de consumo da população de uma cidade da Nigéria, a fim de otimizar a geração, transmissão e distribuição da energia elétrica. Já Çamurdan e Ganiz (2017) desenvolveram um trabalho em que foram utilizados diferentes algoritmos de aprendizagem de máquinas,

realizando a análise de dados coletados do mercado de eletricidade da Turquia, com dados no período de 2011 a 2016. Nessa análise, o algoritmo de Arvore de Decisão atinge um bom resultado, com 97% de acurácia.

Além desses também encontram-se trabalhos como o de Evers et al. (2018), que desenvolveram modelos de predição para a prever a demanda de supermercados na Holanda, auxiliando nas tomadas de decisões logísticas da empresa. Os autores relatam que a escolha dos algoritmos baseados em árvores deve-se ao fato da sua velocidade de execução e fácil interpretação. Ademais, relatam que a acurácia dos modelos é de aproximadamente 99%. Gkerekos, Lazakis e Papageorgiou (2018) também escolheram esse algoritmo como um dos que seriam utilizados para a previsão de consumo de combustível em embarcações. Os autores realizaram ajustes de parâmetros utilizando Random Search, conseguindo bons resultados com essa técnica. Killada (2017) em sua tese buscou realizar a predição no mercado de seguros em saúde, utilizando dados de 2014 a 2016 para treinamento e 2017 para análise dos resultados, concluindo que o modelo construído com base em uma árvore de decisão (Adaboost) possui o melhor desempenho.

3.2.7 Random Forest

O algoritmo de Random Forest também é bastante citado na literatura, em diversas aplicações. No seu trabalho Lahouar e Slama (2015) utilizam esse método para prever o consumo de energia elétrica a curto prazo, nas próximas 24h, a partir de dados de uma distribuidora de energia na Tunísia. Os autores atribuem a escolha dessa abordagem ao fato de as Florestas Aleatórias serem imunes às variações de parâmetros e possibilidade de utilização de validação cruzada interna. Nessa abordagem o modelo proposto atinge um erro médio de 2,3%. Johannesen, Kolhe e Goodwin (2019) também utilizaram o algoritmo de RF como uma das técnicas para prever o consumo de energia elétrica. Dentre os algoritmos analisados o RF obteve o melhor resultado para prever a carga em um curto prazo (30 minutos). Ahmad e Chen (2019) utiliza uma variação do Random Forest, conhecida como LSBoost, que é utilizado para reduzir o erro médio quadrático, demonstrando um bom desempenho a curto prazo.

Além do contexto anterior, Feng e Wang (2017) utilizaram o algoritmo de Floresta Aleatória para a previsão de demanda em um sistema de compartilhamento/aluguel de bicicletas, em Washington, DC. Nesse trabalho os pesquisadores concluem que a utilização de Random Forest alcança um bom desempenho, quando comparado a outros modelos lineares. No trabalho desenvolvido por Dyer (2018), é proposto um modelo utilizando RF para a previsão de demanda por moradias na cidade do Cabo. Dentre um dos resultados alcançados, o autor ressalta a importância da utilização desse algoritmo como a seleção das variáveis mais importantes para o modelo, melhorando o desempenho do modelo. Lauer, Legner e Henke (2019) desenvolveram um trabalho para realizar a predição da demanda em uma cadeia de suprimentos de semicondutores. Para isso utilizaram também

a metodologia CRISP-DM, utilizando dados reais de um fabricante de semicondutores.

3.2.8 XGBoost

Como foi dito no capítulo anterior, o algoritmo de XGBoost vem sendo amplamente utilizado em diversos contextos, quando o objetivo é resolver problemas de mineração de dados e aprendizagem de máquinas. No trabalho desenvolvido por Tiwari, Sabzehgar e Rasouli (2018), o cenário que houve a aplicação desse método foi a previsão de de irradiância solar, a fim de auxiliar na tomada de decisão quanto a demanda de energia elétrica através da geração solar. Nesse trabalho, foram realizadas etapas de pré-processamento dos dados, como a normalização e posteriormente o XGBoost para a modelagem do sistema de previsão. Para a validação do modelo desenvolvido foram utilizados dados reais da cidade de San Diego, Estados Unidos, sendo obtido um erro mínimo de 2,1% em um dos casos analisados. Seguindo no mesmo contexto, Fonteijn et al. (2019) propuseram a implementação prática de um sistema de previsão de carga elétrica a curto prazo, para áreas residenciais. Para isso, realizaram a comparação de modelos desenvolvidos a partir da aplicação de Regressão Linear, Random Forest e XGBoost. Após análises, concluíram que dentre os modelos avaliados, o XGBoost apresentou a melhor performance e informaram que a aplicação de técnicas de decomposição de séries temporais poderiam melhorar ainda mais os resultados alcançados.

Ma et al. (2018) apresentam um trabalho em que o objetivo principal é o desenvolvimento de modelos de previsão para prever a quantidade de futuros visitantes que um restaurante viria a ter. Para essa análise também foi realizado a comparação de modelos de regressão, dentre eles, Random Forest e XGBoost. Após a aplicação dos regressores desenvolvidos em dados coletados de sites de reservas de restaurante, ficou-se concluído que o XGBoost apresenta o melhor desempenho quando comparado individualmente aos demais modelos, atingindo um erro de 0,502 (RMSLE). Os autores ainda destacam que um dos motivos da escolha desse algoritmo é a sua eficiência, mesmo sendo aplicado em um contexto de baixo poder computacional. A fim de prever o consumo de um determinado serviço também, Vanichrujee et al. (2018) implementaram um modelo que buscava a previsão de demanda de táxis da cidade de Bangkok, Tailândia, dividindo a localidade em diversas áreas. Os dados foram coletados durante 4 meses, sendo gerado por mais de 5000 táxis. Para resolver o problema os autores utilizaram algoritmos como LSTM e o próprio XGBoost. Dentro de um dos contextos (áreas) analisados, o XGBoost se destaca diante dos outros algoritmos analisados. Já Chandel et al. (2019) desenvolveram um trabalho no qual propôs a aplicação desse algoritmo para realizar a previsão de vendas de uma empresa, nos Estados Unidos. Para isso utilizaram algoritmos de Regressão Linear e Árvore de Decisão também, constatando-se que a aplicação do XGBoost alcança melhores resultados do que os métodos anteriores.

Outra forma de utilização dessa técnica e que auxilia no processo de previsão de

demanda é a predição de preços de produtos. Dentro dessa ideia Gumus e Kiran (2017) realizam o desenvolvimento de um modelo que previa o preço do petróleo bruto, como um suporte a tomada de decisão de empresas que dependem da compra desse material, já que esse serve de matéria-prima para diversos contextos, como a geração de energia elétrica, por exemplo.

3.3 CONCLUSÃO

Após uma pesquisa, com o auxílio da ferramenta *Google Scholar*, utilizando como filtro o termo "Forecast Demand" e os algoritmos estudados, além de um período especificado dos últimos 10 anos (2010 a 2020), fica evidenciado as diversas abordagens, que diversos autores tem aplicado, quando o objetivo é realizar a previsão de demanda como um suporte à tomada de decisão. Dentre todos os algoritmos apresentados nas seções anteriores, dos considerados mais simples aos mais sofisticados, nota-se que todos eles são aplicados em diversos contextos e áreas, obtendo desempenhos muito bons quando aplicados com dados reais. Entre eles observamos que o algoritmo ARIMA, apesar das diversas técnicas mais atuais, continua sendo aplicado como uma das principais soluções para a previsão de Séries Temporais.

Dentre essas diferentes aplicações, destacam-se fatos como a necessidade do tratamento e transformações dos dados das Séries Temporais, além da seleção de atributos que melhor representem o problema a ser modelado. Quanto a seleção das variáveis que serão utilizados na modelagem, fica evidenciado também que em algumas pesquisas, a utilização de variáveis externas, como de contexto social, climático, demográficas etc, auxiliam na melhoria do desempenho do modelo. Além disso, os algoritmos de regressão analisados demonstraram ser aptos para resolver problemas que possuem amostras de análises de diferentes tamanhos.

No próximo capítulo, será apresentado a metodologia de experimentação realizada para implementação de modelos de previsão de demanda em uma empresa de bebidas, que atua com diferentes produtos, em diferentes unidades.

4 METODOLOGIA DE EXPERIMENTAÇÃO

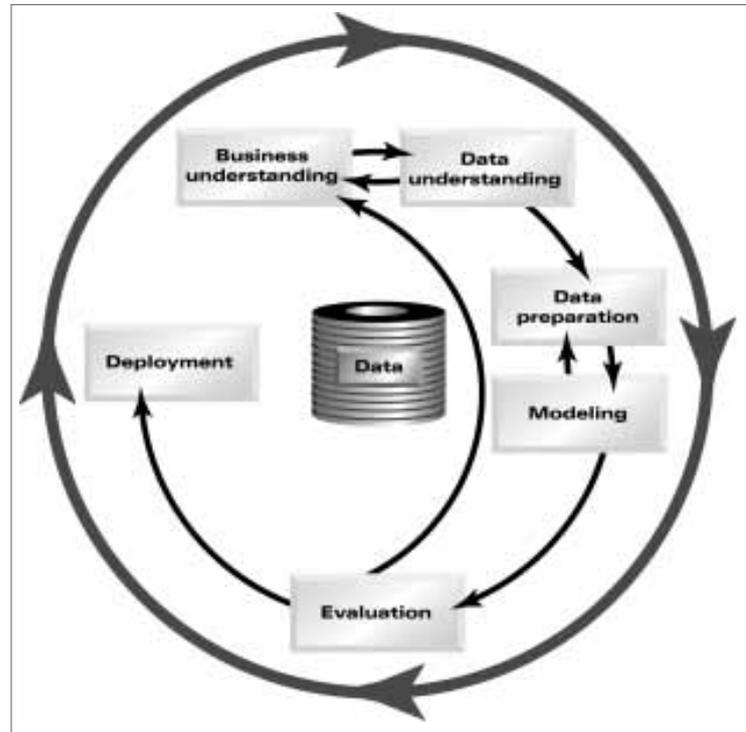
Neste capítulo será apresentado todas as análises investigativas e metodologia utilizada para o desenvolvimento dessa pesquisa. Como já foi citado, o presente trabalho tem como objetivo realizar a análise, predição e comparação de modelos preditivos de Séries Temporais, buscando realizar a Previsão de Demanda de uma empresa de bebidas multinacional. Para que esse objetivo seja alcançado, será utilizado como base a metodologia Cross Industry Standard for Data Mining (CRISP-DM). O CRISP-DM é uma metodologia considerada abrangente na área de mineração de dados. Um de seus pontos positivos é o fornecimento de um plano completo para a realização de um projeto de mineração de dados (SHEARER, 2000).

Em seu trabalho, Shearer (2000) cita que o CRISP-DM se divide em 6 fases:

1. **Entendimento dos negócios:** Considerada a fase mais importante, pois nela se concentra as atividades propostas para entender os objetivos do projeto, principalmente do ponto de vista de negócios. Logo, nessa fase inicial é definido o problema de mineração de dados, objetivos do negócio, determinação de metas etc.
2. **Entendimento dos dados:** Nessa fase é iniciada com a coleta dos dados iniciais do projeto. Dentro desse momento são descobertas as informações iniciais sobre os dados, sua descrição, exploração e verificação de qualidade.
3. **Preparação dos dados:** Essa fase é caracterizada por um conjunto de atividades que tem como objetivo construir o conjunto final de dados que serão utilizados na fase de modelagem. Dentre essas atividades são citadas: seleção de dados, limpeza dos dados, formatação dos dados etc.
4. **Modelagem:** Nessa etapa são selecionadas as técnicas de modelagem para a aplicá-las no conjunto de dados selecionados e tratados na fase anterior. Nesse momento, também os modelos são treinados, ajustados e avaliados.
5. **Avaliação:** Antes do modelo ser colocado em produção, é realizada uma avaliação mais detalhada dele, buscando garantir que ele atinja os objetivos do negócio. Nessa etapa são desenvolvidas atividades como avaliação dos resultados, revisão do processo e determinação das próximas etapas.
6. **Implantação:** Por fim, após a construção e avaliação do modelo, esse deve ser disponibilizado para poder ser aplicado. Nessa fase incluem-se atividades como o monitoramento do modelo e produções de relatórios.

Na figura 29 podemos observar as fases citadas anteriormente, seguindo o ciclo desse método.

Figura 16 – Fases do CRISP-DM



Fonte: (SHEARER, 2000)

Nesse trabalho será executada as 5 primeiras fases do processo. Essas serão apresentadas a seguir, sendo divididas em seções para um melhor entendimento da abordagem seguida na pesquisa. O aprofundamento da fase de avaliação será apresentado no capítulo seguinte, em que os modelos gerados serão comparados com modelos já existentes.

4.1 ENTENDIMENTO DO NEGÓCIO E ENTENDIMENTO DOS DADOS

Nessa primeira seção serão apresentadas, de forma conjunta, como a primeira e segunda etapa listadas no método CRISP-DM foram aplicadas no desenvolvimento da pesquisa apresentada nesse trabalho.

Inicialmente, na primeira etapa do trabalho, foram realizadas reuniões com um especialista da área de negócio, a fim de entender quais seriam os objetivos e pontos a serem alcançados nos modelos que viriam a ser desenvolvidos. Esse analista também possuía um conhecimento sobre os modelos que foram desenvolvidos anteriormente a desenvolvimento dessa pesquisa, além de todas as características da base de dados que viria a ser utilizada. Logo, ele apresentou quais critérios foram utilizados para a construção do conjunto de dados que seria aplicado aos modelos que já haviam sido desenvolvidos. Essa apresentação foi importante para que fosse seguido as mesmas etapas, a fim de poder ser realizado uma comparação entre esses modelos e os que foram desenvolvidos nesse trabalho. Nessas reuniões também foram ditas que alguns dos modelos já existentes foram construídos uti-

lizando modelos ARIMA, com o auxílio da ferramenta *auto.arima*, que seleciona o melhor modelo ARIMA, disponibilizado como um módulo da linguagem de programação *R*.

Com isso, nesse primeiro momento, foram apresentadas a base de dados da empresa de bebidas que seriam realizadas as previsões de demanda, junto com um arquivo que realizaria a identificação das categorias contidas nessa base de dados. O *dataset* analisado possui 10 atributos que serão apresentados na tabela a seguir, seguido de uma imagem que mostra esse conjunto de dados. A base de dados contém um total de 683981 instâncias distribuídas numa periodicidade semanal, partindo da primeira semana de 2015 até a semana 36 de 2017.

Tabela 1 – Atributos do Dataset

Atributo	Descrição
SEMANA	Semana no ano em que se obteve um volume de vendas
PROD	Produto
COD_MERCADO	Mercado. Ex.: supermercado, mercadinho, bar etc
UNIDADE	Unidade de venda regional
TAMANHO	Tamanho do produto. Ex.: 2L, 500ml etc
MRC	Marca do Produto
SBR	Sabor do produto
TPO	Tipo do Produto. Ex.: Refrigerante, cerveja, água etc
PREVISAO	Previsão de volume de vendas pela (modelo já desenvolvido)
VENDA_REAL	Volume de vendas real

Figura 17 – Amostra do Dataset

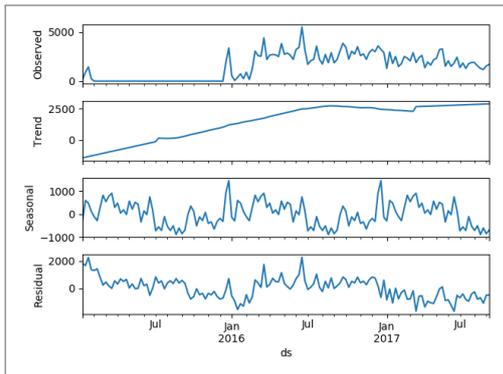
	SEMANA	PROD	COD_MERCADO	UNIDADE	TAMANHO	MRC	SBOR	TPO	PREVISAO	VENDA_REAL
0	201501	34	1	A	8	3	26.0	2	418	469.9400
1	201501	34	2	A	8	3	26.0	2	251	444.0717
2	201501	34	3	A	8	3	26.0	2	2230	4013.8911
3	201501	34	4	A	8	3	26.0	2	586	655.3291
4	201501	34	5	A	8	3	26.0	2	1126	474.2513
5	201501	34	6	A	8	3	26.0	2	254	302.8742
6	201501	38	1	A	9	4	8.0	3	396	774.8209
7	201501	38	3	A	9	4	8.0	3	0	0.0000
8	201501	38	4	A	9	4	8.0	3	1908	3338.4246
9	201501	38	5	A	9	4	8.0	3	429	1325.8280
10	201501	38	6	A	9	4	8.0	3	520	1181.3599

Ainda na reunião, ficou-se definido que os modelos a serem desenvolvidos contemplariam apenas uma marca e uma das unidades (marca = 10 e unidade = 'A') e que o conjunto de testes seriam as 6 últimas semanas do conjunto de dados, já que os modelos existentes haviam sido desenvolvidos dessa forma. Diante disso os produtos escolhidos

para ser realizado a modelagem estão identificados pelas categorias 67, 61, 78, 79 e 82, que nesse trabalho passarão a ser rotulados como Produto 1, 2, 3, 4 e 5, respectivamente. Após esse filtro, cada produto possui uma amostra de 140 instâncias, na mesma periodicidade.

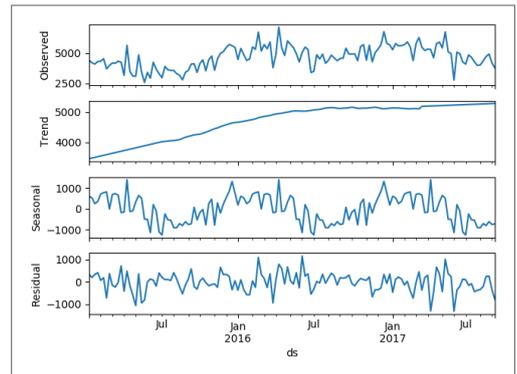
Posteriormente com os dados coletados, foram realizadas análises com auxílio de gráficos para observar o comportamento das séries temporais, realizando a sua decomposição em seus componentes. As imagens a seguir mostrarão o resultado dessas análises.

Figura 18 – Decomposição da Série Temporal - Produto 1



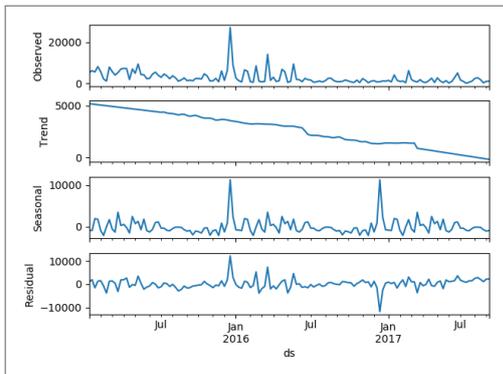
Fonte: Autor

Figura 19 – Decomposição da Série Temporal - Produto 2



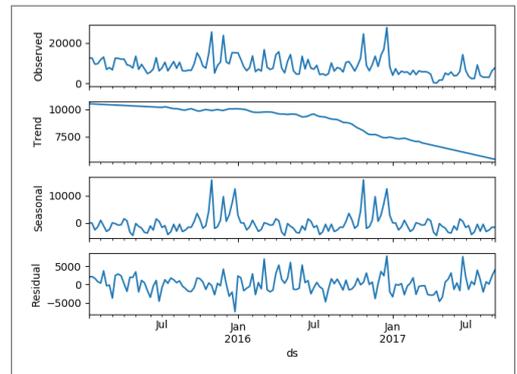
Fonte: Autor

Figura 20 – Decomposição da Série Temporal - Produto 3



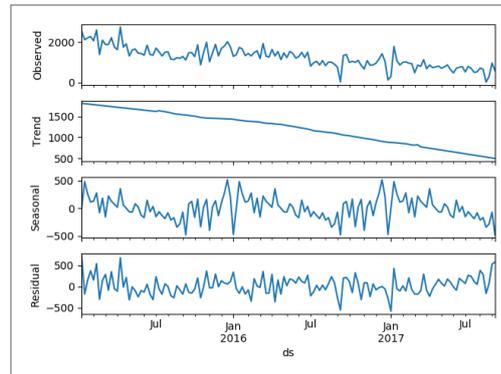
Fonte: Autor

Figura 21 – Decomposição da Série Temporal - Produto 4



Fonte: Autor

Figura 22 – Decomposição da Série Temporal - Produto 5



Fonte: Autor

Diante da observação da decomposição das Séries Temporais dos produtos, é possível concluir que cada produto possui uma particularidade, com diferentes formas de apresentação de seus componentes, não existindo um padrão para os 5 produtos investigados. Após a análise da decomposição das séries temporais dos produtos em seus componentes de Trend, Sazonalidade e Ruído, foram analisados a estacionariedade dessas, a fim verificar a necessidade de possíveis transformações e tratamentos dos dados. Como foi dito no capítulo 2, a verificação da estacionariedade é importante, pois auxilia na definição do método de previsão que será empregado, além de que tornar uma série "estacionarizada" é uma das técnicas utilizadas para contribuir com melhorias no desempenho dos regressores. Para essa avaliação, será utilizado o teste estatístico de Dickey-Fuller aumentado ou ADF. Nesse método é utilizado um método chamado teste de raiz unitária ou *unit root tests*, que tem um importante papel na análise de séries temporais não estacionárias, já que determina a intensidade que uma série temporal é definida pelo componente Trend. Para isso, é utilizado um teste de hipótese onde a hipótese nula é que a série temporal é representada por uma raiz unitária, sendo considerada não estacionária (FULLER, 1976) (REISEN et al., 2017). Na tabela a seguir, serão apresentados os resultados do teste ADF para os 5 produtos estudados.

Tabela 2 – Resultados ADF test

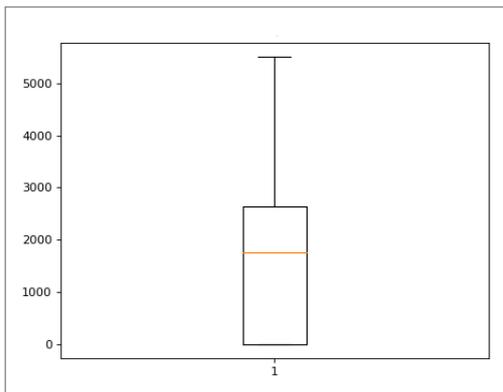
Produto	Estatística ADF	p-value
Produto 1	-1,52	0,52
Produto 2	-2,17	0,22
Produto 3	-2,15	0,22
Produto 4	-3,06	0,03
Produto 5	-1,73	0,41

Para interpretar os resultados apresentados na tabela 2 é necessário saber que os

valores dos nível de significância 5% é -2,88. Ciente desses valores podemos concluir que os produtos 1, 2, 3 e 5 podem ser considerados Séries Temporais não estacionárias, enquanto que o produto 4 seria uma Série Temporal estacionária. Tal conclusão é obtida já que os valores *p-value* abaixo de 0,05 significam a rejeição da hipótese nula. Além disso, valores menores que o nível de significância analisado, também implicam que a hipótese nula pode ser rejeitada.

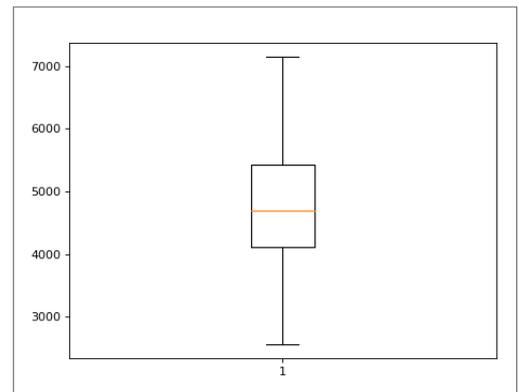
Ademais, foram feitas observações com gráficos *boxplot*, a fim de identificar possíveis outliers nos dados que seriam utilizados para o treinamento do modelo. A seguir pode-se observar os gráficos gerados nessa análise.

Figura 23 – Boxplot - Produto 1



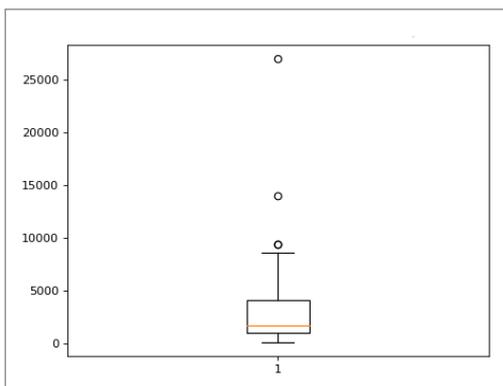
Fonte: Autor

Figura 24 – Boxplot - Produto 2



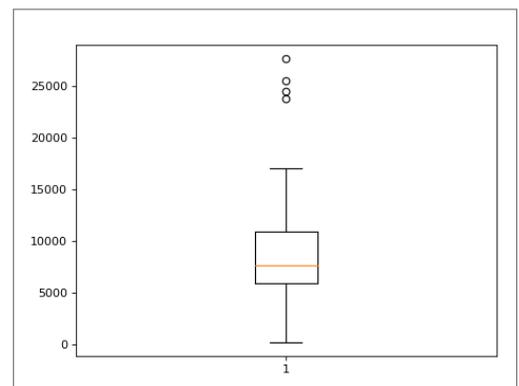
Fonte: Autor

Figura 25 – Boxplot - Produto 3



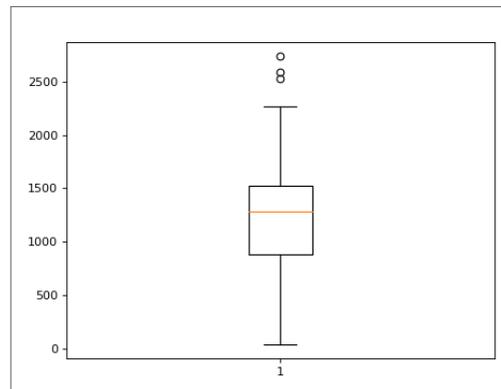
Fonte: Autor

Figura 26 – Boxplot - Produto 4



Fonte: Autor

Figura 27 – Boxplot - Produto 5



Fonte: Autor

Analisando os gráficos acima, podemos observar que as séries dos produtos 1 e 2 (Figuras 23 e 24) não possuem *outliers* (valores discrepantes), enquanto que os produtos 3, 4 e 5 (Figuras 25, 26 e 27) apresentam tais pontos. Tais informações são importantes para a fase seguinte, que é a preparação dos dados para o treinamento dos modelos de regressão. A utilização de valores que não condizem com a realidade estatisticamente podem prejudicar o desenvolvimento de um bom modelo.

Para a construção dos gráficos apresentados acima, foram utilizados módulos disponíveis na linguagem de programação *Python*. Para a análise dos componentes das Séries Temporais e identificação da estacionariedade foi utilizado o módulo *Statsmodels*, que é uma biblioteca para análise estatística e econométrica (SEABOLD; PERKTOLD, 2010). Já para a construção dos gráficos Boxplots, foi utilizado o módulo *Matplotlib* (HUNTER, 2007).

4.2 PREPARAÇÃO DOS DADOS

Nessa seção serão apresentadas todas as técnicas utilizadas no passo anterior à modelagem dos dados, com o auxílio dos métodos de regressão. Nesse estágio, foi utilizada a linguagem de programação *Python*, mais uma vez, com o auxílio do módulo *Pandas*. O *Pandas* é uma ferramenta avançada aplicada em conjuntos de dados estruturados, fornecendo métodos capazes de manipular e analisar esses dados (MCKINNEY, 2011). Com essa ferramenta foi possível realizar o filtro nas bases de dados de acordo com as unidades, produto e marcas da série temporal, que foram definidas na etapa anterior.

Após esses filtros iniciais, executou-se a etapa de pré-processamento, a fim de se criar as variáveis ou atributos que serão utilizados na etapa de modelagem. O primeiro campo a passar por transformações em sua estrutura foi o de "SEMANA", que representa a semana no ano em que se foi registrado um volume de vendas. Para que seja possível extrair mais informações desse campo, foi feita uma transformação, informando uma data correspondente àquela semana. A data escolhida na conversão foi o último dia da semana

processada. Na imagem a seguir, poderemos observar essa etapa de transformação dos dados.

Figura 28 – Exemplo da transformação do campo "Semana" em "Data"

	semana	data	y
0	201501	2015-01-10	2532.2998
1	201502	2015-01-17	2110.0384
2	201503	2015-01-24	2202.6063
3	201504	2015-01-31	2268.5453
4	201505	2015-02-07	2061.8526
5	201506	2015-02-14	2588.0941
6	201507	2015-02-21	1387.2489
7	201508	2015-02-28	2088.4817
8	201509	2015-03-07	1875.4489
9	201510	2015-03-14	1877.9848
10	201511	2015-03-21	2210.2147

Fonte: Autor

Após esse novo campo ser gerado ("data"), foi possível a construção de novos atributos, que melhor representariam aquela instância em um modelo de regressão. A partir dele novas variáveis puderam ser extraídas, através do seu processamento, que se executou a partir da divisão dos componentes contidos em uma data. A escolha dessas novas variáveis buscavam a melhor granularidade possível dos dados, aumentando a incidência de padrões dos dados, a fim de que os modelos pudessem identificar o comportamento dos valores estudados. Diante desse objetivo foram criados os seguintes campos: *semana_mes*, que representaria qual semana no mês aquele volume foi coletado, *mes*, que informaria o mês em que a coleta foi realizada e *feriados*, uma variável binária que informaria se na semana daquela data existia um feriado ou não. Para que a informação *feriado* fosse coletada, foi criado um dicionário com o auxílio da linguagem de programação *Python*, com um dos principais feriados do calendário brasileiro.

Em seguida, com esses novos campos gerados foi aplicado o método conhecido como One-Hot Encode. Essa técnica consiste na representação de dados categóricos, que não possuem um relacionamento ordinal, em uma nova variável binária, também conhecidas como *dummy* (BROWNLEE, 2017). Logo, informações como mês passou a contar com 12 novas variáveis, sendo preenchida com "1" caso aquela data pertencesse a aquele mês, ou "0" caso contrário, por exemplo. Com isso, os valores dos campos deixaram de apresentar um caráter categórico (numérico) e passaram a ser representados por atributos binários. Essa técnica foi aplicada nos campos citados anteriormente (*semana_mes* e *mes*). Nos

modelos desenvolvidos, os atributos utilizados foram as variáveis resultantes dessa etapa de pré-processamento. Esse exemplo pode ser observado na imagem a seguir.

Figura 29 – Exemplo da transformação One Hot Encode

semana	data	mes	x0_1.0	x0_2.0	x0_3.0	x0_4.0	x0_5.0	x0_6.0	x0_7.0	x0_8.0	x0_9.0	x0_10.0	x0_11.0	x0_12.0
201501	2015-01-10	1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201502	2015-01-17	1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201503	2015-01-24	1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201504	2015-01-31	1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201505	2015-02-07	2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201506	2015-02-14	2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201507	2015-02-21	2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201508	2015-02-28	2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201509	2015-03-07	3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201510	2015-03-14	3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
201511	2015-03-21	3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Fonte: Autor

Após a criação das variáveis, foram realizadas transformações nos valores do atributo "VENDA_REAL", que será o alvo da predição nos modelos desenvolvidos. Como se pode observar nos gráficos gerados na seção anterior, os intervalos da produção dos produtos variam bastante no tempo, com valores máximos e mínimos muito amplos. Diante disso, foi feita uma transformação nessas informações, a fim de diminuir essa diferença. O método escolhido para esse propósito foi a Normalização, que modifica o intervalo dos dados, deixando-os entre 0 e 1. A equação 4.1, mostra a fórmula utilizada para essa conversão.

$$Norm(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4.1)$$

Além da normalização, também foram removidos os valores considerados discrepantes (outliers), que também foram identificados na etapa de análise de dados, com o auxílio dos gráficos boxplots. Por fim, foram também retirados os valores de sazonalidade da séries temporais, identificados na análise de decomposição das séries. Posteriormente, esses valores seriam adicionados à serie, no momento da predição dos modelos.

Com todas essas transformações executadas, informações externas passaram a ser coletadas a fim de verificar a influência de variáveis temporais externas ao contexto do conjunto de dados coletado, procurando melhoria de resultados como os citados no Capítulo 3. Baseando-se nos trabalhos citados anteriormente e com o auxílio do especialista consultado, foram utilizadas variáveis climáticas, que poderiam influenciar no consumo do produto demandado e variáveis econômicas, que também podem interferir na produção. Para as variáveis climáticas, foram consideradas as seguintes informações: temperatura média, umidade média, insolação média, temperatura máxima e temperatura mínima. Esses dados foram obtidos a partir do portal do governo, Instituto Nacional de Meteorologia

(INMET)¹. Inicialmente os dados foram coletados de acordo com a região, representado pela variável "UNIDADE", que informa a qual cidade aquela informação pertence e posteriormente filtrados para a data correspondente, gerando um arquivo *.csv*. Já os dados econômicos escolhidos foram a Taxa de Juros Selic e Índice Nacional de Preços ao Consumidor Amplo (IPCA), pois são taxas utilizadas como instrumento da política monetária do país, simbolizando como encontra-se o momento econômico naquele período (BRASIL, 2020). Esses dados também foram coletados de um portal do governo, do Instituto de Pesquisa Econômica Aplicada (IPEA)². Desse portal foi feito o download de um arquivo *.csv*, que foi filtrado, obtendo apenas as informações referentes às datas presentes no conjunto de dados analisado. Para essas variáveis, foram utilizados seus valores contínuos. Finalizada essas preparações, deu-se início a etapa de modelagem, que será descrita na próxima seção desse capítulo.

4.3 MODELAGEM DOS DADOS

Esta seção terá o objetivo de apresentar os algoritmos escolhidos para a realização da modelagem do problema. Esses já foram apresentados e listados nos capítulos 2 e 3, sendo exposto toda a sua base teórica com vantagens e desvantagens e também trabalhos que os utilizaram com o mesmo objetivo de prever demanda, em diferentes contextos. Logo, os algoritmos que serão utilizados para a modelagem do problema são: Linear Regression, SVR, MLP, Decision Tree, SGD, Random Forest e XGBoost Regressor. Como foi apresentado no Capítulo 2, cada algoritmo escolhido possui uma particularidade na forma como aborda a solução da modelagem de dados. O mais simples como o Linear Regression, com a aplicação de funções de regressão simples ou múltipla, o SVR que realiza transformações nos dados através da aplicação de uma função Kernel, o MLP com a utilização de Redes Neurais, o SGD com a atualização de vetores gradientes alinhados com processos estocásticos durante o treinamento e por fim métodos que utilizam a estrutura de árvores para solucionar os problemas, desde a aplicação de Decision Tree, na sua forma simples, um conjunto de Árvores com o Random Forest e um conjunto de árvores otimizadas com o XGBoost. Além disso, esses são os algoritmos comumente utilizados e citados na solução de Previsão de Demanda, como foi evidenciado no Capítulo 3. Diante disso, tais algoritmos foram escolhidos para serem analisados para solução do problema proposto nesse trabalho.

Da mesma forma que aconteceu nas etapas anteriores, nessa também será utilizada a linguagem de programação *Python*. Para se atingir o objetivo de usar algoritmos de aprendizagem de máquinas será utilizada a biblioteca Scikit-learn. O Scikit-learn é uma biblioteca que engloba um grande conjunto de algoritmos de Machine Learning considerados de alto nível, tanto para resolver problemas supervisionados, como não superviso-

¹ <http://www.inmet.gov.br/portal/>

² <http://www.ipeadata.gov.br/Default.aspx>

nados. Seus autores destacam que uma de suas vantagens é a facilidade de uso, já que se concentra em levar valores tanto para especialistas, como não especialistas (PEDREGOSA et al., 2011).

Perante aos dados devidamente preparados e algoritmos de regressão selecionados, iniciou-se a fase de ajustes de parâmetros para os modelos a serem desenvolvidos. Para essa finalidade foi utilizada a técnica Random Search, que consiste na definição de um intervalo de valores para os parâmetros que serão ajustados, sendo selecionados de forma aleatória. Após essas combinações aleatórias, observa-se qual delas apresentou o melhor desempenho, sendo essa a escolhida para o modelo. Uma das vantagens desse método, quando comparado com a conhecida técnica Grid Search, é a sua eficiência para a otimização dos parâmetros, pois na escolha os hiperparâmetros nem sempre apresentam o mesmo grau de importância, ou seja, existem situações que a alteração de um parâmetro apresenta uma sensibilidade maior do que outros. Portanto, o Random Search aborda um intervalo maior de possibilidades, aumentando a chance de encontrar o melhor desempenho (BERGSTRÄ; BENGIO, 2012).

Em paralelo à técnica de Random Search, também foi utilizada a técnica de Cross-validation, que tem o objetivo de auxiliar na avaliação dos modelos ajustados, dividindo os dados em partes destinadas ao treinamento, teste e validação do algoritmo. Para isso define-se o número de iterações em que os dados serão divididos, conhecido como folds, onde em cada iteração o algoritmo será treinado com um conjunto de treino, teste e validação diferente (BISHOP, 2006) (DUDA; HART; STORK, 2012).

Nas tabelas a seguir, serão apresentadas os parâmetros ajustados em cada algoritmo. Nesses ajustes o número de folds definido é igual a 10 e número de iterações no Random Search igual a 1000.

Tabela 3 – Parâmetros ajustados - Linear Regression

Parâmetro	Intervalo ou Valores
fit_intercept	True ou False
normalize	True ou False
copy_X	True ou False
n_jobs	1 a 10

Tabela 4 – Parâmetros ajustados - SVR

Parâmetro	Intervalo ou Valores	Distribuição
kernel	rbf, linear, poly, sigmoid	Não Possui
degree	1 a 10	Não possui
coef0	0 a 1	Uniforme
tols	10^{-4} a 10^{-2}	Uniforme
C	0 a 50	Uniforme
epsilon	10^{-3} a 10^{-1}	Uniforme
shrinking	True ou False	Não possui
max_iter	100 a 500	Não possui

Tabela 5 – Parâmetros ajustados - SGD

Parâmetro	Intervalo ou Valores	Distribuição
epsilon	10^{-3} a 10^{-1}	Uniforme
tols	10^{-4} a 10^{-2}	Uniforme
loss	squared_loss, huber, epsilon_insensitive, squared_epsilon_insensitive	Não possui
penalty	none, l1, l2, elasticnet	Não possui
alpha	10^{-4} a 10^{-3}	Uniforme
l1_ratio	0 a 1	Uniforme
fit_intercept	True ou False	Não possui
shuffle	True ou False	Não possui
learning_rate	constant, optimal, invscaling	Não possui
n_iter_no_change	10 a 100	Não possui

Tabela 6 – Parâmetros ajustados - Decision Tree

Parâmetro	Intervalo ou Valores
criterion	mse, friedman_mse, mae
splitter	best ou random
max_depth	1 a 10
min_samples_split	2 a 10
min_samples_leaf	1 a 10

Tabela 7 – Parâmetros ajustados - MLP

Parâmetro	Intervalo ou Valores	Distribuição
activation	identity, logistic, tanh, relu	Não possui
solver	lbfgs, sgd, adam	Não possui
alpha	0 a 1	Uniforme
learning_rate	constant, invscaling, adaptive	Não possui
batch_size	10 a 100	Não possui
learning_rate_init	0 a 1	Uniforme
hidden_layer_sizes	2 a 100	Não possui

Tabela 8 – Parâmetros ajustados - Random Forest

Parâmetro	Intervalo ou Valores
n_estimators	10 a 300
max_depth	1 a 10
min_samples_split	2 a 10
min_samples_leaf	1 a 10

Tabela 9 – Parâmetros ajustados - XGBoost

Parâmetro	Intervalo ou Valores	Distribuição
max_depth	3 a 7	Não possui
learning_rate	0.1 a 0.3	Uniforme
n_estimators	100 a 1000	Uniforme
booster	gbtree, gblinear, dart	Não possui
gamma	0 a 1	Uniforme
min_child_weight	1 a 10	Não possui
subsample	0 a 1	Não possui
colsample_bytree	0 a 1	Uniforme
reg_alpha	0 a 1	Uniforme
reg_lambda	0 a 1	Uniforme
scale_pos_weight	0 a 1	Uniforme
importance_type	gain, weight, cover, total_gain, total_cover	Não possui

A escolha dos intervalos dos parâmetros numéricos baseou-se no tamanho da amostra disponível para a modelagem. Já nos categóricos, todas as opções disponíveis foram utilizadas, a fim de se avaliar possíveis melhorias nos resultados. Com os parâmetros e intervalos definidos foi seguida a estratégia de realizar a modelagem seguindo 3 abordagens diferentes. Na primeira abordagem os modelos seriam treinados com o valor de y na sua forma sem alterações ou transformações. Na segunda, seria aplicado a transformação de normalização e dessazonalização (subtração do componente de sazonalidade da série) do y . Na terceira e última, foram adicionadas as variáveis externas aos modelos.

Diante dessas três abordagens, foram desenvolvidos modelos com diferentes combinações de atributos, que foram construídos e selecionados na seção anterior. A escolha dos atributos foi realizada de forma iterativa, buscando os melhores resultados. Ou seja, inicialmente foi utilizado um conjunto de variáveis e posteriormente foram sendo adicionadas novos atributos a fim de verificar o comportamento do modelo. Na tabela a seguir podemos ver os modelos que foram desenvolvidos.

Tabela 10 – Modelos e atributos

Modelo	Atributos
1	Semana mes
2	Semana mes, mes
3	Semana mes, mes, feriados
4	Semana mes, feriados
5	Semana mes, juros selic
6	Semana mes, temperatura max, temperatura minima
7	Semana mes, inflacao ipca
8	Semana mes, temperatura média, umidade média, insolação média

Após as combinações realizadas, como é possível notar na Tabela 10, foram desenvolvidos 8 modelos diferentes para cada algoritmo de regressão, com as 3 abordagens diferentes, citadas anteriormente. Observa-se também que os modelos 5, 6, 7 e 8 utilizam as variáveis externas citadas anteriormente, como resultado da última abordagem investigativa do trabalho. A etapa de Avaliação, que apresentará os resultados alcançados pelos modelos e a comparação com os já existentes serão expostos no próximo capítulo dessa dissertação.

4.4 AVALIAÇÃO

Nesta penúltima etapa do CRISP-DM, as avaliações dos modelos desenvolvidos serão realizadas com o auxílio das métricas MAPE, MAE, RMSE e RMSLE. Todas essas são métricas de erro, que tem o objetivo de obter um resumo da distribuição dos erros, comumente calculando uma média. Nas equações a seguir, podemos observar como são calculadas todas essas métricas.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \frac{y_t - \hat{y}_t}{y_t} \right| \quad (4.2)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (4.3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (4.4)$$

$$RMSLE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\log_e(y_t + 1) - \log_e(\hat{y}_t + 1))^2} \quad (4.5)$$

Nas equações acima, os valores representados por y_t são os reais, enquanto que os \hat{y}_t são os previstos pelos modelos. A equação 4.2, que calcula a métrica MAPE, apresenta o erro de forma percentual, sendo considerado independente da escala. Já as equações

4.3, 4.4 e 4.5 são consideradas métricas que dependem diretamente da escala dos dados, ajustando-se de acordo com a necessidade da avaliação, penalizando mais ou menos a diferença do erro. Ou seja, a métrica MAE por ser uma análise da média absoluta do erro, as diferenças da previsão para o valor real são ponderadas igualmente. No caso do RMSE, ao obter a raiz quadrada dos erros quadráticos médios, tende a penalizar as pequenas diferenças e também destaca valores discrepantes. Já o RMSLE tende a não penalizar as grandes diferenças devido ao uso de funções logarítmicas tanto na previsão, quanto no valor real.

5 EXPERIMENTOS E ANÁLISES DE RESULTADOS

Neste capítulo será apresentado as análises dos resultados alcançados pelos modelos construídos nessa dissertação. Como foi dito anteriormente, nesse trabalho os modelos desenvolvidos serão comparados a modelos já existentes, que realizaram a previsão de demanda dos produtos analisados. Esses modelos foram desenvolvidos por duas empresas distintas, que serão identificadas como Empresa A e Empresa B. A Empresa A é especialista em *Data Science* e utilizou modelos ARIMA, com o apoio da ferramenta *auto.arima* para realizar a realização da predição da produção. Já a Empresa B é a própria produtora de bebidas, que desenvolveu um modelo interno para realizar a previsão das demandas. Nas tabelas a seguir poderemos observar os valores alcançados por esses modelos.

Tabela 11 – Resultados Empresa A

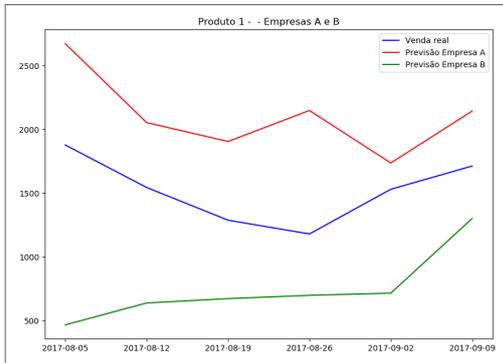
Produto	MAPE	MAE	RMSE	RMSLE
1	40,66	588,20	637,87	0,36
2	8,05	342,17	383,01	0,08
3	82,72	948,04	1024,69	0,74
4	27,04	1337,8	1598,54	0,32
5	413,98	255,13	324,40	1,34

Tabela 12 – Resultados Empresa B

Produto	MAPE	MAE	RMSE	RMSLE
1	49,94	773,70	718,73	0,82
2	50,62	2174,85	2497,89	0,99
3	197,13	1537,30	1712,98	1,89
4	94,9	3598,21	5141,81	0,85
5	133,68	329,94	431,18	1,32

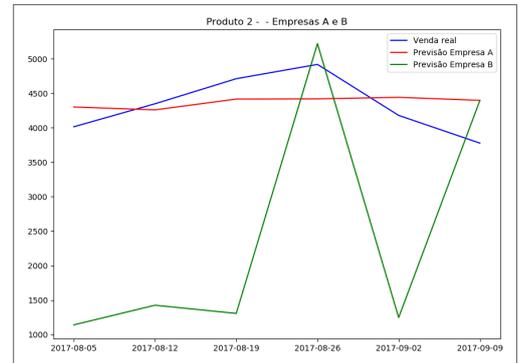
Nas imagens a seguir, serão apresentadas as análises através de gráficos.

Figura 30 – Previsões - Produto 1



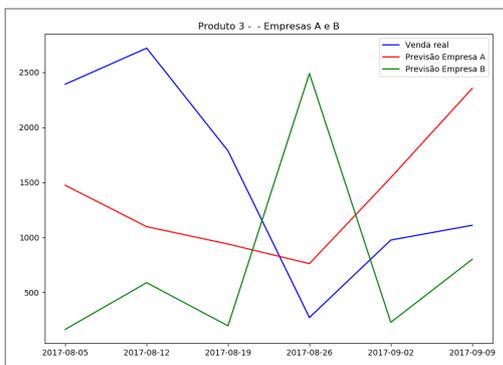
Fonte: Autor

Figura 31 – Previsões - Produto 2



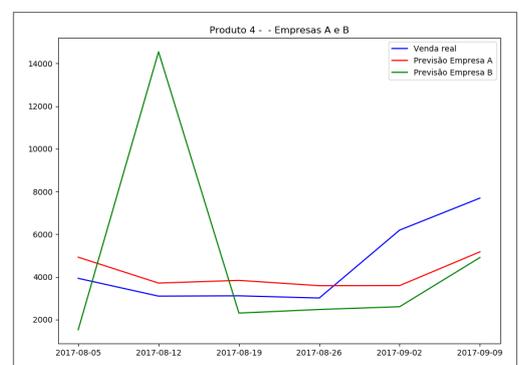
Fonte: Autor

Figura 32 – Previsões - Produto 3



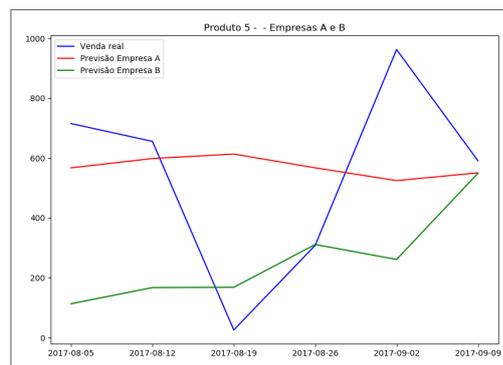
Fonte: Autor

Figura 33 – Previsões - Produto 4



Fonte: Autor

Figura 34 – Previsões - Produto 5



Fonte: Autor

Nas imagens acima, é observado em azul o valor real produzido, enquanto que a linha em vermelho simboliza a previsão realizada pela empresa A e em verde a previsão da Empresa B para as seis últimas semanas do conjunto de dados. Através dos gráficos é possível se analisar o comportamento do conjunto de testes dos produtos, complementando a análise com as métricas presentes nas tabelas 11 e 12.

Conhecido os resultados alcançados pelos modelos ARIMA e modelos internos desenvolvidos anteriormente, a seguir serão apresentados os valores obtidos pelos modelos construídos nesse trabalho. Nas seções a seguir serão mostrados os melhores resultados atingidos, separados por produtos e abordagens propiciando uma melhor análise e comparação dos resultados.

5.1 ABORDAGEM INICIAL

Seguindo a primeira abordagem, foram utilizados os valores das previsões sem as aplicações de métodos de transformação, com o intuito de observar se os métodos de regressão obteriam resultados eficientes sem a necessidade de aplicações de outras técnicas. Os melhores resultados para cada produto poderão ser observados nas tabelas a seguir. Tal abordagem será identificada como "Inicial". Para essa abordagem, foram utilizados os Modelos 1, 2, 3 e 4, apresentados no Capítulo anterior.

Tabela 13 – Resultados Produto 1 - Abordagem Inicial

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	40,66	588,20	637,87	0,36
Empresa B	-	49,94	773,70	718,73	0,82
4	XGBoost	11,85	170,85	184,85	0,12
2	MLP	13,72	195,10	239,73	0,16
1	XGBoost	14,68	205,70	246,86	0,17
2	XGBoost	15,55	215,35	256,63	0,17
1	MLP	16,21	268,11	328,12	0,22

Para a Produto 1, podemos observar que logo de início os resultados alcançados pelos modelos foram superiores aos desenvolvidos anteriormente. O resultado que se destaca é o alcançado pelo Modelo 4, com a utilização do algoritmo XGBoost.

Tabela 14 – Resultados Produto 2 - Abordagem Inicial

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	8,05	342,17	383,01	0,08
Empresa B	-	50,62	2174,85	2497,89	0,99
4	SGD	8,19	349,48	415,63	0,09
4	XGBoost	10,1	415,53	473,39	0,11
2	MLP	10,6	433,31	505,29	0,12
1	XGBoost	10,8	440,51	519,21	0,12
4	MLP	10,92	444,73	527,61	0,12

No caso do Produto 2, os resultados alcançados na abordagem inicial não foram superiores aos desenvolvidos pela Empresa A, atingindo um resultado similar ao apresentado pelo ARIMA. Quando a comparação é realizada com os resultados da Empresa B, todos os resultados alcançados foram superiores. O algoritmo que se destacou nessa abordagem foi o SGD, com o modelo 4.

Tabela 15 – Resultados Produto 3 - Abordagem Inicial

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	82,72	948,04	1024,69	0,74
Empresa B	-	197,13	1537,3	1712,98	1,89
2	MLP	71,96	941,26	1147,95	0,90
3	MLP	72,56	883,61	1091,33	0,85
4	SGD	86,92	953,96	1174,39	0,97
4	SVR	87,8	955,84	1158,63	0,95
3	SGD	90,23	908,75	1112,84	0,92

Os resultados alcançados na modelagem do Produto 3 também foi possível alcançar resultados superiores quando comparados às empresas A e B. O algoritmo MLP, com os modelos 2 e 3, destacou-se nos resultados alcançados.

Tabela 16 – Resultados Produto 4 - Abordagem Inicial

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	27,04	1337,8	1598,54	0,32
Empresa B	-	94,9	3598,21	5141,81	0,85
1	MLP	40,08	1685,53	1812,36	0,38
3	MLP	40,3	1690,18	1813,81	0,39
2	SVR	40,41	1689,38	1805,66	0,38
3	SVR	41,58	1651,04	1705,66	0,38
1	SVR	41,82	1702,28	1791,28	0,39

Para o Produto 4, a abordagem inicial apresentou resultados superiores apenas quando comparados com os modelos desenvolvidos pela Empresa B. Com destaque para os algoritmos de MLP e SVR. Quando comparados a Empresa A, os resultados alcançados foram inferiores.

Tabela 17 – Resultados Produto 5 - Abordagem Inicial

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	413,98	255,13	324,4	1,34
Empresa B	-	133,68	329,94	431,18	1,32
3	SGD	391,54	349,4	379,07	1,38
1	MLP	571,79	330,02	414,72	1,47
2	SGD	600,72	347,9	436,99	1,5
3	SVR	606,83	372,04	453,11	1,51
2	SVR	618,91	370,69	458,36	1,52

Para o Produto 5, a abordagem inicial apresentou resultados superiores apenas quando comparados com a Empresa A, com destaque para o algoritmo SGD com o modelo 3. Quando comparado aos resultados alcançados pela Empresa B, os resultados foram inferiores.

5.2 ABORDAGEM TRANSFORMAÇÃO DOS DADOS

Na segunda abordagem foram realizadas a aplicação dos métodos de Normalização e Dessazonalização nos valores de previsão. Como foi dito no capítulo anterior, a primeira técnica tem o objetivo de retirar a diferença de escala presente nos valores da Série Temporal. Logo, sua aplicação tem o objetivo de analisar possíveis melhorias na capacidade de predição dos algoritmos, quando essa diferença é retirada. Na segunda técnica, é realizada a retirada do componente de Sazonalidade da Série Temporal, que devido ao seu comportamento cíclico pode ser re-adicionado à série no final das previsões. Da mesma forma, sua retirada tem o objetivo investigativo de observação de possíveis melhorias no desempenho dos algoritmos de regressão com a presença ou ausência de tal componente. Com a utilização dessas técnicas, os melhores resultados alcançados serão apresentados nas tabelas a seguir.

Tabela 18 – Resultados Produto 1 - Transformação dos Dados

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	40,66	588,20	637,87	0,36
Empresa B	-	49,94	773,70	718,73	0,82
4	XGBoost	11,85	170,85	184,85	0,12
1 Normalização	MLP	13,62	201,20	239,33	0,16
2 Normalização	MLP	16,86	256,02	284,65	0,19
4 Normalização	Decision Tree	17,43	253,80	302,58	0,20
2 Normalização	Random Forest	18,11	264,25	304,50	0,20
3 Normalização	Random Forest	18,12	263,71	304,65	0,20
1 Normalização	SVR	19,12	289,16	313,58	0,21
4 Dessazonalização	XGBoost	20,49	335,44	447,12	0,35
2 Dessazonalização	XGBoost	22,76	362,11	435,60	0,31
1 Dessazonalização	XGBoost	23,04	338,08	395,50	0,26

Com a aplicação das técnicas de transformação dos dados, os resultados alcançados para o Produto 1 continuaram sendo superiores quando comparados aos modelos desenvolvidos pelas duas empresas. Destaque para a técnica de Normalização, com o modelo 1 e o algoritmo de MLP. Apesar de obterem resultados superiores, o destaque com o melhor resultado alcançado ainda foi o obtido na abordagem inicial, como podemos observar na tabela 18.

Tabela 19 – Resultados Produto 2 - Transformação dos Dados

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	8,05	342,17	383,01	0,08
Empresa B	-	50,62	2174,85	2497,89	0,99
4	SGD	8,19	349,48	415,63	0,09
2 Normalização	SGD	7,53	330,30	431,31	0,10
4 Normalização	XGBoost	7,56	334,62	407,19	0,09
3 Normalização	SGD	8,34	363,03	482,33	0,11
4 Normalização	SVR	9,83	415,46	458,34	0,11
3 Dessazonalização	SVR	8,48	379,94	427,22	0,10
4 Dessazonalização	Decision Tree	9,00	411,80	559,55	0,13
4 Dessazonalização	Random Forest	9,21	420,39	580,98	0,13
4 Dessazonalização	SGD	9,34	416,53	522,93	0,12

No caso do Produto 2, alguns dos resultados alcançados foram superiores aos desenvolvidos pelas duas empresas. Os resultados alcançados pelos algoritmos SGD, com o Modelo 2 e técnica de Normalização, e o XGBoost, com o modelo 4 e a técnica de Normalização, apresentaram destaque, melhorando os resultados alcançados na abordagem inicial.

Tabela 20 – Resultados Produto 3 - Transformação dos Dados

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	82,72	948,04	1024,69	0,74
Empresa B	-	197,13	1537,3	1712,98	1,89
2	MLP	71,96	941,26	1147,95	0,90
1 Normalização	MLP	70,95	1038,58	1251,31	1,03
3 Normalização	SGD	72,86	877,39	1096,79	0,86
4 Normalização	SGD	74,38	821,67	1049,40	0,83
1 Normalização	SGD	75,41	866,94	1073,74	0,85
2 Normalização	SGD	85,66	882,46	1073,98	0,88
1 Dessazonalização	Random Forest	95,57	666,67	808,40	0,77
2 Dessazonalização	Random Forest	95,68	654,36	798,87	0,77
4 Dessazonalização	Linear Regression	96,31	611,09	788,89	0,77
1 Dessazonalização	Linear Regression	96,34	612,07	789,07	0,77

Para o Produto 3, o comportamento foi similar ao apresentado no Produto 2. A aplicação das técnicas de transformação trouxeram resultados aos modelos desenvolvidos na etapa inicial. Na maioria dos casos, os resultados apresentados são superiores aos desenvolvidos pelas empresas A e B. Nessa etapa, destacou-se o algoritmo de MLP, com o modelo 1 de forma conjunta com a técnica de Normalização.

Tabela 21 – Resultados Produto 4 - Transformação dos Dados

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	27,04	1337,8	1598,54	0,32
Empresa B	-	94,9	3598,21	5141,81	0,85
1	MLP	40,08	1685,53	1812,36	0,38
4 Normalização	SGD	70,56	2405,82	2703,94	0,57
1 Normalização	SVR	73,06	2500,12	2850,32	0,59
1 Dessazonalização	MLP	53,56	2148,25	2725,11	0,53

Para o produto 4, o comportamento manteve-se similar ao observado na abordagem inicial. Os resultados permaneceram superiores quando comparados com a Empresa B.

Porém com a aplicação da transformação dos dados, não houve uma melhoria nos resultados, quando comparados com a abordagem anterior, destacando-se apenas os resultados apresentados na tabela 21.

Tabela 22 – Resultados Produto 5 - Transformação dos Dados

Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	413,98	255,13	324,4	1,34
Empresa B	-	133,68	329,94	431,18	1,32
3	SGD	391,54	349,4	379,07	1,38
3 Normalização	SVR	423,69	239,79	321,33	1,35
2 Normalização	SGD	439,86	342,98	384,14	1,50

Na previsão do Produto 5, a aplicação das técnicas de transformação dos dados não trouxeram melhorias para os resultados. Nessa abordagem, os dois melhores resultados alcançados estão presentes na tabela 22, porém nenhum deles é superior aos resultados apresentados pelas empresas A e B, nem ao alcançado na abordagem inicial.

5.3 ADIÇÃO DAS VARIÁVEIS EXTERNAS E RESULTADOS FINAIS

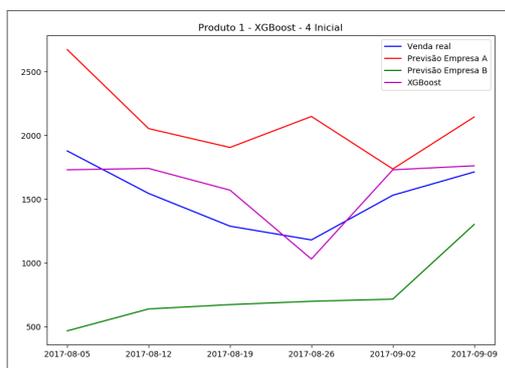
A terceira e última abordagem foi a adição das variáveis externas, referentes a dados climáticos e econômicos, aos experimentos realizados. Nessa etapa serão apresentados os 10 melhores modelos alcançados para cada produto, a fim de observar as possíveis melhorias trazidas pelas variáveis externas. Além disso serão apresentados gráficos, com a finalidade de melhor descrever as previsões. Assim sendo, será possível analisar quais abordagens e modelos conseguiram se destacar diante das diversas técnicas aplicadas.

Nas tabelas a seguir, serão apresentados entre parênteses o quanto de ganho ou perda de desempenho o modelo teve em relação aos desenvolvidos pela Empresa A e B. Na primeira posição dos parênteses será observado a comparação com o valor alcançado pela Empresa A, enquanto que na segunda os da Empresa B. Já nos gráficos, os valores em azul representarão os valores reais, em vermelho as previsões da Empresa A, em verde da Empresa B e os em roxo os valores preditos pelos modelos desenvolvidos nesse trabalho.

Tabela 23 – Melhores Resultados Produto 1

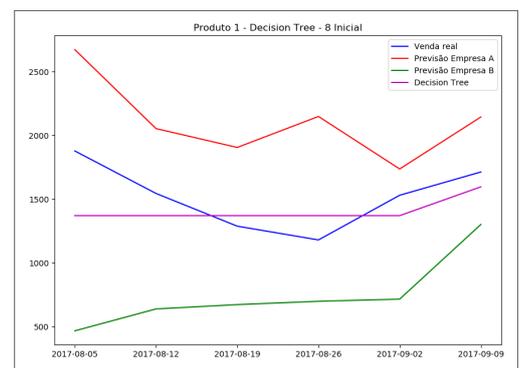
Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	40,66	588,20	637,87	0,36
Empresa B	-	49,94	773,70	718,73	0,82
4 Inicial	XGBoost	11,85 (-71%, -76%)	170,85 (-71%, -78%)	184,85 (-71%, -74%)	0,12 (-67%, -85%)
8 Inicial	Decision Tree	13,02 (-68%, -74%)	205,24 (-65%, -73%)	248,41 (-61%, -65%)	0,16 (-56%, -80%)
7 Normalização	MLP	13,52 (-67%, -73%)	200,27 (-66%, -74%)	237,81 (-63%, -67%)	0,16 (-56%, -80%)
1 Normalização	MLP	13,62 (-67%, -73%)	201,20 (-66%, -74%)	239,33 (-62%, -67%)	0,16 (-56%, -80%)
7 Normalização	SGD	13,7 (-66%, -73%)	196,26 (-67%, -75%)	247,23 (-61%, -66%)	0,17 (-53%, -79%)
2 Inicial	MLP	13,72 (-66%, -73%)	195,10 (-67%, -75%)	239,73 (-62%, -67%)	0,16 (56%, 80%)
5 Normalização	XGBoost	14,55 (-64%, -71%)	226,24 (-62%, -71%)	257,54 (-60%, -64%)	0,17 (-53%, -79%)
1 Inicial	XGBoost	14,68 (-64%, -71%)	205,70 (-65%, -73%)	246,86 (-61%, -66%)	0,17 (-53%, -79%)
5 Inicial	XGBoost	14,96 (-63%, -70%)	236,63 (-60%, -69%)	271,37 (-57%, -62%)	0,18 (-53%, -79%)
2 Inicial	XGBoost	15,55 (-62%, -69%)	215,35 (-63%, -72%)	256,63 (-60%, -64%)	0,17 (-50%, -78%)

Figura 35 – Previsão Modelo 4 Inicial (XGBoost) - Produto 1



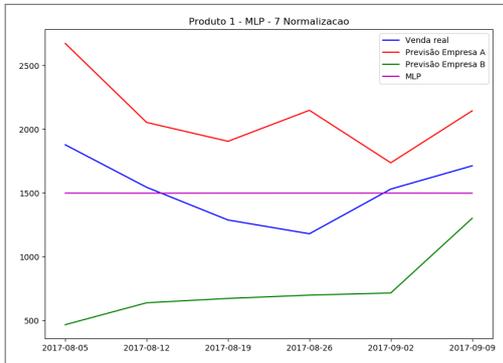
Fonte: Autor

Figura 36 – Previsão Modelo 8 Inicial (Decision Tree) - Produto 1



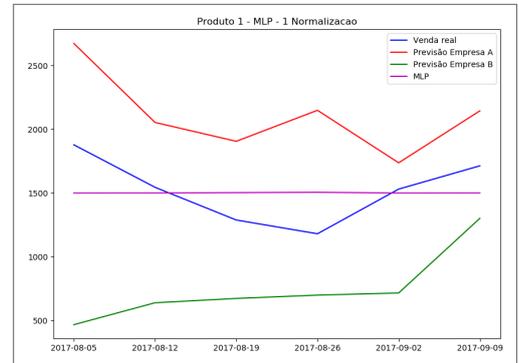
Fonte: Autor

Figura 37 – Previsão Modelo 7 Normalização (MLP) - Produto 1



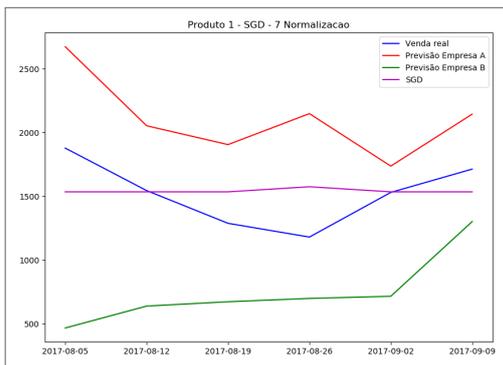
Fonte: Autor

Figura 38 – Previsão Modelo 1 Normalização (MLP) - Produto 1



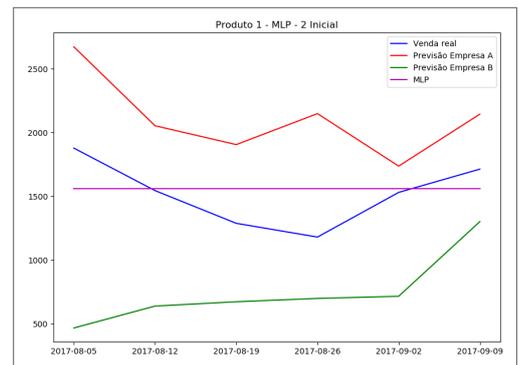
Fonte: Autor

Figura 39 – Previsão Modelo 7 Normalização (SGD) - Produto 1



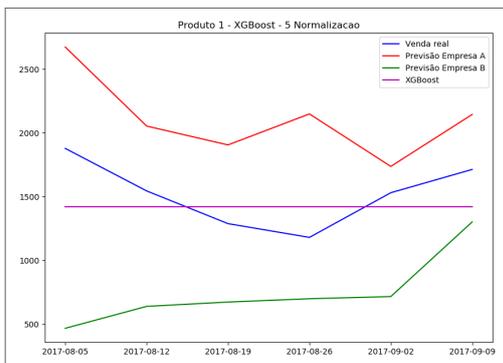
Fonte: Autor

Figura 40 – Previsão Modelo 2 Inicial (MLP) - Produto 1



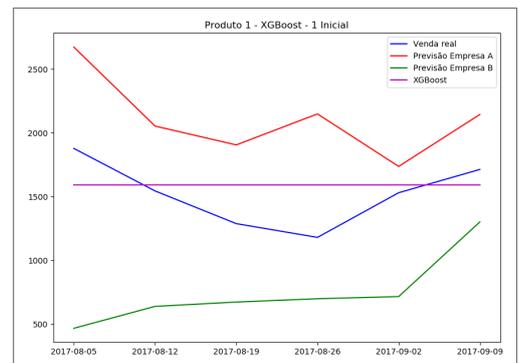
Fonte: Autor

Figura 41 – Previsão Modelo 5 Normalização (XGBoost) - Produto 1



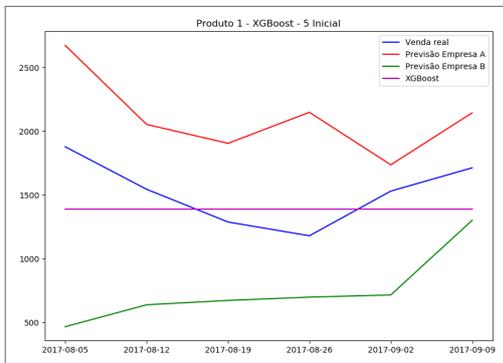
Fonte: Autor

Figura 42 – Previsão Modelo 1 Inicial (XGBoost) - Produto 1



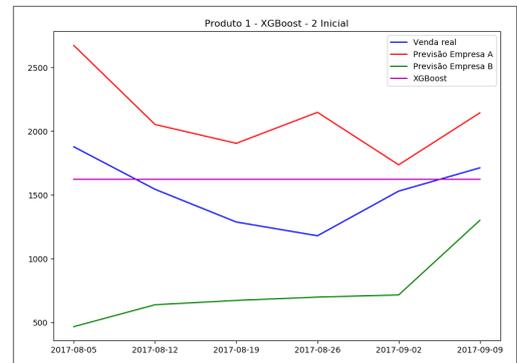
Fonte: Autor

Figura 43 – Previsão Modelo 5 Inicial (XGBoost) - Produto 1



Fonte: Autor

Figura 44 – Previsão Modelo 2 Inicial (XGBoost) - Produto 1



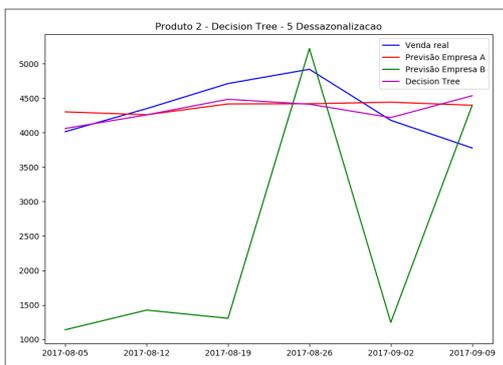
Fonte: Autor

Observando a tabela 23 é possível concluir que os modelos de regressão obtêm resultados eficientes, quando comparados com os promovidos pelos modelos desenvolvidos pelas duas empresas, para o Produto 1. As métricas de avaliação encontram-se com resultados superiores, com baixas taxas de erro nos melhores modelos de regressão desenvolvidos nesse trabalho. Dentre os algoritmos se destacam o de XGBoost, com o melhor resultado (modelo 4), acompanhado pelo de MLP, Random Forest, Decision Tree e SGD. Complementando essa análise, os gráficos demonstram a capacidade dos melhores regressores identificarem o comportamento das séries temporais, especialmente o XGBoost. Além disso, nota-se também que a técnica de Normalização contribuiu bastante para uma boa previsão dos modelos, aparecendo entre 4 dos 10 melhores modelos, enquanto que a técnica de Dessazonalização não aparece como um dos melhores resultados. Quando analisado a influência que as variáveis externas trazem a previsão, observa-se que os modelos que as contém também se destacam, como por exemplo o modelo 8 que apresenta o 2º melhor resultado alcançado nas previsões. Dentre os 10 melhores resultados, 5 apresentam a utilização de variáveis externas.

Tabela 24 – Melhores Resultados Produto 2

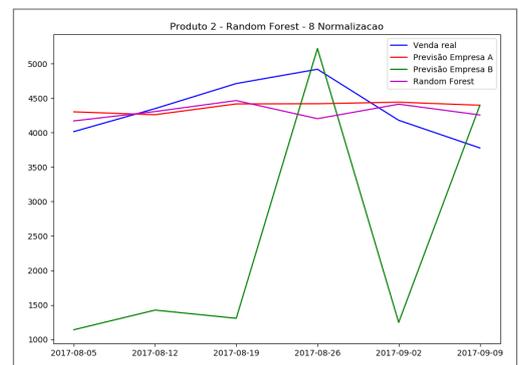
Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	8,05	342,17	383,01	0,08
Empresa B	-	50,62	2174,85	2497,89	0,99
5 Dessazonaliz.	Decision Tree	6,56 (-19%, -87%)	278,02 (-19%, -87%)	386,06 (+1%, -85%)	0,09 (+13%, -91%)
8 Normalização	Random Forest	7,15 (-11%, -86%)	312,21 (-9%, -86%)	383,81 (0%, -85%)	0,09 (+13%, -91%)
5 Inicial	MLP	7,52 (-7%, -85%)	334,62 (-2%, -85%)	412,67 (+8%, -83%)	0,09 (+13%, -91%)
2 Normalização	SGD	7,53 (-6%, -85%)	330,30 (-3%, -85%)	431,31 (+13%, -83%)	0,10 (+25%, -90%)
4 Normalização	XGBoost	7,56 (-6%, -85%)	334,62 (-2%, -85%)	407,19 (+6%, -84%)	0,09 (+13%, -91%)
8 Normalização	MLP	7,56 (-6%, -85%)	334,62 (-2%, -85%)	406,08 (+6%, -84%)	0,09 (+13%, -91%)
8 Normalização	XGBoost	7,58 (-6%, -85%)	334,62 (-2%, -85%)	404,17 (+6%, -84%)	0,09 (+13%, -91%)
8 Inicial	Random Forest	7,61 (-5%, -85%)	329,61 (-4%, -85%)	416,58 (+9%, -83%)	0,10 (+25%, -90%)
8 Normalização	SVR	7,83 (-3%, -85%)	339,21 (-1%, -84%)	394,07 (+3%, -84%)	0,09 (+13%, -91%)
8 Inicial	Decision Tree	7,85 (-2%, -84%)	336,24 (-2%, -85%)	392,63 (+3%, -84%)	0,09 (+13%, -91%)

Figura 45 – Previsão Modelo 5 Dessazonalização (Decision Tree) - Produto 2



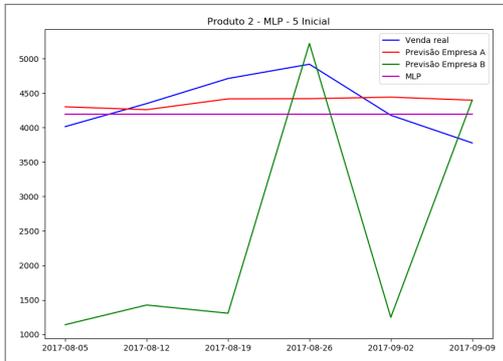
Fonte: Autor

Figura 46 – Previsão Modelo 8 Normalização (Random Forest) - Produto 2



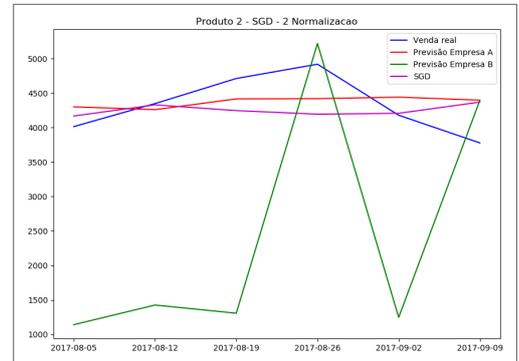
Fonte: Autor

Figura 47 – Previsão Modelo 5 Inicial (MLP) - Produto 2



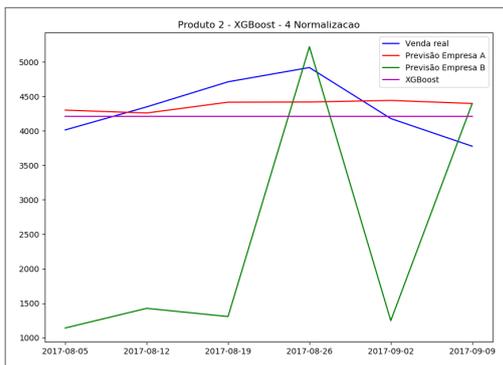
Fonte: Autor

Figura 48 – Previsão Modelo 2 Normalização (SGD) - Produto 2



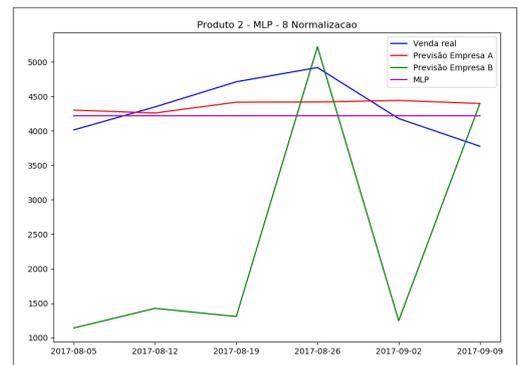
Fonte: Autor

Figura 49 – Previsão Modelo 4 Normalização (XGBoost) - Produto 2



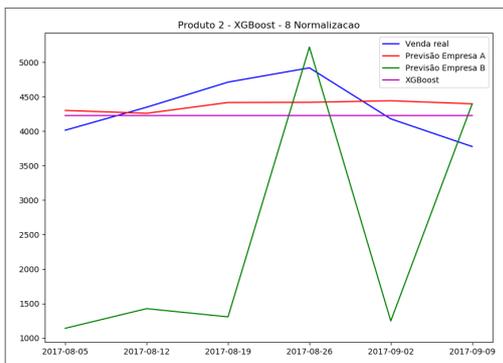
Fonte: Autor

Figura 50 – Previsão Modelo 8 Normalização (MLP) - Produto 2



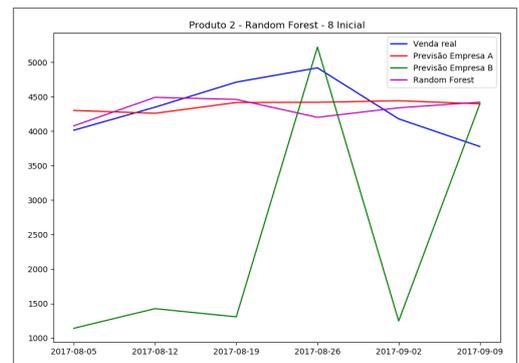
Fonte: Autor

Figura 51 – Previsão Modelo 8 Normalização (XGBoost) - Produto 2



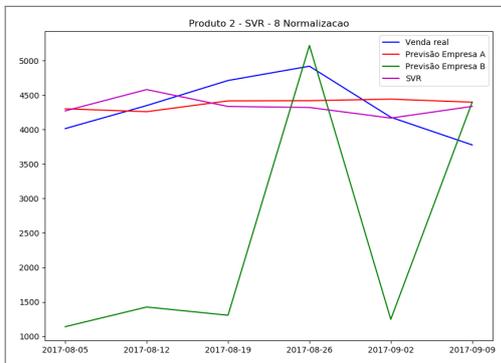
Fonte: Autor

Figura 52 – Previsão Modelo 8 Inicial (Random Forest) - Produto 2



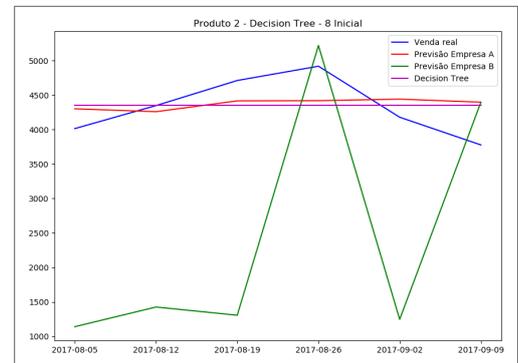
Fonte: Autor

Figura 53 – Previsão Modelo 8 Normalização (SVR) - Produto 2



Fonte: Autor

Figura 54 – Previsão Modelo 8 Inicial (Decision Tree) - Produto 2



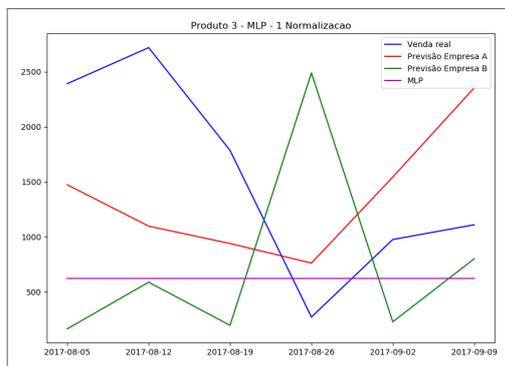
Fonte: Autor

Para o Produto 2, a tabela 24 mostra que os resultados alcançados pelos modelos propostos nesse trabalho também encontram-se com resultados superiores quando comparado aos desenvolvidos pela Empresa A e B. Tal observação é encontrada principalmente quando se comparam as métricas MAPE e MAE. Ao analisar as outras métricas, observa-se que essas estão próximas, porém minimamente acima dos modelos da Empresa A, revelando que em alguns pontos as previsões destoam um pouco em determinadas semanas do conjunto de teste. Esse comportamento pode ser observado nos gráficos. Nos modelos desenvolvidos para esse produto também é observado que as técnicas de transformação dos dados, com a normalização dos dados e retirada da sazonalidade, contribuíram para um melhor desempenho na previsão da demanda. Dentre os algoritmos o de Decision Tree apresenta o melhor resultado (modelo 5 + dessazonalização), acompanhado pelos de MLP, Random Forest, SGD e SVR. Como observado no Produto 1, as variáveis externas também se destacaram nos modelos desenvolvidos para o Produto 2. O Modelo 5, que conta com a utilização de variáveis econômicas, alcançou melhor resultado, enquanto que o Modelo 8, que conta com a aplicação de variáveis climáticas, aparece em 6 dos casos dos 10 melhores resultados.

Tabela 25 – Melhores Resultados Produto 3

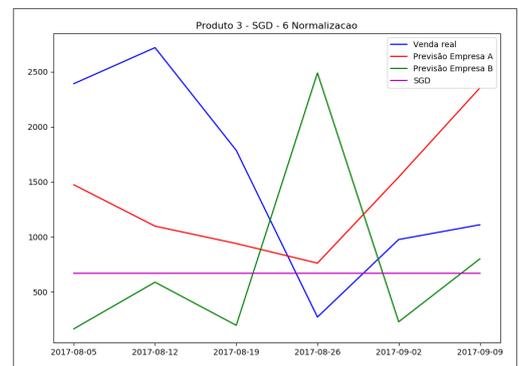
Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	82,72	948,04	1024,69	0,74
Empresa B	-	197,13	1537,3	1712,98	1,89
1 Normalização	MLP	70,95 (-14%,-64%)	1038,58 (+10%,-32%)	1251,31 (+22%,-27%)	1,03 (+39%,-46%)
6 Normalização	SGD	71,30 (-14%,-64%)	1005,53 (+6%,-35%)	1215,23 (+19%,-29%)	0,98 (+32%,-48%)
8 Inicial	MLP	71,76 (-13%,-64%)	961,09 (+1%,-37%)	1168,27 (+14%,-32%)	0,92 (+24%,-51%)
2 Inicial	MLP	71,96 (-13%,-63%)	941,26 (-1%,-39%)	1147,95 (+12%,-33%)	0,90 (+22%,-52%)
6 Inicial	MLP	72,31 (-13%,-63%)	908,08 (-4%,-41%)	1114,89 (+9%,-35%)	0,87 (+18%,-54%)
5 Inicial	MLP	72,55 (-12%,-63%)	885,23 (-7%,-42%)	1092,85 (+7%,-36%)	0,85 (+15%,-55%)
3 Inicial	MLP	72,56 (-12%,-63%)	883,61 (-7%,-43%)	1091,33 (+7%,-36%)	0,85 (+15%,-55%)
3 Normalização	SGD	72,86 (-12%,-63%)	877,39 (-7%,-43%)	1096,79 (+7%,-36%)	0,86 (+16%,-54%)
8 Inicial	SGD	73,98 (-11%,-62%)	834,36 (-12%,-46%)	1042,00 (+2%,-39%)	0,82 (+11%,-57%)
4 Normalização	SGD	74,38 (-10%,-62%)	821,67 (-13%,-47%)	1049,40 (+2%,-39%)	0,83 (+12%,-56%)

Figura 55 – Previsão Modelo 1 Normalização (MLP) - Produto 3



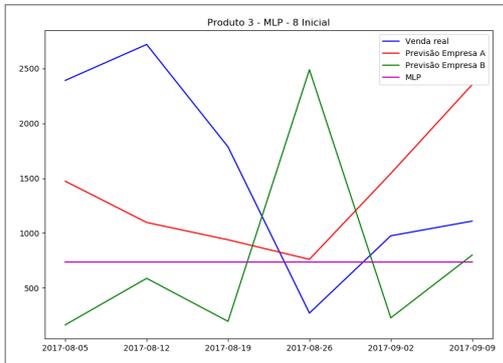
Fonte: Autor

Figura 56 – Previsão Modelo 6 Normalização (SGD) - Produto 3



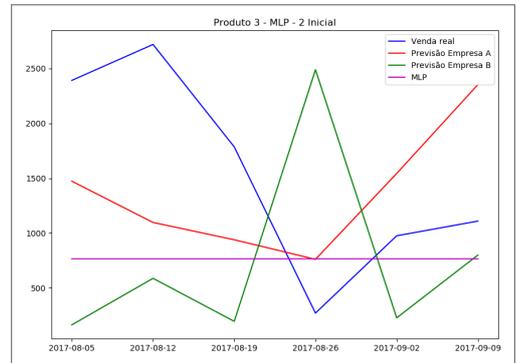
Fonte: Autor

Figura 57 – Previsão Modelo 8 Inicial (MLP) - Produto 3



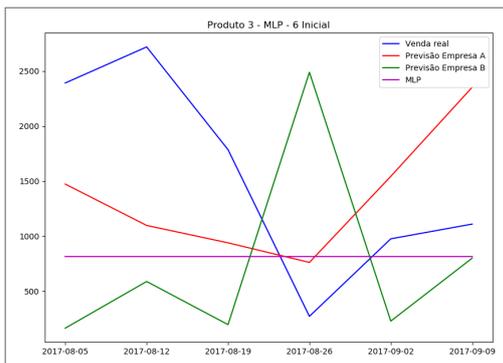
Fonte: Autor

Figura 58 – Previsão Modelo 2 Inicial (MLP) - Produto 3



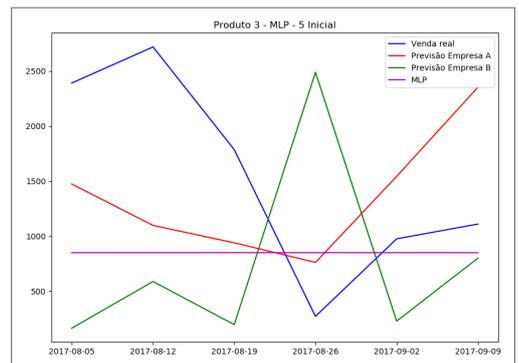
Fonte: Autor

Figura 59 – Previsão Modelo 6 Inicial (MLP) - Produto 3



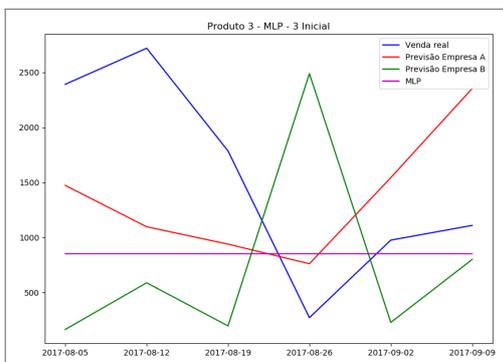
Fonte: Autor

Figura 60 – Previsão Modelo 5 Inicial (MLP) - Produto 3



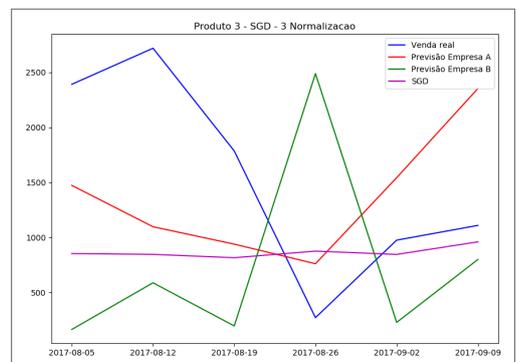
Fonte: Autor

Figura 61 – Previsão Modelo 3 Inicial (MLP) - Produto 3



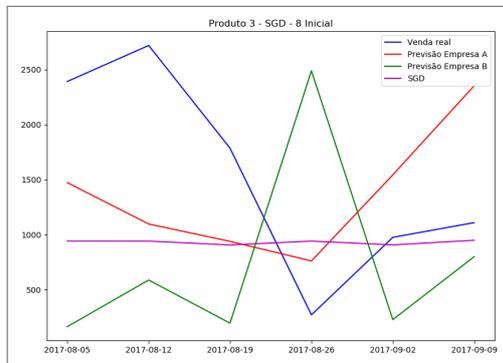
Fonte: Autor

Figura 62 – Previsão Modelo 3 Normalização (SGD) - Produto 3



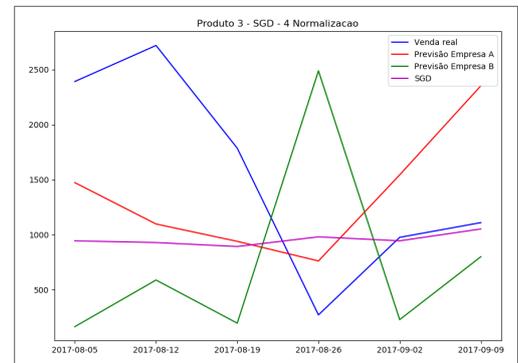
Fonte: Autor

Figura 63 – Previsão Modelo 8 Inicial (SGD) - Produto 3



Fonte: Autor

Figura 64 – Previsão Modelo 4 Normalização (SGD) - Produto 3



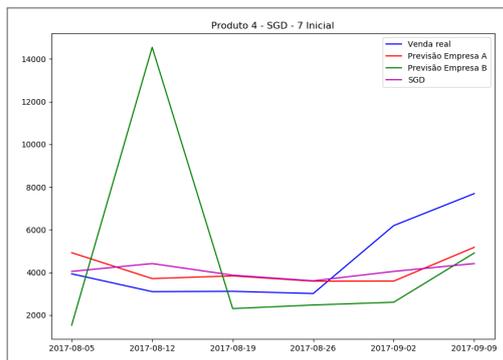
Fonte: Autor

No Produto 3, na tabela 25, encontra-se um comportamento similar ao observado nos produtos 1 e 2. Os modelos de regressão de forma geral tem apresentado resultados melhores quando comparado o percentual de erro, representado pelo MAPE. Quando analisados as demais métricas, como o MAE, também se encontram resultados melhores, porém algumas taxas de erro um pouco acima dos alcançados pelo modelo ARIMA. O RMSE e RMSLE apontam que em alguns deles, existem pontos específicos em que as previsões encontram-se discrepantes. Porém de forma geral, o percentuais de erro encontram-se abaixo dos obtidos no modelo da Empresa A e principalmente da Empresa B, em que todos os valores encontram-se acima dos alcançados nesse trabalho. Nesse produto, a técnica de Normalização se destacou quando aplicada com os algoritmos de MLP e SGD. A dessazonalização não apresentou resultados relevantes. Já as variáveis externas demonstraram ser um bom recurso para a melhoria dos resultados alcançados pelos modelos, com destaque para o modelo 6, que utiliza variáveis climáticas.

Tabela 26 – Melhores Resultados Produto 4

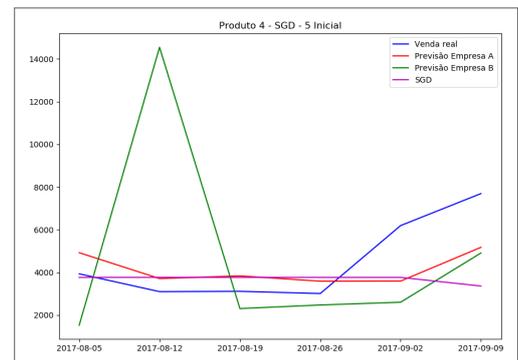
Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	27,04	1337,8	1598,54	0,32
Empresa B	-	94,9	3598,21	5141,81	0,85
7 Inicial	SGD	27,82 (+3%,-71%)	1368,73 (+2%,-62%)	1734,45 (+9%,-66%)	0,34 (+6%,-60%)
5 Inicial	SGD	27,92 (+3%,-71%)	1501,18 (+12%,-58%)	2087,81 (+31%,-59%)	0,42 (+31%,-51%)
7 Inicial	MLP	33,41 (+24%,-71%)	1544,42 (+15%,-57%)	1822,57 (+14%,-65%)	0,37 (+16%,-56%)
8 Normalização	SGD	34,57 (+28%,-64%)	1598,25 (+19%,-56%)	3486762,13 (+36%,-87%)	0,38 (+19%,-55%)
7 Inicial	SVR	34,86 (+29%,-63%)	1547,52 (+16%,-57%)	1867,29 (+17%,-64%)	0,36 (+13%,-58%)
5 Inicial	SVR	35,53 (+31%,-63%)	1644,78 (+23%,-54%)	1959,12 (+23%,-62%)	0,40 (+25%,-53%)
6 Inicial	SGD	39,42 (+46%,-58%)	1725,49 (+29%,-52%)	1915,22 (+20%,-63%)	0,40 (+25%,-53%)
1 Inicial	MLP	40,08 (+48%,-58%)	1685,52 (+26%,-53%)	1812,36 (+13%,-65%)	0,38 (+19%,-55%)
3 Inicial	MLP	40,30 (+49%,-58%)	1690,18 (+26%,-53%)	1813,81 (+13%,-65%)	0,39 (+22%,-54%)
2 Inicial	SVR	40,41 (+49%,-57%)	1689,37 (+26%,-53%)	1805,66 (+13%,-65%)	0,38 (+19%,-55%)

Figura 65 – Previsão Modelo 7 Inicial (SGD) - Produto 4



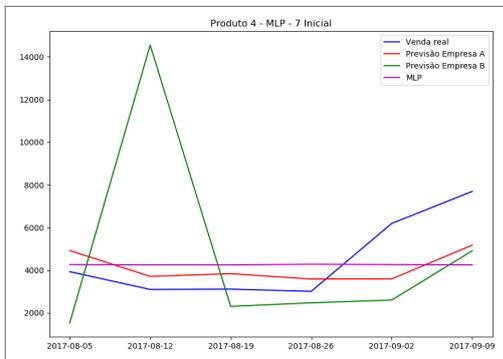
Fonte: Autor

Figura 66 – Previsão Modelo 5 Inicial (SGD) - Produto 4



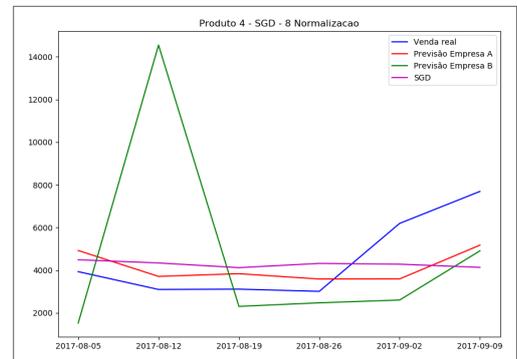
Fonte: Autor

Figura 67 – Previsão Modelo 7 Inicial (MLP) - Produto 4



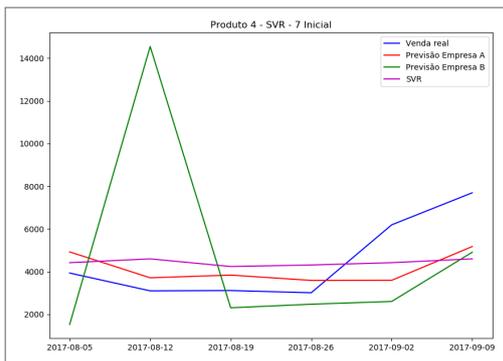
Fonte: Autor

Figura 68 – Previsão Modelo 8 Normalização (SGD) - Produto 4



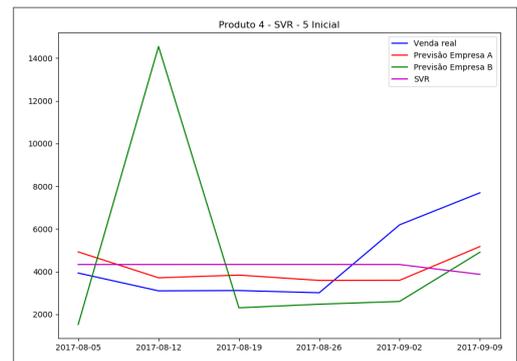
Fonte: Autor

Figura 69 – Previsão Modelo 7 Inicial (SVR) - Produto 4



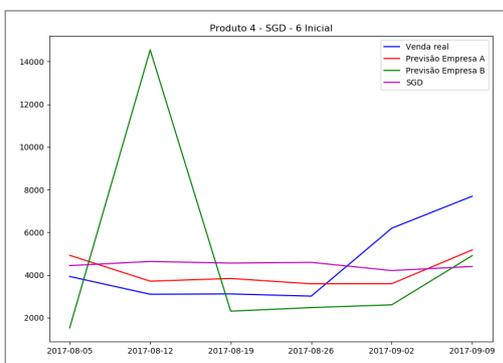
Fonte: Autor

Figura 70 – Previsão Modelo 5 Inicial (SVR) - Produto 4



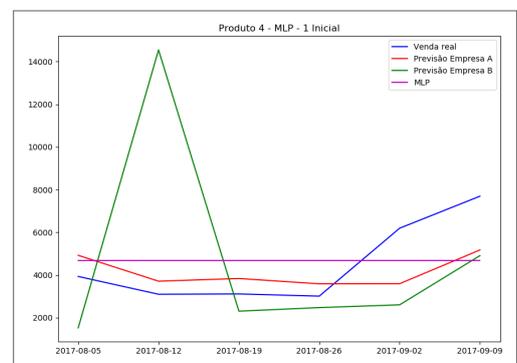
Fonte: Autor

Figura 71 – Previsão Modelo 6 Inicial (SGD) - Produto 4



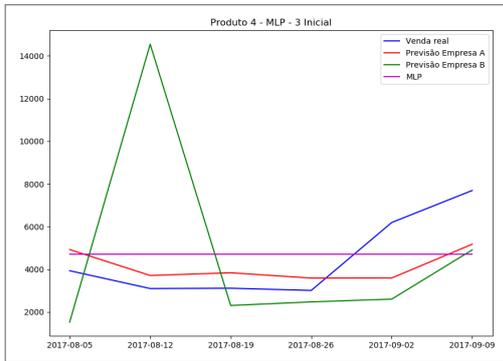
Fonte: Autor

Figura 72 – Previsão Modelo 1 Inicial (MLP) - Produto 4



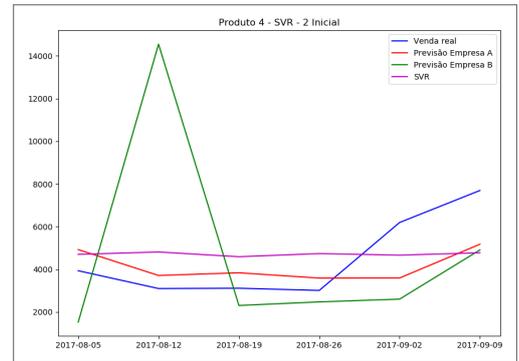
Fonte: Autor

Figura 73 – Previsão Modelo 3 Inicial (MLP) - Produto 4



Fonte: Autor

Figura 74 – Previsão Modelo 2 Inicial (SVR) - Produto 4



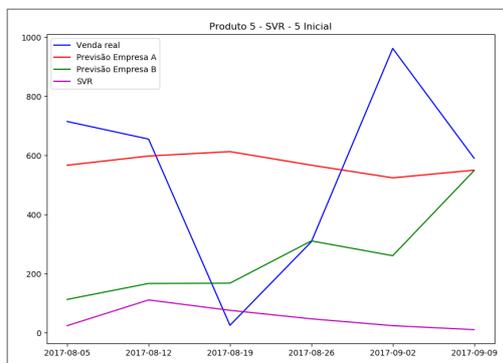
Fonte: Autor

No caso do Produto 4, observando a tabela 26, os resultados alcançados não foram tão bons quando comparados com os modelos de previsão da Empresa A. Dos modelos desenvolvidos nesse trabalho, apenas 2 apresentaram resultados melhores ou similares, ambos utilizando o algoritmo de SGD. Dentre os outros métodos que se destacaram, apesar de não alcançar resultados similares aos anteriores, estão o MLP e SVR. Porém, quando comparados com os resultados alcançados pelo modelo interno da Empresa B, todos os modelos desenvolvidos com métodos de regressão obtiveram resultados melhores. Para esse produto, as técnicas de transformação não demonstraram ser tão eficazes. Contudo, os modelos que utilizaram as variáveis externas alcançaram os melhores resultados entre os modelos desenvolvidos nessa pesquisa, destacando-se os modelos 7 e 5, que utilizam variáveis econômicas entre seus atributos.

Tabela 27 – Melhores Resultados Produto 5

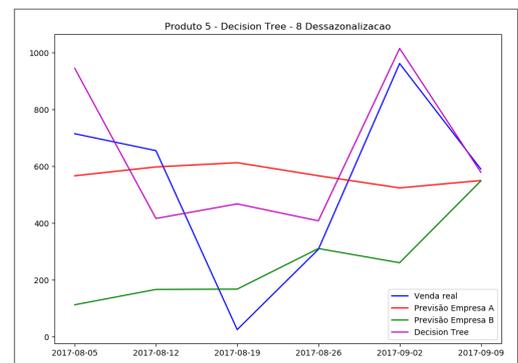
Modelo	Algoritmo	MAPE	MAE	RMSE	RMSLE
Empresa A	ARIMA	413,98	255,13	324,4	1,34
Empresa B	-	133,68	329,94	431,18	1,32
5 Inicial	SVR	110,0 (-73%, -18%)	510,92 (+100%, +55%)	586,17 (+81%, +36%)	2,82 (+110%, +114%)
8 Dessaz.	Decision Tree	309,11 (-25%, +131%)	179,48 (-30%, -46%)	230,65 (-29%, -47%)	1,20 (-10%, -9%)
5 Normaliz.	Decision Tree	363,97 (-12%, +172%)	263,97 (+3%, -20%)	311,75 (-4%, -28%)	1,29 (-4%, -2%)
5 Inicial	Decision Tree	363,97 (-12%, +172%)	263,97 (+3%, -20%)	311,75 (-4%, -28%)	1,29 (-4%, -2%)
7 Normaliz.	SVR	366,84 (-11%, +174%)	234,80 (-8%, -29%)	297,66 (-8%, -31%)	1,28 (-4%, -3%)
3 Inicial	SGD	391,54 (-5%, +193%)	349,03 (+37%, +6%)	379,07 (+17%, -12%)	1,38 (+3%, +5%)
7 Dessaz.	SGD	398,92 (-4%, +198%)	371,67 (+46%, +13%)	412,25 (+27%, -4%)	1,77 (+32%, +34%)
6 Inicial	SVR	399,01 (-4%, +198%)	313,89 (+23%, -5%)	367,60 (+13%, -15%)	1,36 (+1%, +3%)
7 Dessaz.	Decision Tree	420,35 (+2%, +214%)	436,68 (+71%, +32%)	462,85 (+43%, +7%)	1,67 (+25%, +27%)
3 Normaliz.	SVR	423,69 (+2%, +217%)	239,79 (-6%, -27%)	321,33 (-1%, -25%)	1,35 (+1%, +2%)

Figura 75 – Previsão Modelo 5 Inicial (SVR) - Produto 5



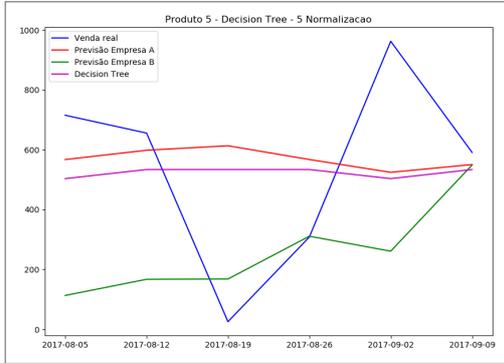
Fonte: Autor

Figura 76 – Previsão Modelo 8 Dessazonalização (Decision Tree) - Produto 5



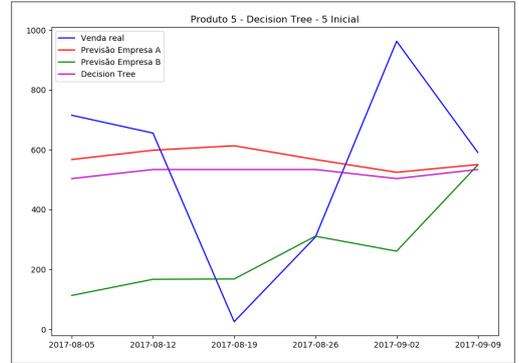
Fonte: Autor

Figura 77 – Previsão Modelo 5 Normalização (Decision Tree) - Produto 5



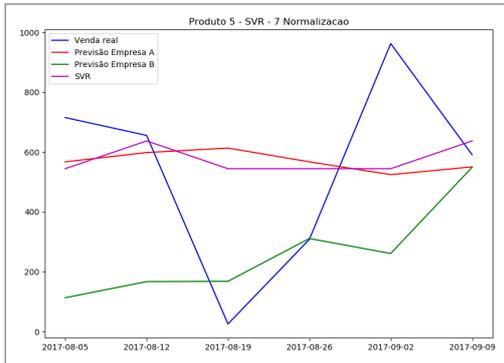
Fonte: Autor

Figura 78 – Previsão Modelo 5 Inicial (Decision Tree) - Produto 5



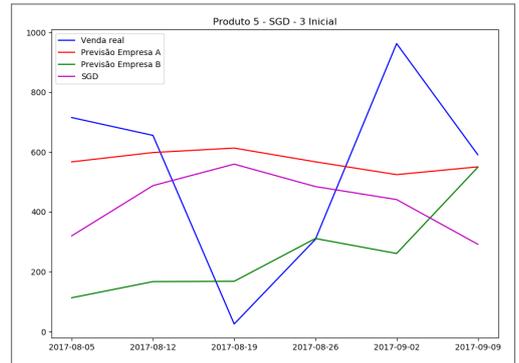
Fonte: Autor

Figura 79 – Previsão Modelo 7 Normalização (SVR) - Produto 5



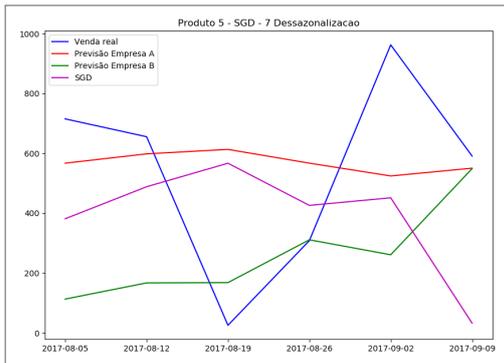
Fonte: Autor

Figura 80 – Previsão Modelo 3 Inicial (SGD) - Produto 5



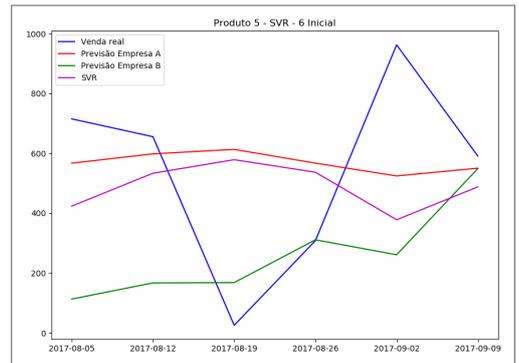
Fonte: Autor

Figura 81 – Previsão Modelo 7 Des-sazonalização (SGD) - Produto 5



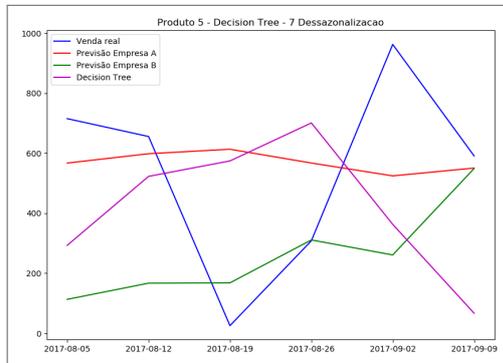
Fonte: Autor

Figura 82 – Previsão Modelo 6 Inicial (SVR) - Produto 5



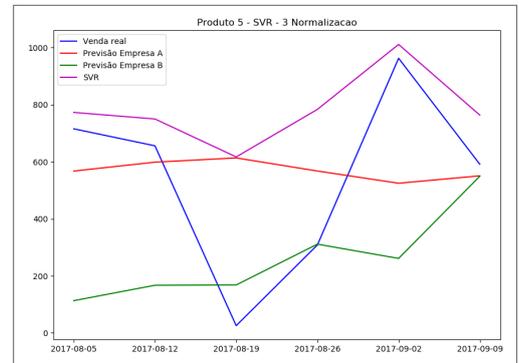
Fonte: Autor

Figura 83 – Previsão Modelo 7 Dessazonalização (Decision Tree) - Produto 5



Fonte: Autor

Figura 84 – Previsão Modelo 3 Normalização (SVR) - Produto 5



Fonte: Autor

Por fim, analisando a tabela 27, para o Produto 5 os modelos desenvolvidos conseguiram atingir resultados superiores, principalmente quando a comparação foi a métrica MAPE, estando bem abaixo dos modelos que utilizaram o método ARIMA, desenvolvido pela Empresa A. Quando comparado com a Empresa B, apenas alguns modelos conseguiram ter um desempenho melhor. Nesse quesito, o algoritmo SVR (modelo 5) apresentou o melhor resultado, porém ao analisar as outras métricas não demonstrou os mesmos índices. Quando analisado as outras métricas, o modelo que obteve o melhor o melhor desempenho foi o que utilizou Árvore de Decisão (modelo 8 + dessazonização). Nesse produto, também é possível observar que as técnicas de Normalização e Dessazonização se apresentaram como um importante fator para uma melhoria no desempenho dos modelos. Quando analisado a adição das variáveis externas, observa-se que sua utilização trouxe melhorias nos modelos, com o destaque para os dois melhores resultados alcançados. Além disso, observa-se também que entre os 10 melhores resultados apresentados na tabela, 8 deles contaram com a utilização de variáveis externas.

5.4 CONCLUSÕES

De forma geral, ao analisar os resultados dos modelos de regressão desenvolvidos nesse trabalho, observa-se que eles apresentam resultados superiores ou similares quando comparados com métodos que utilizaram ARIMA. Além disso, a utilização de variáveis externas mostrou ser uma abordagem que agrega bastante valor aos modelos. Dos modelos que se destacaram, listam-se os 5, 6, 7 e 8, que utilizaram variáveis climáticas e econômicas como entrada nos modelos. Dentre as abordagens, a inicial e a Normalização dos dados alcançaram os melhores resultados. A Dessazonização apresentou bons resultados apenas em casos bastante específicos, como nos melhores predições alcançadas para os Produtos 2 e 5.

Outro ponto observado é o desafio na modelagem do problema, devido a uma pequena amostra de dados, já que quando realizado os filtros apresentados no Capítulo 4 passam a conter 140 instâncias. Com isso, um conjunto de teste contendo apenas 6 semanas de avaliação também dificultam uma melhor investigação do problema, já que apenas um valor discrepante pode acabar influenciando no resultado final. Observa-se esse comportamento ao analisar os resultados do Produto 5, que apresenta um valor discrepante no conjunto de teste.

Em paralelo à abordagem seguida nesse trabalho, uma outra pesquisa estava sendo executada, realizando a aplicação de Métodos de Aprendizagem Profunda para a previsão de demanda dos mesmos produtos, com a mesma base de dados. A seguir poderemos observar os melhores resultados alcançados pelo trabalho de Dissertação de Mestrado desenvolvido por Soares (2020). Nele, o autor utilizou dois algoritmos de Aprendizagem Profunda, o LSTM e Convolutional Neural Network - Long short-term memory (CNN-LSTM). Nas tabelas abaixo, podemos comparar os desempenhos desses algoritmos, com os melhores resultados alcançados nessa pesquisa, utilizando a métrica MAPE.

Tabela 28 – Resultados Comparação Produto 1

Algoritmo	MAPE
ARIMA	40,66
Modelo Interno	49,94
LSTM	11,52
CNN-LSTM	9,95
XGBoost	11,85

Tabela 29 – Resultados Comparação Produto 2

Algoritmo	MAPE
ARIMA	8,05
Modelo Interno	50,62
LSTM	8,18
CNN-LSTM	7,66
Decision Tree	6,56

Tabela 30 – Resultados Comparação Produto 3

Algoritmo	MAPE
ARIMA	82,72
Modelo Interno	197,13
LSTM	47,81
CNN-LSTM	52,47
MLP	70,95

Tabela 31 – Resultados Comparação Produto 4

Algoritmo	MAPE
ARIMA	27,04
Modelo Interno	94,9
LSTM	27,05
CNN-LSTM	21,32
SGD	27,82

Tabela 32 – Resultados Comparação Produto 5

Algoritmo	MAPE
ARIMA	413,98
Modelo Interno	133,68
LSTM	356,83
CNN-LSTM	339,19
SVR	110,0

Ao analisarmos os resultados alcançados por Soares (2020), podemos concluir que as duas abordagens, com Métodos de Regressão e Algoritmos de Aprendizagem Profunda, apresentam desempenhos superiores ou similares aos desenvolvidos pela Empresa A e B.

6 CONCLUSÃO

6.1 CONSIDERAÇÕES FINAIS

Este trabalho tem como objetivo principal realizar uma investigação ampla, utilizando um conjunto de algoritmos de Aprendizagem de Máquinas, especificamente Métodos de Regressão, para a realização da previsão de demanda, através da modelagem e análises de Séries Temporais, tornando-se uma ferramenta de apoio a decisão de gestores de organizações empresariais. Além disso, também propôs-se a realização da comparação desses modelos desenvolvidos, com outros já existentes implementados e aplicados por outras empresas, utilizando ARIMA e modelos internos. Ademais, investigar a utilização de variáveis externas ao contexto da empresa, a fim de observar a influência desses atributos de entrada nos modelos de previsão, também foi um dos objetivos propostos.

A fim de que esse objetivo fosse atingido, foi utilizado a metodologia CRISP-DM, em que suas primeiras 5 etapas foram desenvolvidas e demonstradas nos dois últimos capítulos dessa dissertação. Para isso foram utilizados módulos/bibliotecas disponíveis na linguagem de programação *Python*, sendo eles o *Scikit-learn* para a utilização dos algoritmos de Aprendizagem de Máquinas, *Statsmodels* para análises estatísticas e *Matplotlib* para a construção dos gráficos.

No capítulo anterior foram mostrados e interpretados os resultados alcançados. Para a previsão da demanda de 5 produtos da empresa, os modelos de regressão demonstraram ser uma abordagem eficiente para a realização de previsão de séries temporais, através da Previsão de Demanda.

No total, foram desenvolvidos 8 modelos para cada produto, que utilizaram 3 abordagens diferentes, sendo elas a utilização dos dados de treinamento sem transformação na variável alvo ("Venda Real"), a aplicação da técnica de Normalização e retirada do componente de sazonalidade na série temporal (Dessazonalização) na variável alvo e por fim a adição de variáveis externas nos modelos.

Diante dessas diferentes abordagens, observou-se que a aplicação de métodos de regressão na previsão de séries temporais é uma abordagem eficiente, alcançando resultados superiores ou similares aos desenvolvidos anteriormente. Quando aplicada as técnicas de Normalização e Dessazonalização houve uma contribuição para melhorias nos modelos desenvolvidos. Além disso, a utilização das variáveis externas, de clima e econômicas, demonstraram ser atributos que podem contribuir para uma boa previsão da demanda, mostrando que informações externas ao contexto podem ser um bom recurso na modelagem, auxiliando na solução do problema proposto.

Dos modelos desenvolvidos, os que visavam a previsão dos Produtos 1 e 2 apresentaram os melhores resultados, com a utilização dos algoritmos de XGBoost, Decision Tree e

Random Forest. Para o Produto 1 o MAPE alcançado foi de 11,85% com a utilização do XGBoost, enquanto que para o Produto 2 um MAPE de 6,56% com Decision Tree e 7,15% com Random Forest, utilizando técnicas de Dessazonalização e Normalização, além das variáveis externas. Os modelos de predição dos demais produtos apresentaram desempenhos superiores ou similares ao desenvolvidos pela Empresa A e B. Essa conclusão se deve ao fato de que na comparação da métrica MAPE, eles apresentam resultados melhores, mas ao comparar com as demais métricas apresentam um desempenho similar ou inferior, principalmente na qualidade da predição pois apresentam alguns valores da previsão que destoam. Com isso, no Produto 3, os algoritmos que se destacaram foram o MLP e SGD, com um MAPE de 70,95% e 71,30%, respectivamente. Para esses dois, foram utilizados a técnica de Normalização, com o a aplicação de variáveis climáticas no último citado. Para o Produto 4, o algoritmo SGD também se destacou com um MAPE de 27,82%, sem transformação dos dados, porém com a aplicação da variável econômica "inflacao_ipca". Já para o Produto 5, destacam-se o SVR com um MAPE de 110% e o Decision Tree com um MAE mais baixo do que o ARIMA e o modelo interno desenvolvido pela Empresa B. Nesses dois casos o primeiro contou com a utilização da variável externa "juros_selic", enquanto que o segundo contou com a aplicação da técnica de Dessazonalização e utilização de um conjunto de variáveis climáticas. Dos algoritmos escolhidos, apenas o Linear Regression não alcançou resultados para estar entre os 10 melhores para a predição do volume de produção dos 5 produtos. Os demais, que não possuíam uma base linear, apresentaram resultados superiores ou similares aos apresentados pelos modelos que utilizaram ARIMA.

Como dito anteriormente, não houve um destaque para um algoritmo em específico, em todas as abordagens utilizadas e nas predições dos diversos produtos. Com isso, conclui-se que para a solução de problemas similares, aconselha-se a aplicação de um conjunto de métodos de aprendizagem de máquinas, a fim de encontrar aquele que obtém o melhor resultado para a solução daquele problema específico. Atualmente, diversas ferramentas de Tecnologia da Informação tornam a aplicação dessa abordagem viável.

Essas diferenças nas análises se deve a uma dos desafios encontradas para o desenvolvimento do trabalho, que foi a pequena amostra dos dados para a modelagem do problema. Após a realização dos filtros para que fossem desenvolvidos os modelos específicos para cada produto, os dados passaram a contar com 140 registros semanais, sendo os 6 últimos destinados para a avaliação do modelo. Ou seja, 134 semanas foram utilizadas para o conjunto de treinamento e validação dos modelos e 6 semanas para o conjunto de teste. Logo, torna-se difícil a generalização das previsões realizadas. Em um conjunto que conta com 6 semanas, um valor discrepante em apenas uma das semanas torna a qualidade da predição ruim, principalmente quando as métricas utilizadas são dependentes da escala dos valores, não representando da melhor forma os modelos. Portanto, diante dessa dificuldade, alguns dos modelos não conseguiram contornar esse obstáculo. Aqueles que se

destacaram foram os que melhor lograram êxito em se adaptar a essa situação. Além dessa limitação, uma amostra pequena também traz empecilhos para encontrar combinações de atributos que formem padrões, favorecendo a aplicação dos métodos de Aprendizagem de Máquinas, melhorando o seu desempenho. Apesar dessa dificuldades, os resultados mostrados no capítulo anterior, mostram a eficiência dos algoritmos de regressão para a solução do problema.

6.2 TRABALHOS FUTUROS

Diante das considerações anteriores, os bons resultados são encorajadores e demonstram a oportunidade para a realização de novas análises e realizações de trabalhos futuros. Dentre essas possibilidades estão:

- **Abranger o número de algoritmos de regressão:** Pesquisar na literatura algoritmos de regressão que não foram contemplados nesse trabalho, abrangendo a investigação do desempenho de algoritmos de regressão na previsão de Séries Temporais.
- **Aplicação de métodos ensembles:** Investigar se a aplicação da técnica de ensemble, combinando algoritmos de regressão, propiciaria uma melhoria nos resultados alcançados.
- **Busca por novas variáveis externas:** Com os resultados alcançados utilizando variáveis externas, agregando valor ao modelos que as utilizaram, a busca por outras fontes de dados públicos, para extração de variáveis também demonstra ser uma boa oportunidade de melhorias futuras.
- **Aplicação de outras técnicas de Ajustes de Parâmetros:** A fim de ampliar a investigação da aplicação de algoritmos de regressão, a utilização de outro conjunto de técnicas de ajuste dos parâmetros observando a possibilidade possíveis melhorias no resultado.
- **Análise Qualitativa dos resultados:** além da análise quantitativa, como já foi realizada nesse trabalho, a avaliação qualitativa das previsões, a fim de observar a qualidade da predição, com o auxílio de especialistas das empresas que produzem os produtos.
- **Publicação do trabalho:** submissão e publicação do estudo realizado, no formato de artigo.

REFERÊNCIAS

- ABOLGHASEMI, M.; BEH, E.; TARR, G.; GERLACH, R. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering*, Elsevier, p. 106380, 2020.
- AHMAD, T.; CHEN, H. Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems. *Sustainable cities and society*, Elsevier, v. 45, p. 460–473, 2019.
- AL-MUSAYLH, M. S.; DEO, R. C.; ADAMOWSKI, J. F.; LI, Y. Short-term electricity demand forecasting with mars, svr and arima models using aggregated demand data in queensland, australia. *Advanced Engineering Informatics*, Elsevier, v. 35, p. 1–16, 2018.
- ALTUNKAYNAK, A.; NIGUSSIE, T. A. Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models. *Urban Water Journal*, Taylor & Francis, v. 15, n. 2, p. 177–181, 2018.
- AYELE, T. *Ethiopian Health Commodity Demand Prediction*. Tese (Doutorado) — ASTU, 2019.
- BAKKER, M.; DUIST, H. V.; SCHAGEN, K. V.; VREEBURG, J.; RIETVELD, L. Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering*, Elsevier, v. 70, p. 93–102, 2014.
- BARRETO, A. F.; ANTONOVZ, T. A má gestão de custos influência na mortalidade das empresas no brasil. *Anais do V SINGEP, São Paulo*, 2016.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, n. Feb, p. 281–305, 2012.
- BIANCO, V.; MANCA, O.; NARDINI, S. Linear regression models to forecast electricity consumption in italy. *Energy Sources, Part B: Economics, Planning, and Policy*, Taylor & Francis, v. 8, n. 1, p. 86–93, 2013.
- BIBAUD-ALVES, J.; THOMAS, P.; BRIL, H. Demand forecasting using artificial neuronal networks and time series: application to a french furniture manufacturer case study. In: . [S.l.: s.n.], 2019.
- BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006.
- BONIFÁCIO, A. L. et al. Estratégia computacional para avaliação de propriedades mecânicas de concreto de agregado leve. Universidade Federal de Juiz de Fora (UFJF), 2017.
- BOTTOU, L. Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010*. [S.l.]: Springer, 2010. p. 177–186.
- BOURLARD, H.; KAMP, Y. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, Springer, v. 59, n. 4-5, p. 291–294, 1988.

- BRASIL, B. C. do. *Política monetária*. 2020. Disponível em: <<https://www.bcb.gov.br/>>.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: springer, 2016.
- BROWNLEE, J. How to one hot encode sequence data in python. *Machine Learning Mastery*, 2017.
- ÇAMURDAN, Z.; GANIZ, M. C. Machine learning based electricity demand forecasting. In: IEEE. *2017 International Conference on Computer Science and Engineering (UBMK)*. [S.l.], 2017. p. 412–417.
- CANDELIERI, A.; GIORDANI, I.; ARCHETTI, F.; BARKALOV, K.; MEYEROV, I.; POLOVINKIN, A.; SYSOYEV, A.; ZOLOTYKH, N. Tuning hyperparameters of a svm-based water demand forecasting system through parallel global optimization. *Computers & Operations Research*, Elsevier, v. 106, p. 202–209, 2019.
- CHANDEL, A.; DUBEY, A.; DHAWALE, S.; GHUGE, M. Sales prediction system using machine learning. *International Journal of Scientific Research and Engineering Development*, v. 2, 2019.
- CHANG, Y.-W.; TSAI, C.-Y. Apply deep learning neural network to forecast number of tourists. In: IEEE. *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. [S.l.], 2017. p. 259–264.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: ACM. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.], 2016. p. 785–794.
- CHEN, T.; HE, T.; BENESTY, M.; KHOTILOVICH, V.; TANG, Y. Xgboost: extreme gradient boosting. *R package version 0.4-2*, p. 1–4, 2015.
- CORTES, C.; VAPNIK, V. Support vector machine. *Machine learning*, v. 20, n. 3, p. 273–297, 1995.
- DEB, C.; ZHANG, F.; YANG, J.; LEE, S. E.; SHAH, K. W. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 74, p. 902–924, 2017.
- DHALIWAL, S.; NAHID, A.-A.; ABBAS, R. Effective intrusion detection system using xgboost. *Information*, Multidisciplinary Digital Publishing Institute, v. 9, n. 7, p. 149, 2018.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.
- DYER, R. *Predicting residential demand: applying random forest to predict housing demand in Cape Town*. Tese (Doutorado) — University of Cape Town, 2018.
- EHLERS, R. S. Análise de séries temporais. *Universidade Federal do Paraná*, 2007.

-
- EVERS, J.; TAVASSZY, L.; DUIN, J. van; SCHOTT, D.; GORTE, F. Demand forecast models for online supermarkets. 2018.
- FANG, T.; LAHDELMA, R. Evaluation of a multiple linear regression model and sarima model in forecasting heat demand for district heating system. *Applied energy*, Elsevier, v. 179, p. 544–552, 2016.
- FENG, Y.; WANG, S. A forecast for bicycle rental demand based on random forests and multiple linear regression. In: IEEE. *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. [S.l.], 2017. p. 101–105.
- FONTEIJN, R.; CASTELIJNS, T.; GROND, M.; NGUYEN, P. H.; MORREN, J.; SLOOTWEG, H. Short-term load forecasting on mv/lv transformer level. AIM, 2019.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.
- FULLER, W. A. Introduction to statistical time series, new york: Johnwiley. *FullerIntroduction to Statistical Time Series1976*, 1976.
- FUNDERA. *What Percentage of Small Businesses Fail? (And Other Need-to-Know Stats)*. 2020. Disponível em: <<https://www.fundera.com/blog/what-percentage-of-small-businesses-fail>>.
- GKEREKOS, C.; LAZAKIS, I.; PAPAGEORGIU, S. Leveraging big data for fuel oil consumption modelling. Technische Universitat Hamburg-Hamburg, 2018.
- GUMUS, M.; KIRAN, M. S. Crude oil price forecasting using xgboost. In: IEEE. *2017 International Conference on Computer Science and Engineering (UBMK)*. [S.l.], 2017. p. 1100–1103.
- GUO, Q.; FENG, Y.; SUN, X.; ZHANG, L. Power demand forecasting and application based on svr. *Procedia computer science*, Elsevier, v. 122, p. 269–275, 2017.
- HAMBALI, A.; AKINYEMI, M.; JYUSUF, N. Electric power load forecast using decision tree algorithms. *Comput. Inf. Syst. Dev. Inform. Allied Res. J*, v. 4, p. 29–42, 2016.
- HARRINGTON, P. *Machine learning in action*. [S.l.]: Manning Publications Co., 2012.
- HONG, T.; GUI, M.; BARAN, M. E.; WILLIS, H. L. Modeling and forecasting hourly electric load by multiple linear regression with interactions. In: IEEE. *IEEE PES General Meeting*. [S.l.], 2010. p. 1–8.
- HUNTER, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007.
- JOHANNESSEN, N. J.; KOLHE, M.; GOODWIN, M. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of cleaner production*, Elsevier, v. 218, p. 555–564, 2019.
- JUNIOR, A. P. S.; SEGATTO, M.; CALIA, R. C.; FIOROTTO, J. A.; BERTON, L. T. A tecnologia da informação como suporte ao ajuste da previsão da demanda: um estudo de caso em uma empresa de bebidas carbonatadas. *Revista Produção Online*, v. 10, n. 3, p. 621–648, 2010.

- KATSATOS, A.; MOUSTRIS, K. Application of artificial neuron networks as energy consumption forecasting tool in the building of regulatory authority of energy, athens, greece. *Energy Procedia*, Elsevier, v. 157, p. 851–861, 2019.
- KAZEMI, A.; HOSSEINZADEH, M. et al. A multi-level fuzzy linear regression model for forecasting industry energy demand of iran. *Procedia-Social and Behavioral Sciences*, Elsevier, v. 41, p. 342–348, 2012.
- KILLADA, P. *Data Analytics using Regression Models for Health Insurance Market place Data*. Tese (Doutorado) — University of Toledo, 2017.
- KRESS, G.; SNYDER, J. *Forecasting and market analysis techniques: A practical approach*. [S.l.]: Praeger Pub Text, 1994.
- LAHOUAR, A.; SLAMA, J. B. H. Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, Elsevier, v. 103, p. 1040–1051, 2015.
- LAUER, T.; LEGNER, S.; HENKE, M. Application of machine learning on plan instability in master production planning of a semiconductor supply chain. *IFAC-PapersOnLine*, Elsevier, v. 52, n. 13, p. 1248–1253, 2019.
- LIAW, A.; WIENER, M. et al. Classification and regression by randomforest. *R news*, v. 2, n. 3, p. 18–22, 2002.
- LUSTOSA, L. J.; MESQUITA, M. A. de; OLIVEIRA, R. J. *Planejamento e controle da produção*. [S.l.]: Elsevier Brasil, 2008.
- MA, X.; TIAN, Y.; LUO, C.; ZHANG, Y. Predicting future visitors of restaurants using big data. In: IEEE. *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*. [S.l.], 2018. v. 1, p. 269–274.
- MADDALA, G.; LAHIRI, K. *Introduction to econometrics. 4. utg.* [S.l.]: Chichester, Wiley, 2009.
- MAGERMAN, D. M. Statistical decision-tree models for parsing. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. [S.l.], 1995. p. 276–283.
- MARINO, D. L.; AMARASINGHE, K.; MANIC, M. Building energy load forecasting using deep neural networks. In: IEEE. *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. [S.l.], 2016. p. 7046–7051.
- MARTÍNEZ-ÁLVAREZ, F.; TRONCOSO, A.; ASECIO-CORTÉS, G.; RIQUELME, J. A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, Multidisciplinary Digital Publishing Institute, v. 8, n. 11, p. 13162–13193, 2015.
- MCKINNEY, W. pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, v. 14, 2011.
- MENDONÇA, S. A. T. de; RUZZÃO, A. P. de A.; SANTOS, J. E. dos; AZADINHO, L. Z. O planejamento estratégico como ferramenta: Um estudo sobre a eficiência das micro e pequenas empresas brasileiras. *Administração de Empresas em Revista*, v. 16, n. 17, p. 50–68, 2017.

- MIRANDA, L. B. S.; FRANÇA, I. M.; FERREIRA, J. A. D.; MARTINS, M. R.; MIRANDA, L. M. Ferramentas administrativas utilizadas nas micro e pequenas empresas: Uma análise realizada no centro comercial da cidade de viçosa-mg. *Revista Conbrad [ISSN 2525-6815] Qualis B5*, v. 1, n. 2, p. 39–52, 2016.
- MIRANDA, R. G. de; ANDRADE, G. J. P. O. de; GERBER, J. Z.; BORNIA, A. C. Método estruturado para o processo de planejamento da demanda nas organizações. *Revista ADMpg*, v. 4, n. 2, 2011.
- MLXTEND. *Gradient Descent and Stochastic Gradient Descent*. 2019. Disponível em: <http://rasbt.github.io/mlxtend/user_guide/general_concepts/gradient-optimization/>.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. [S.l.]: John Wiley & Sons, 2012. v. 821.
- Noureen, S.; Atique, S.; Roy, V.; Bayne, S. Analysis and application of seasonal arima model in energy demand forecasting: A case study of small scale agricultural load. In: *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. [S.l.: s.n.], 2019. p. 521–524.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, v. 12, n. Oct, p. 2825–2830, 2011.
- PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. *Production*, SciELO Brasil, v. 11, n. 1, p. 43–64, 2001.
- PHAN, H.; MAASS, M.; MAZUR, R.; MERTINS, A. Random regression forests for acoustic event detection and classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, IEEE, v. 23, n. 1, p. 20–31, 2014.
- POPADIUK, S.; SANTOS, A. E. M. D. Conhecimentos tácito, explícito e cultural no planejamento da demanda. *JISTEM: Journal of Information Systems and Technology Management*, Universidade de São Paulo, v. 7, n. 1, p. 205–226, 2010.
- RAI, S.; GUPTA, A.; ANAND, A.; TRIVEDI, A.; BHADAURIA, S. Demand prediction for e-commerce advertisements: A comparative study using state-of-the-art machine learning methods. *10th ICCCNT 2019*, 2019.
- REISEN, V. A.; LEVY-LEDUC, C.; BOURGUIGNON, M.; BOISTARD, H. Robust dickey–fuller tests based on ranks for time series with additive outliers. *Metrika*, Springer, v. 80, n. 1, p. 115–131, 2017.
- RIZKYA, I.; SYAHPUTRI, K.; SARI, R.; SIREGAR, I.; UTAMININGRUM, J. Autoregressive integrated moving average (arima) model of forecast demand in distribution centre. In: IOP PUBLISHING. *IOP Conference Series: Materials Science and Engineering*. [S.l.], 2019. v. 598, n. 1, p. 012071.
- ROSENBLATT, F. *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*. [S.l.], 1961.

- ROSSETO, M. et al. Técnicas qualitativas de previsão de demanda: um estudo multicase com empresas do ramo de alimentos. *SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA-SEGeT*, v. 8, 2011.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. *Learning internal representations by error propagation*. [S.l.], 1985.
- SAPUTRA, Y. M.; HOANG, D. T.; NGUYEN, D. N.; DUTKIEWICZ, E.; MUECK, M. D.; SRIKANTESWARA, S. Energy demand prediction with federated learning for electric vehicle networks. *arXiv preprint arXiv:1909.00907*, 2019.
- SCIKIT-LEARN. *Linear Models*. 2019. Disponível em: <https://scikit-learn.org/stable/modules/linear_model.html#linear-model>.
- SCOTT, A. J. *Principles of forecasting: a handbook for researchers and practitioners*. [S.l.]: Kluwer Academic Publishers, 2001.
- SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: *9th Python in Science Conference*. [S.l.: s.n.], 2010.
- SEBRAE. *Sobrevivência das empresas no Brasil*. 2016. Disponível em: <<https://goo.gl/KxhndK>>.
- SEBRAE. *Sobrevivência das empresas no Brasil*. 2017. Disponível em: <<https://blog.sebrae-sc.com.br/previsao-de-demanda/>>.
- SHARMA, A. *Understanding Activation Functions in Neural Networks*. 2017. Disponível em: <<https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>>.
- SHEARER, C. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, THE DATA WAREHOUSE INSTITUTE, v. 5, n. 4, p. 13–22, 2000.
- SLIMANI, I.; FARISSI, I. E.; AL-QUALSADI, S. A. Configuration of daily demand predicting system based on neural networks. In: IEEE. *2016 3rd International conference on logistics operations management (GOL)*. [S.l.], 2016. p. 1–5.
- SOARES, E. *Modelos de Redes Neurais Recorrentes para Previsão de Demanda de Curto Prazo*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Pernambuco, 2020.
- SON, H.; KIM, C. Short-term forecasting of electricity demand for the residential sector using weather and social variables. *Resources, conservation and recycling*, Elsevier, v. 123, p. 200–207, 2017.
- SZOPLIK, J. Forecasting of natural gas consumption with artificial neural networks. *Energy*, Elsevier, v. 85, p. 208–220, 2015.
- TIWARI, S.; SABZEHGAR, R.; RASOULI, M. Short term solar irradiance forecast using numerical weather prediction (nwp) with gradient boost regression. In: IEEE. *2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*. [S.l.], 2018. p. 1–8.

TOWARDS-DATA-SCIENCE. *Random Forest Regression - Along with its implementation in Python*. 2019. Disponível em: <<https://towardsdatascience.com/random-forest-and-its-implementation-71824ced454f>>.

TSO, G. K.; YAU, K. K. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, Elsevier, v. 32, n. 9, p. 1761–1768, 2007.

TSURUOKA, Y.; TSUJII, J.; ANANIADOU, S. Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. [S.l.], 2009. p. 477–485.

VANICHRUJEE, U.; HORANONT, T.; PATTARA-ATIKOM, W.; THEERAMUNKONG, T.; SHINOZAKI, T. Taxi demand prediction using ensemble model based on rnns and xgboost. In: IEEE. *2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES)*. [S.l.], 2018. p. 1–6.

VIDHYA, A. *A Complete Tutorial on Time Series Modeling in R*. 2015. Disponível em: <<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>>.

VINAGRE, E.; PAZ, J. F. D.; PINTO, T.; VALE, Z.; CORCHADO, J. M.; GARCIA, O. Intelligent energy forecasting based on the correlation between solar radiation and consumption patterns. In: IEEE. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. [S.l.], 2016. p. 1–7.

VU, D. H.; MUTTAQI, K. M.; AGALGAONKAR, A. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, Elsevier, v. 140, p. 385–394, 2015.

WAN, X.-l.; ZHANG, Z.; RONG, X.-x.; MENG, Q.-c. Exploring an interactive value-adding data-driven model of consumer electronics supply chain based on least squares support vector machine. *Scientific Programming*, Hindawi Publishing Corp., v. 2016, p. 4, 2016.

WEISBERG, S. *Applied linear regression*. [S.l.]: John Wiley & Sons, 2005. v. 528.

WERNER, L.; RIBEIRO, J. L. D. Modelo composto para prever demanda através da integração de previsões. *Production*, SciELO Brasil, v. 16, n. 3, p. 493–509, 2006.

XU, M.; WATANACHATURAPORN, P.; VARSHNEY, P. K.; ARORA, M. K. Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, Elsevier, v. 97, n. 3, p. 322–336, 2005.

ZHANG, B.; PU, Y. Study on tourist volume forecasting based on aba-svr model within network environment. In: ATLANTIS PRESS. *3rd International Conference on Judicial, Administrative and Humanitarian Problems of State Structures and Economic Subjects (JAHP 2018)*. [S.l.], 2018.