



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
NÚCLEO DE GESTÃO
CURSO DE GRADUAÇÃO EM ADMINISTRAÇÃO

JEOVÁ RIBEIRO FALCÃO FILHO

**PREVISÃO DE DEMANDA: ANÁLISE DE MODELOS PREDITIVOS EM UMA
ORGANIZAÇÃO DO SETOR DE MOTO PEÇAS EM CARUARU-PE**

Caruaru

2023

JEOVÁ RIBEIRO FALCÃO FILHO

PREVISÃO DE DEMANDA: ANÁLISE DE MODELOS PREDITIVOS EM UMA ORGANIZAÇÃO DO SETOR DE MOTO PEÇAS EM CARUARU-PE

Trabalho de Conclusão de Curso apresentado ao Curso de Administração da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Bacharel em Administração.

Área de concentração: Administração e predição.

Orientadora: Prof^ª. Dr^ª. Regilda da Costa e Silva Menezes.

Caruaru

2023

Ficha de identificação de obra elaborada pelo autor,
Através do programa de geração automática do SIB/UFPE

Falcão Filho, Jeová Ribeiro.

Previsão de demanda: Análise de modelos preditivos em uma organização do setor de moto peças em Caruaru-PE / Jeová Ribeiro Falcão Filho. – Caruaru, 2023.

40

Orientador(a): Regilda da Costa e Silva Menezes

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Pernambuco, Centro Acadêmico do Agreste, Administração, 2023.

Inclui referências, anexos.

1. Previsão de demanda. 2. Modelos Preditivos. 3. Estratégia. 4. Análise de modelos. I. Menezes, Regilda da Costa e Silva. (orientação). II. Título.

310 CDD (22.ed.)

JEOVÁ RIBEIRO FALCÃO FILHO

**PREVISÃO DE DEMANDA: ANÁLISE DE MODELOS PREDITIVOS EM UMA
ORGANIZAÇÃO DO SETOR DE MOTO PEÇAS EM CARUARU-PE**

Trabalho de Conclusão de Curso apresentado à coordenação do Curso de Graduação em Administração do Campus Agreste da Universidade Federal de Pernambuco - UFPE, na modalidade de monografia, como requisito parcial para a obtenção do título de Bacharel em Administração.

Aprovada em: 27/09/2023.

BANCA EXAMINADORA

Prof^a. Dra. REGILDA DA COSTA E SILVA MENÊZES (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. JOSÉ CICERO DE CASTRO (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. MARCOS LUIZ HENRIQUE (Examinador Interno)
Universidade Federal de Pernambuco

Dedico esse trabalho aos meus pais, irmãos e amigos, que estiveram comigo em todo meu ciclo de vida me auxiliando, me alegrando e me motivando em todo o meu processo educacional.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer a toda minha família e amigos que estiveram próximos durante todo meu ciclo de vida, pois com eles adquiri mais vontade e determinação durante toda minha trajetória.

Também gostaria de agradecer a professora Regilda por ter me orientado nesse projeto de conclusão do curso de graduação em Administração.

Gostaria de agradecer também a todos os outros professores e colegas que participaram dessa minha trajetória no curso de Administração, acabei aprendendo bastante com cada um que esteve presente.

RESUMO

Para que uma empresa consiga sucesso competitivo é necessário um planejamento adequado tendo em vista seu faturamento e demanda pré-definida. Com isso, a previsão de demanda desempenha um papel fundamental nas operações de uma organização, pois compreender as tendências de receita é essencial para sua gestão financeira e estratégica. Nesse contexto, este estudo tem como objetivo analisar a eficácia de diferentes modelos preditivos em um cenário empresarial específico, focando em uma empresa do setor de moto peças situada em Caruaru-PE. Para alcançar esse objetivo, foram utilizados os dados de faturamento da empresa de agosto de 2022 a julho de 2023 e foram aplicados os modelos de média móvel, regressão linear, árvore de decisão e floresta aleatória. A avaliação do desempenho desses modelos foi conduzida por meio do cálculo do erro médio absoluto, proporcionando uma medida precisa da eficiência de cada abordagem preditiva. Essa análise proporciona às empresas informações valiosas para tomar decisões mais assertivas, otimizar seus recursos e enfrentar desafios comerciais com maior confiança, contribuindo assim para o alcance do sucesso organizacional.

Palavras-chave: Previsão de demanda; Modelos preditivos; Estratégia; Análise de modelos.

ABSTRACT

In order for a company to achieve competitive success, adequate planning is necessary, considering its revenue and pre-defined demand. In this regard, demand forecasting plays a crucial role in an organization's operations because understanding revenue trends is essential for financial and strategic management. In this context, this study aims to analyze the effectiveness of different predictive models in a specific business scenario, focusing on a motorcycle parts company located in Caruaru-PE. To achieve this objective, the company's revenue data from August 2022 to July 2023 were utilized, and the following models were applied: moving average, linear regression, decision tree, and random forest. The evaluation of these models' performance was conducted by calculating the mean absolute error, providing a precise measure of the efficiency of each predictive approach. This analysis provides valuable information to companies for making more informed decisions, optimizing their resources, and facing business challenges with greater confidence, thereby contributing to the achievement of organizational success.

Keywords: Demand forecasting; Predictive models; Strategy; Model analysis.

LISTA DE FIGURAS

Figura 1 –	Etapas para construção de um modelo preditivo	15
Figura 2 –	Implementação da média móvel feito no livro de Lustosa	16
Figura 3 –	Gráfico das vendas do produto	18
Figura 4 –	Gráfico da demanda do produto com suas médias móveis	19
Figura 5 –	Gráfico de dispersão com a reta da regressão linear	21
Figura 6 –	Árvore de decisão regressiva	22
Figura 7 –	Uma ilustração de uma Random Forest	24
Figura 8 –	Etapas do projeto	25
Figura 9 –	Média móvel com o período de três meses	26
Figura 10 –	Média móvel com o período de seis meses	27
Figura 11 –	Regressão linear com os dados diários	28
Figura 12 –	Distribuição dos dados de treino após a diferenciação	29
Figura 13 –	Distribuição dos dados de teste após a diferenciação	29
Figura 14 –	Aplicação do modelo de árvore de decisão	29
Figura 15 –	Aplicação do modelo de floresta aleatória	30

LISTA DE TABELAS

Tabela 1 –	Resultados dos métodos de predição utilizados	30
------------	---	----

LISTA DE SIGLAS

R^2	Coeficiente de determinação
MM	Média Móvel
D_i	Demanda ocorrida no período i
n	Número de períodos da amostra
α	Intercepto
β	Coeficiente de inclinação da reta
COV	Covariância
VAR	Variância
MSE	Erro quadrático médio
Y	Valor observado para cada amostra
\hat{Y}	Valor previsto pelo modelo para cada amostra

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	13
2.1	GERAL	13
2.2	ESPECÍFICOS	13
3	REFERENCIALTEÓRICO.....	14
3.1	PREVISÃO DE DEMANDA	14
3.2	MÉTODOS PREDITIVOS	16
3.2.1	MÉDIAS MÓVEIS	17
3.2.2	REGRESSÃO LINEAR	19
3.2.3	ÁRVORE DE DECISÃO	21
3.2.4	FLORESTA ALEATÓRIA	23
4	ESTUDO DE CASO	24
5	ANÁLISE DOS RESULTADOS.....	26
6	CONCLUSÃO	31
7	TRABALHOS FUTUROS	31
	REFERÊNCIAS	32
	ANEXO A – DIVISÃO DE DADOS PARA FLORESTA ALEATÓRIA E ÁRVORE DE DECISÃO	34
	ANEXO B – ALGORITMO DA FLORESTA ALEATÓRIA	35
	ANEXO C – ALGORITMO DA ÁRVORE DE DECISÃO	36
	ANEXO D – ALGORITMO DA REGRESSÃO LINEAR	37
	ANEXO E – DEMANDA DO PRODUTO	38
	ANEXO F – DEMANDA COM SUAS MÉDIAS MÓVEIS DE PERÍODOS IGUAIS A 3 E 6.....	39
	ANEXO G – AMOSTRA COM VALORES DE X E Y	40

1 INTRODUÇÃO

Prever o faturamento de uma empresa ajuda a fornecer uma visão clara e realista das metas financeiras da organização, fazendo assim, com que ela possa estabelecer estratégias e alocar recursos de uma forma mais adequada. A utilização do conceito de séries temporais para previsão de valores futuros auxilia na concepção do planejamento econômico e de negócios; no controle de estoque e produção; e na otimização de seus respectivos processos (BOX; JENKINS, 2016).

Compreender as tendências de receita por meio da predição de demanda desempenha um papel crucial na gestão financeira e no planejamento estratégico das empresas, permitindo a identificação de oportunidades de crescimento, a avaliação da viabilidade de investimentos e o estabelecimento de metas realistas para o futuro.

Existem inúmeros métodos de previsão de séries temporais, variando desde os mais simples e amplamente empregados, como as médias móveis, até os mais complexos, como as redes neurais. Esses métodos têm como propósito estimar valores futuros, fazendo uso das informações históricas disponíveis, com o intuito de identificar padrões nos dados. Esses padrões podem ser significativos a ponto de gerar uma vantagem, normalmente econômica (WITTEN *et al.*, 2011).

Após a utilização do método preditivo, se tem a etapa mais importante de uma análise de séries temporais, que é a validação. A validação é importante, pois permite avaliar a qualidade e a confiabilidade das previsões geradas. Com isso, é possível avaliar o desempenho do modelo, a partir de dados não utilizados no treinamento, criando assim, uma estimativa realista de sua capacidade de generalização (HYNDMAN; ATHANASOPOULOS, 2018).

Segundo Chopra e Meindl (2015) realizar previsões de demanda desempenha um papel fundamental para o alcance do sucesso organizacional, uma vez que possibilita aprimorar o planejamento da produção, gerenciar os estoques e recursos de uma forma mais eficiente, além de minimizar os custos envolvidos.

Este trabalho busca comparar diferentes métodos preditivos de uma série temporal e analisar qual a melhor solução para a determinada empresa. É evidente que há uma necessidade por encontrar os melhores métodos preditivos que se encaixem de acordo com o seu propósito de negócio e compara-los através de um indicador de desempenho. Para isso, o trabalho foi realizado com base nos dados de uma empresa do setor de moto peças, localizada na cidade de Caruaru-PE. Para a aplicação destes métodos foram considerados os dados

históricos do faturamento da empresa compreendidos entre os meses de agosto de 2022 a julho de 2023.

2 OBJETIVOS

2.1 GERAL

O trabalho tem como objetivo a utilização de modelos preditivos para a análise de uma série temporal, buscando prever e avaliar como diferentes modelos se comportam quando aplicados na análise de demanda do setor de moto peças de uma empresa localizada na cidade de Caruaru-PE.

2.2 ESPECÍFICOS.

- Analisar o desempenho dos modelos de previsão.
- Comparar os modelos tradicionais com os modelos de aprendizagem de máquina.
- Identificar possíveis formas de melhora dos modelos preditivos.

3 REFERENCIAL TEÓRICO

Este capítulo apresenta os conceitos utilizados para a construção desse trabalho, trazendo os modelos preditivos que serão utilizados como comparativo e os conceitos em relação à previsão de demanda.

3.1 PREVISÃO DE DEMANDA

A previsão de demanda desempenha um papel de destaque em várias áreas de gestão organizacional. Não importa o tamanho ou a área de atuação, todas as empresas devem se planejar adequadamente. Os planos de negócios dependem, em grande parte, de previsões de demanda, especialmente para aspectos relacionados à produção, como capacidade, matérias-primas, mão de obra e capital de giro. Essas previsões são amplamente utilizadas no mundo empresarial para o planejamento e gerenciamento tático, estratégico ou operacional, abrangendo diferentes horizontes temporais, desde dias e semanas até meses, com diferentes níveis de detalhes (ACKERMANN; SELLITTO, 2022).

É proporcionado, graças à previsão, um melhor planejamento dos recursos de uma organização. Os erros de uma previsão e planejamento podem acarretar ociosidade de recursos ou ineficiência na prestação de serviços (LUSTOSA *et al.*, 2008).

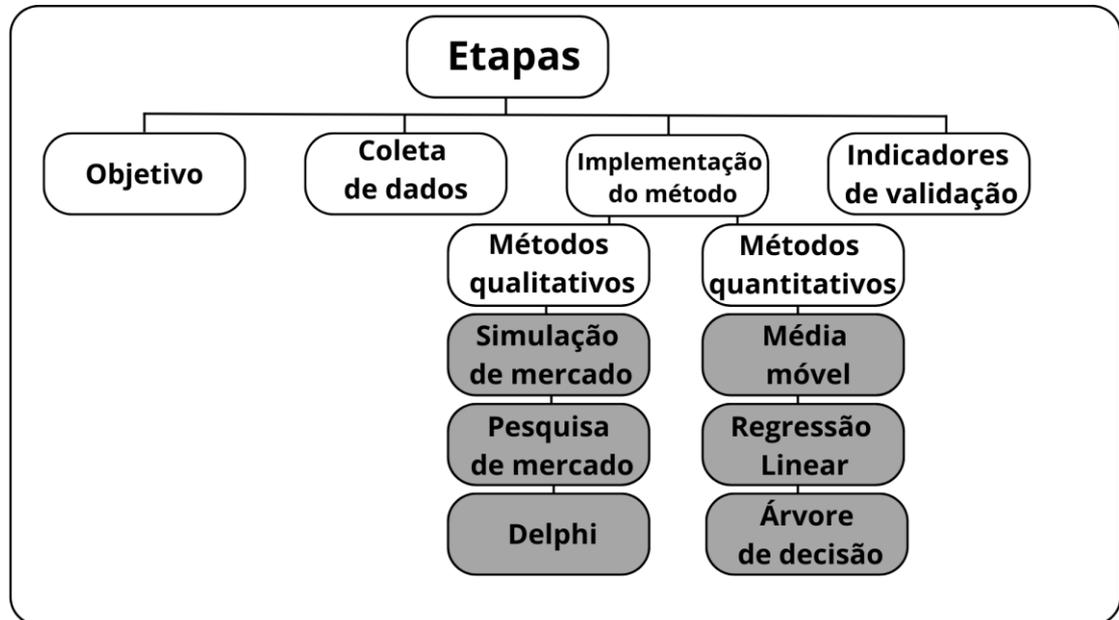
Segundo Chase (2009), o avanço no armazenamento e processamento de dados impulsionou a importância da previsão de demanda, pois ela pode agregar valor em toda a cadeia de suprimentos. Com isso, as empresas estão atualmente utilizando previsões analíticas para identificar padrões no comportamento do consumidor, avaliar a eficácia de suas estratégias de investimento e aprimorar o desempenho financeiro.

Graças às técnicas de previsão, é possível converter informações estratégicas empresariais em previsões de demanda para novos produtos, novas tecnologias, maior participação de mercado ou uma estratégia competitiva mais sólida. Afinal, o planejamento das operações e vendas futuras é uma necessidade inerente às empresas (ACKERMANN; SELLITTO, 2022).

Para realizar o processo de previsão é necessário incrementar quatro etapas, sendo o primeiro processo a definição do objetivo principal, que será associado à necessidade da previsão de demanda. Em seguida, é realizada uma coleta dos dados necessários para que se possa fazer a previsão. Como terceira etapa, se tem a execução do modelo preditivo, que pode ser classificado como qualitativo ou quantitativo. Por fim, será necessária a criação de

indicadores de erro de previsão, que são utilizados inicialmente na calibração dos modelos e, posteriormente, para indicar a precisão das previsões (LUSTOSA *et al.*, 2008).

Figura 1: Etapas para construção de um modelo preditivo

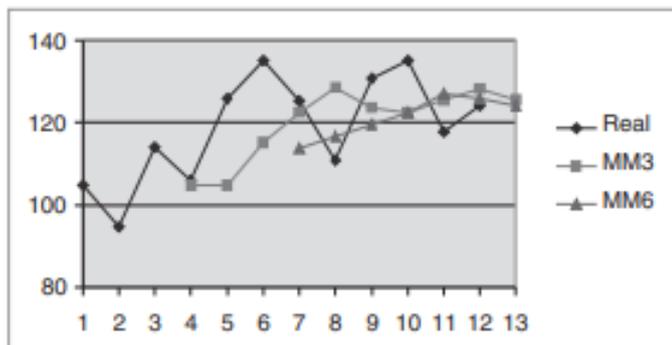


Fonte: De autoria própria.

A Figura 1 mostra o processo que deve ser seguido para construção de um modelo preditivo e alguns métodos utilizados na implementação. Os modelos qualitativos possuindo características baseada no consenso de opiniões e os modelos quantitativos, tendo como característica, o emprego de técnicas estatísticas.

Lustosa (2008) traz um exemplo de um modelo de média móvel com um registro de dados de vendas mensais de um produto, onde ele realiza duas médias móveis uma com um período de três meses e outra com um período de seis meses, onde que para esse caso, com pequena quantidade de amostra, a média móvel com período de seis meses apresentava uma curva mais suave do que à com três meses, refletindo melhor o valor médio da demanda, como mostra na Figura 2.

Figura 2: Implementação da média móvel feito no livro de Lustosa.



Fonte: LUSTOSA, p. 60, 2008.

Fazer uma previsão de demanda eficiente pode acarretar papéis importantes, visto que, a maneira de gerir uma empresa se define a partir do rumo que ela seguirá. Então o planejamento estratégico, seja da produção, vendas e finanças, surge a partir das antecipações futuras e de suas ações (MARQUES; SANTOS, 2018).

3.2 METODOS PREDITIVOS

Um modelo é uma especificação de uma relação matemática entre diferentes variáveis. No campo da predição, os dados fornecem as informações e os modelos podem se desenvolver através delas (GRUS, 2021).

Existem diversos métodos de predição, esses modelos são baseados em algoritmos estatísticos e a performance dos modelos é avaliada através de amostras, comumente é usado critérios de informação apropriados ou feito uma retenção de um conjunto de dados para criar uma amostra de validação (PETROPOULOS *et al.*, 2018).

Diversos modelos de previsão são fundamentados na estatística para sua elaboração. Entre eles o concebido por Perry Kaufman, as médias móveis; a regressão linear, idealizada por Francis Galton; a árvore de decisão, inicialmente criada por J. Ross Quinlan; além da floresta aleatória de Tin Kam Ho, que se baseia nos princípios da árvore de decisão. Além destes, há outros modelos que têm a estatística e a probabilidade como fundamentos para sua criação.

Considerando por exemplo os métodos de regressão linear, árvore de decisão e floresta aleatória, o autor Harrison (2020) utilizou o conjunto de dados habitacionais de Boston e aplicou como validação o coeficiente de determinação (r^2), o resultado para a regressão linear

foi uma pontuação de aproximadamente 71%, já a árvore de decisão a pontuação foi de aproximadamente 84,2% e, por fim, para floresta aleatória o autor obteve um resultado de aproximadamente 86,4%, sendo assim, o melhor método para esse projeto.

Segundo Rueda *et al.* (2023), em sua pesquisa, concluiu que as ferramentas de ensino a distância utilizadas durante a pandemia, utilizando regressão linear e algoritmos de deep learning, influencia positivamente na motivação dos alunos, garantindo mais satisfação e conhecimento para os mesmos.

Em outro estudo de caso, utilizando a regressão linear para previsão de demanda Sales *et al.* (2019) observou a necessidade da utilização de modelos preditivos para melhorar o gastos de uma empresa de confecção em Caruaru-PE e com auxílio da regressão linear pôde avaliar padrões e produtivos que possuem correlação entre tempo e vendas.

O aumento constante da disponibilidade de dados e o contínuo progresso nas técnicas de análise preditiva têm sido fundamentais em várias áreas, incluindo marketing, medicina, finanças e indústria, oferecendo uma vantagem competitiva significativa para as empresas que adotam essas abordagens. (PROVOST, 2013).

A análise preditiva representa uma abordagem poderosa que capacita organizações a reconhecer padrões em dados passados e aplicá-los para fazer previsões precisas em relação a eventos futuros, facilitando, desse modo, o planejamento estratégico e a gestão de riscos. (WITTEN *et al.*, 2016).

3.2.1 MÉDIAS MOVEIS

Segundo Tubino (2000) a média móvel é um modelo que utiliza um período predeterminado, normalmente os mais recentes, de dados para gerar uma previsão. Desta forma, cada novo período que venha a surgir, substitui o mais antigo utilizado no modelo inicial.

A média móvel é uma projeção bastante direta, que utiliza a técnica de replicação do último dado da série histórica, ou seja, ele antecipa que a demanda no próximo período será igual ao valor do período imediatamente anterior. Essa estratégia, embora seja bastante simplificada, tende a gerar estimativas bastante voláteis, uma vez que engloba todas as flutuações da demanda na previsão (LUSTOSA *et al.*, 2008).

$$MM = \frac{\sum_{i+1}^n Di}{n}$$

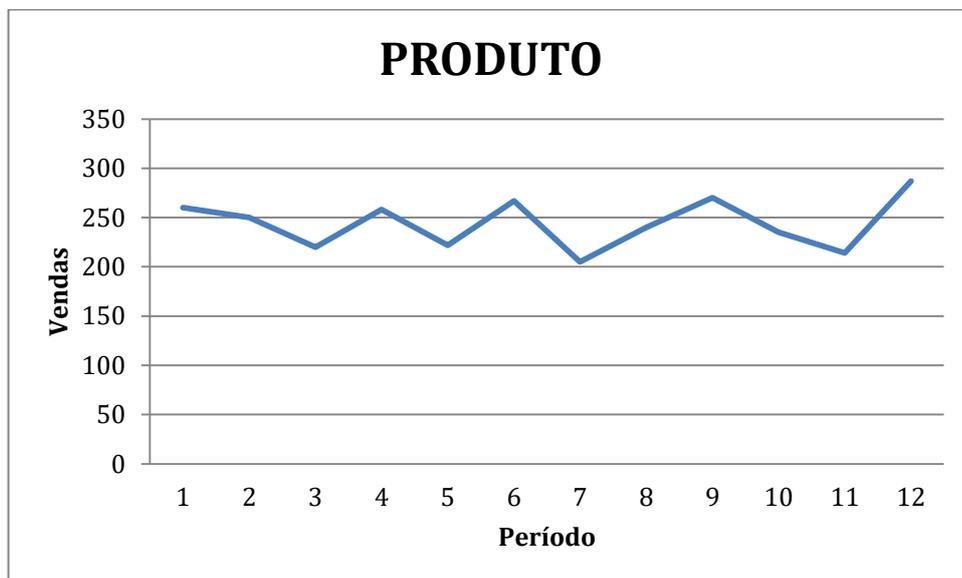
MM = média móvel

D_i = demanda ocorrida no período i

n = número de períodos utilizados no modelo

No Anexo E, apresenta, como exemplo, a demanda de um produto durante os um período de doze meses.

Figura 3: Gráfico das vendas do produto.

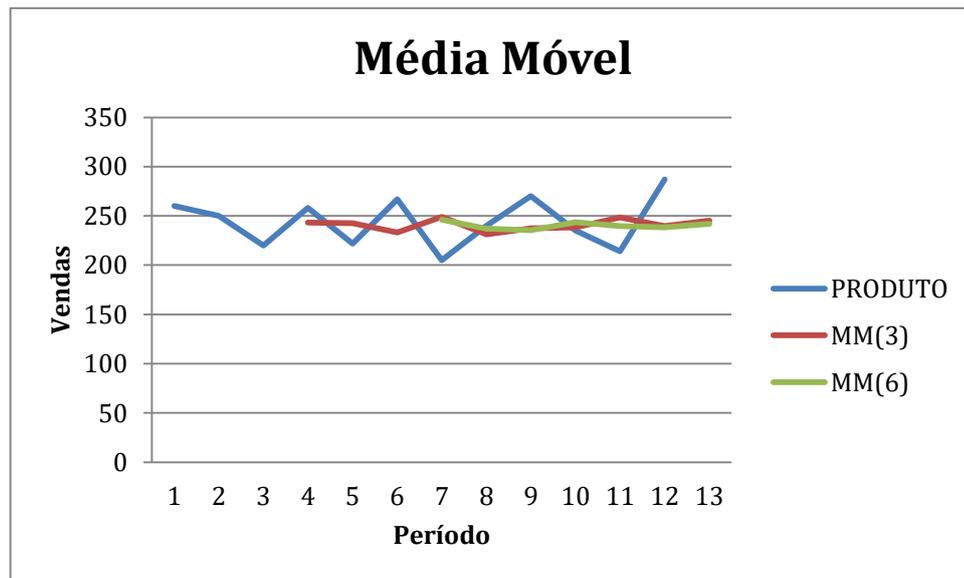


Fonte: De autoria própria.

Em seguida foi aplicada a media móvel, com períodos de previsão $n=3$ e $n=6$. No Anexo F está na primeira coluna o período de tempo da demanda, na segunda coluna a quantidade de demanda, na terceira coluna a média móvel com período igual a 3 e na última coluna a média móvel com período igual a 6.

Na Figura 4, está aplicado graficamente o modelo de média móvel para $n=3$ e $n=6$ em comparação com os valores originais da demanda.

Figura 4: Gráfico da demanda do produto com suas médias móveis.



Fonte: De autoria própria.

3.2.2 REGRESSÃO LINEAR

A análise de regressão consiste em uma análise estatística para determinar a existência de relação funcional entre uma variável dependente e uma ou mais variáveis independentes. Em essência, busca-se obter uma equação que descreva como a variação da variável dependente é influenciada pelas mudanças nos níveis das variáveis independentes (PETERNELLI, 2019).

Segundo Chase (2009) o termo regressão linear simples se refere à regressão de apenas uma variável Y em uma única variável X, que é representada pela equação da reta.

$$Y = \alpha + \beta x$$

α = intercepto

β = coeficiente de inclinação da reta

$$\beta = \frac{Cov(X, Y)}{Var(X)} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\alpha = \frac{(\sum Y_i - \beta \sum X_i)}{n}$$

O coeficiente de correlação é uma função muito importante para a criação de uma regressão linear, pois com ela pode-se encontrar relação entre as variáveis. Além disso, é a partir do coeficiente correlação que podemos encontrar a direção da relação desses dados, se o coeficiente for positivo a inclinação seguirá em um sentido positivo, sendo inversamente caso for negativo. Se a correlação for próxima à zero, a regressão linear não terá um padrão significativo (CHASE, 2009).

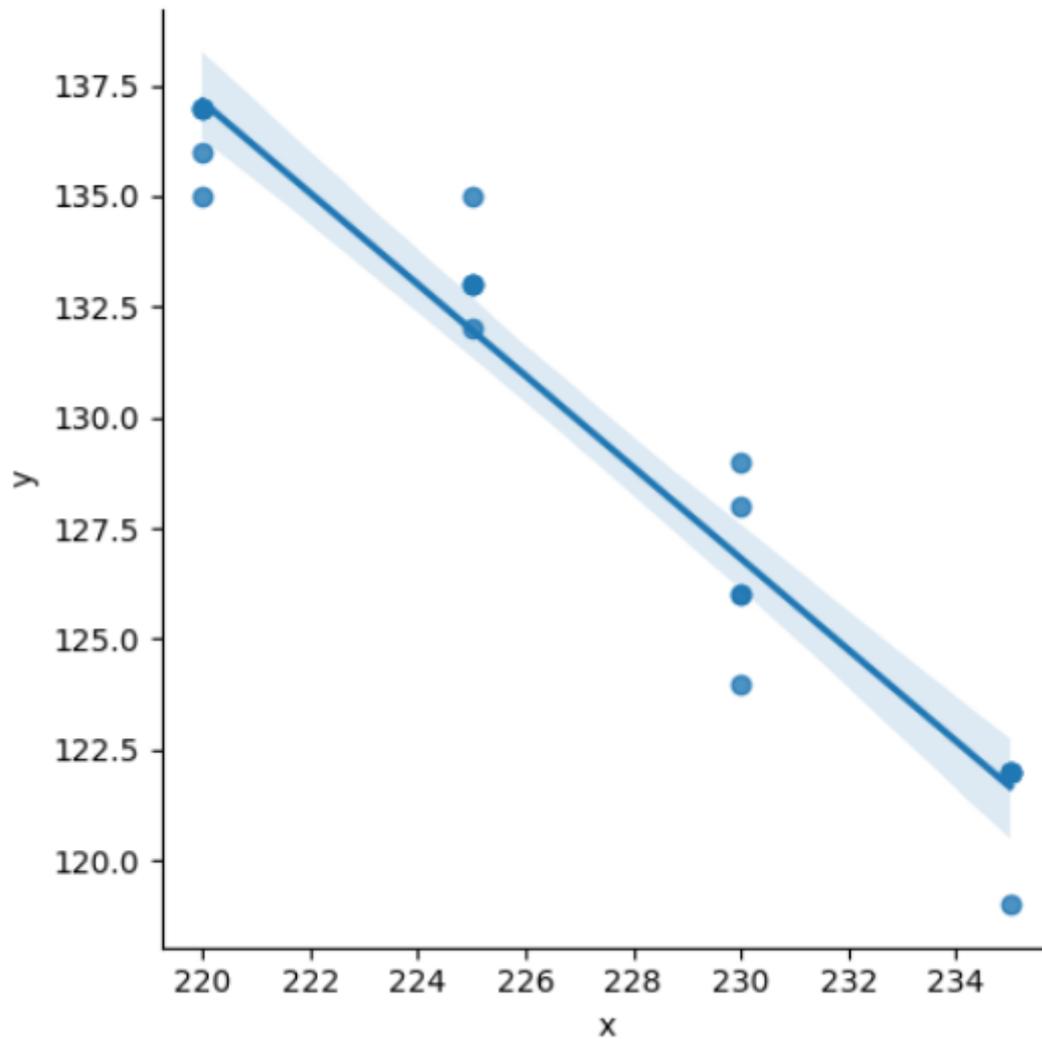
Como exemplo, uma amostra possui vinte variáveis, que estão representados no Anexo G, que serão utilizados para elaborar a regressão linear. Para esses dados temos uma correlação de aproximadamente -97%, ou seja, uma correlação negativa muito forte, o que viabiliza a regressão linear.

Utilizando as fórmulas com os valores de X e Y, é obtido o intercepto com valor de 364,18 e o coeficiente de inclinação da reta sendo -1.032. Então, a equação agora com os valores de α e β é:

$$Y = 364,18 - 1.032X$$

Então a reta, a partir do gráfico de dispersão com o conjunto X e Y será representada pela Figura 5.

Figura 5: Gráfico de dispersão com a reta da regressão linear.



Fonte: De autoria própria.

A partir da Figura 5, pode-se observar que para esse exemplo, a reta é decrescente, já que o coeficiente de correlação era negativo e como sua dispersão possui uma correlação muito forte, o modelo de regressão linear se ajustou no sentido dos dados de amostra.

3.2.3 ÁRVORE DE DECISÃO

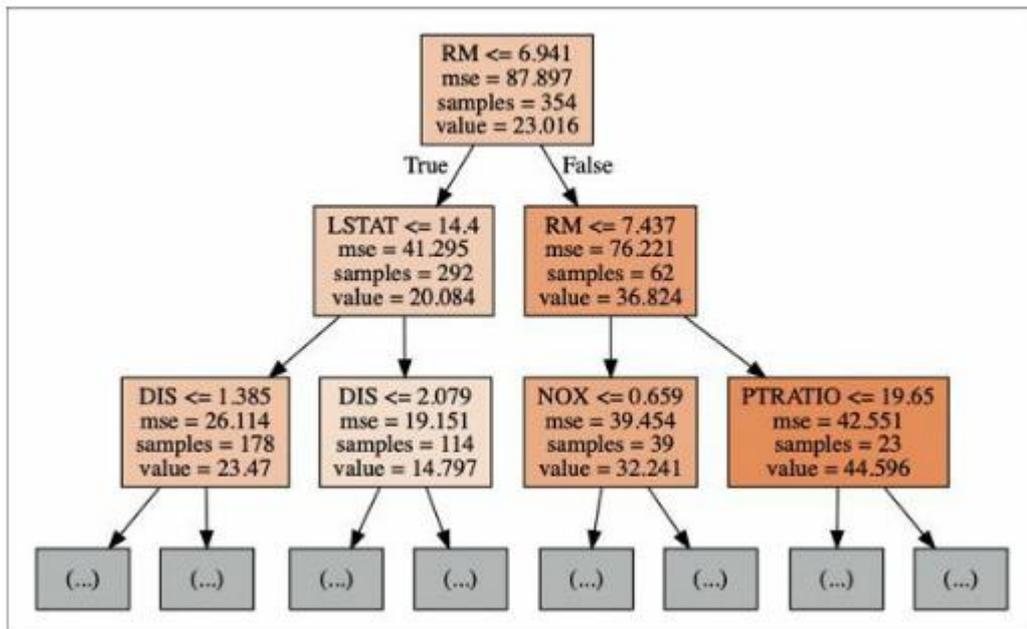
Uma árvore de decisão é representada pela estrutura de árvore, onde possui um determinado número de caminhos possíveis de decisão e há o resultado de cada um deles. É como se buscasse solucionar um problema através de perguntas para prever seu resultado, que

podem ser por dados categóricos (em caso de classificação) ou dados numéricos (GRUS, 2021).

Esse modelo tornou-se referência, desde que J. Ross Quinlan, pesquisador em aprendizado de máquina, o desenvolveu, vem sendo comparado com outros modelos supervisionados. E seus algoritmos (CART e ID3) são desenvolvidos de uma maneira recursiva de divisão e conquista de cima para baixo (HAN *et al.*, 2012).

A árvore de decisão possui diversos pontos positivos como sua facilidade de entendimento e interpretação, além de seu processo de previsão completamente transparente. Como é possível verificar através da Figura 6 (GRUS, 2021).

Figura 6: Árvore de decisão regressiva.



Fonte: HARRISSON, p. 194, 2020.

Outros pontos vantajosos a respeito desse modelo são: a pouca preparação necessária para seus dados, suporte tanto para dados não numéricos como os dados numéricos, suporte para lidar com relacionamentos não lineares (HARRISSON, 2020).

Em relação a sua utilização como estrutura regressiva, a árvore de decisão possui várias separações em seus atributos que são avaliadas. A separação que possuir um menor erro será escolhida. Esse modelo acaba sendo totalmente dependente dos dados de treinamento, já que uma pequena mudança pode acabar modificando toda estrutura da árvore (HARRISSON, 2020).

É possível observar através da Figura 6, que os erros estão representados pelo MSE (Erro quadrático médio), então a “folha” que apresentar o menor MSE será a escolhida.

3.2.4 FLORESTA ALEATÓRIA

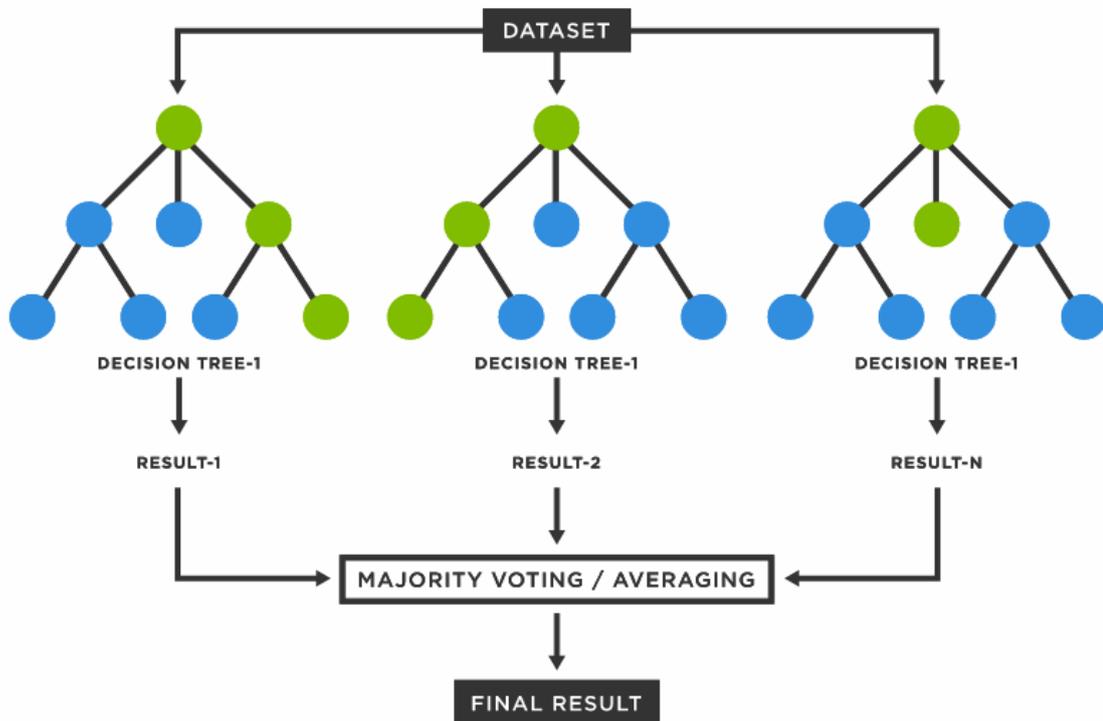
O algoritmo de Floresta Aleatória é um conjunto de árvores de decisões que buscam corrigir a tendência das árvores à super adequação. Quando é criada uma série de árvores e são treinadas com subamostras e atributos aleatórios, a variância é reduzida (HARRISON, 2020).

Para corrigir a tendência da árvore de decisão é usado o bagging ou bootstrap. Segundo Grus (2021) um exemplo da utilização do bootstrap é quando se tem uma amostra cuja média é 100, se todos os pontos estiverem próximos a 100, então a mediana será próxima a 100. Contudo, se metade dos dados estiverem próximos a 0 e a outra metade estiverem próximas a 200, a mediana não parece estar bem determinada.

No caso da floresta aleatória, o bootstrap é usado para corrigir essa tendência da utilização de uma árvore específica. Quando são criadas várias árvores, elas são diferentes umas das outras, trazendo como benefício da viabilidade do uso do conjunto total dos dados (GRUS, 2021).

“A intuição para compreender o bagging vem de um artigo de 1785 do Marquês de Condorcet. Essencialmente, ele diz que, se você estiver criando um júri, deve adicionar qualquer pessoa que tenha uma chance maior que 50% de dar um veredicto correto e então tirar a média das decisões. Sempre que adicionar outro membro (e seu processo de seleção for independente dos demais), um resultado melhor será obtido.” (HARRISON, p. 123, 2020).

Figura 7: Uma ilustração de uma Random Forest.



Fonte: TIBCO.

A Figura 7 é um exemplo de uma Floresta Aleatória, onde é possível observar diversas árvores de decisão, que são processadas para obtenção de um resultado. Em relação a regressão é calculado a média das previsões.

4 ESTUDO DE CASO

O estudo de caso foi desenvolvido a partir dos dados de uma empresa de moto peças do interior de Pernambuco. A empresa possui um ano de existência e trabalha com o varejo de peças e serviço de montagem e revisão de motos. Em relação à sazonalidade do mercado, foi possível encontrar alguns padrões do seu público, como uma demanda maior em meses após festividades, como julho, já que em junho o São João é muito forte na região e em Janeiro, pós Natal e ano novo.

O objetivo é, através da utilização dos dados de faturamento, utilizar os modelos de predição e comparar seus resultados, tentando assim, encontrar as maneiras mais adequadas para prever a demanda da empresa.

Para isso, será utilizada uma base de dados com o faturamento diário, tanto pela venda de peças quanto os serviços que a oficina da loja possui, durante o período de doze meses e serão utilizados os modelos de média móvel, regressão linear, árvore de decisão e floresta aleatória. Em seguida, serão comparados os erros médios absolutos como forma de avaliar o melhor modelo para essa determinada situação.

Fórmula do erro médio absoluto:

$$\text{Erro médio absoluto} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

n = número de amostras

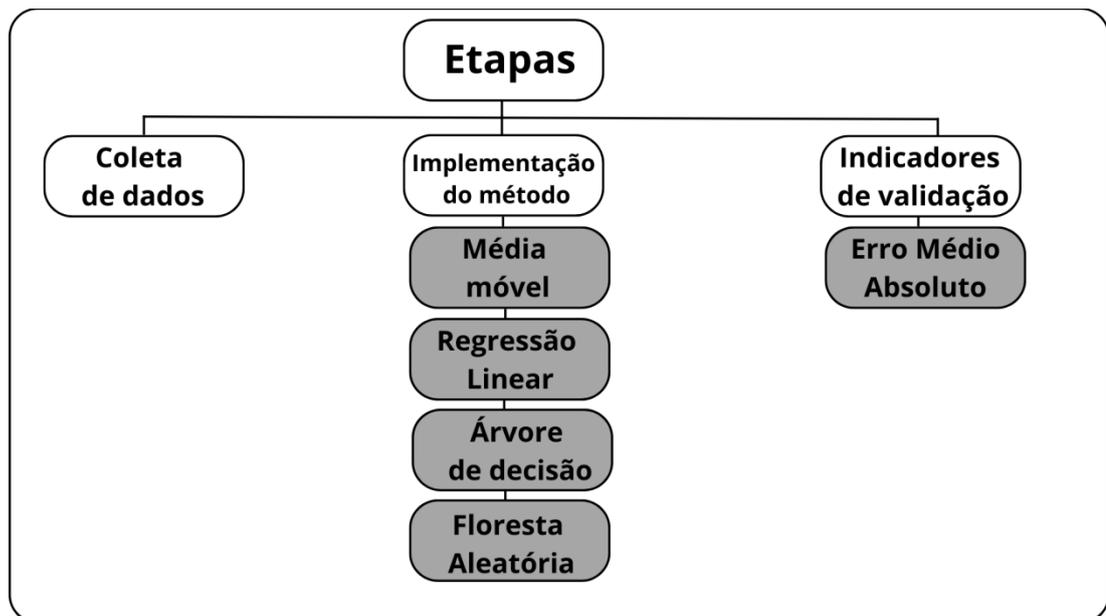
y = valor observado para cada amostra

\hat{Y} = valor previsto pelo modelo para cada amostra

Para fazer a representação gráfica será utilizado o gráfico de linha, que segundo Knaflic (2019) são as representações mais adequadas para tratar de dados em relação a um período.

A figura 8 apresenta todas as etapas realizadas para o projeto, sendo a primeira etapa a coleta de dados, na segunda etapa a implementação dos quatro modelos escolhidos para comparativo, que por fim, será realizada a etapa de validação com o erro médio absoluto.

Figura 8: Etapas do projeto.



Fonte: De autoria própria.

Para a pesquisa foi feita a coleta, onde foram analisados os dados diários de vendas, tanto das peças quanto dos serviços realizados na empresa, em seguida foram aplicados os algoritmos dos métodos (média móvel, regressão linear, árvore de decisão e floresta aleatória), em seguida, foi avaliado seus desempenhos através de erro médio absoluto. Por fim, foi feita uma análise geral para comparar os algoritmos e ver qual método teve o melhor desempenho.

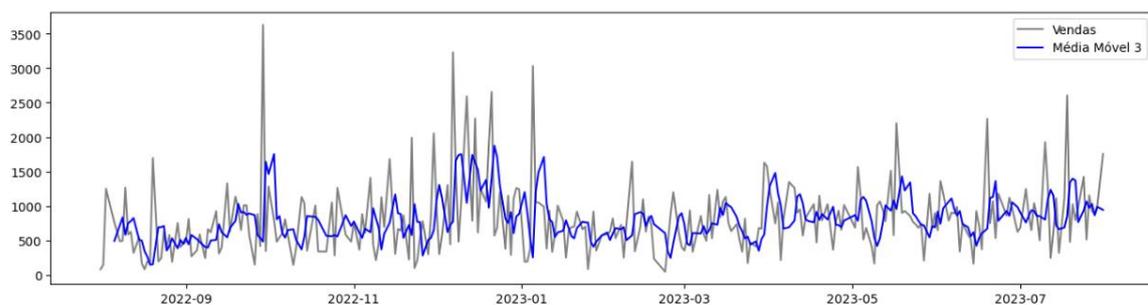
Em relação aos modelos, a média móvel foi feita através do Excel, enquanto as demais foram desenvolvidas a partir da programação em Python. Entre as bibliotecas utilizadas tanto nos modelos quanto nos gráficos foram utilizadas: Pandas, Scikit learn e matplotlib.

5 ANÁLISE DOS RESULTADOS

Foram utilizados os dados do faturamento dos últimos doze meses para fazer uma análise dos resultados da previsão do faturamento com os modelos. O primeiro a ser observado foi o modelo de média móvel, que foi visto como comparação tanto a média móvel com um período de três meses, quanto com seis meses.

Na Figura 9, pode-se verificar a média móvel, como a linha em azul, de três meses, enquanto a linha em cinza são os dados reais.

Figura 9: Média móvel com o período de três meses.

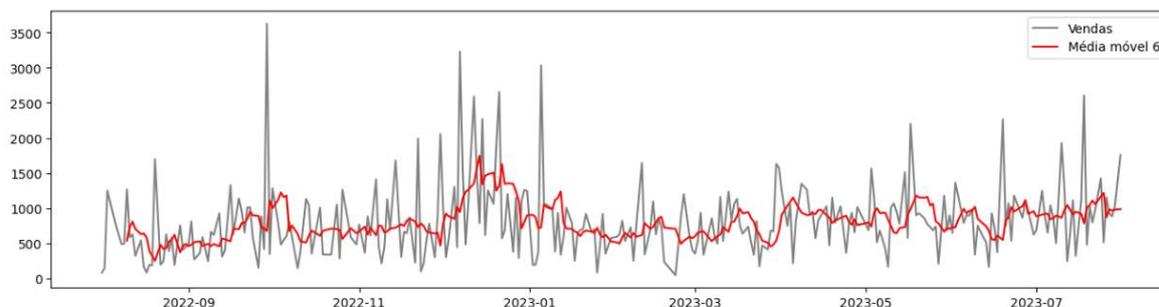


Fonte: De autoria própria.

É possível observar que a trajetória dos dados está correlacionada com os meses anteriores e há ainda presente uma fragilidade aos outliers. Em relação ao seu erro médio absoluto, a média móvel com o período de três meses obteve um valor 407,83, um resultado bem considerável para esse problema específico.

Como um possível comparativo para o próprio modelo, foi utilizado uma média móvel de seis meses, que os resultados podem ser observados a partir da Figura 10, onde a linha de vermelho é o resultado do modelo e a linha em cinza são os dados reais.

Figura 10: Média móvel com o período de seis meses.



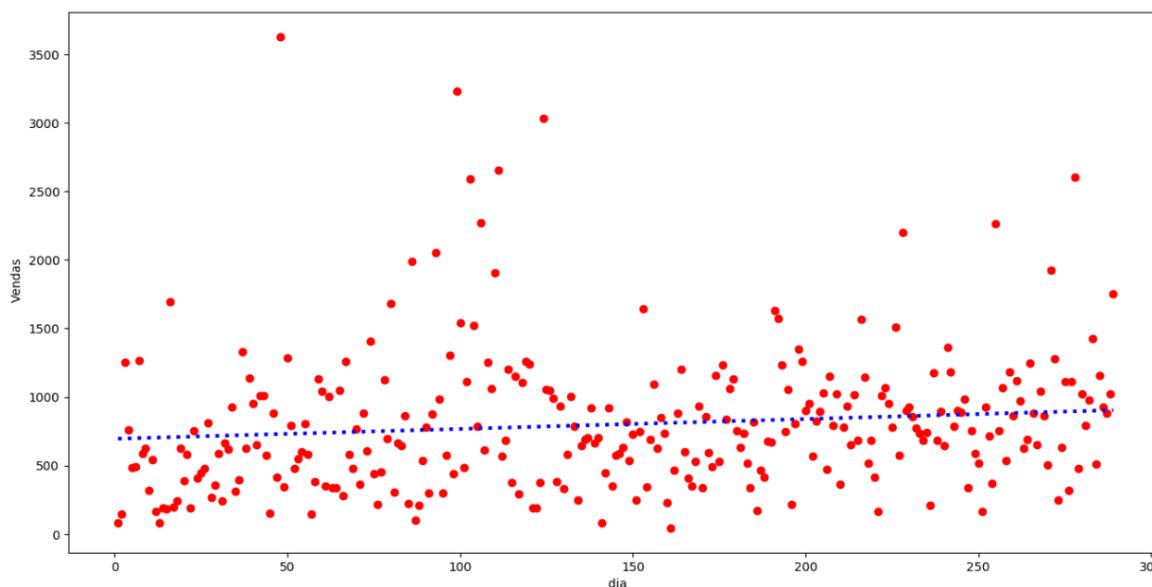
Fonte: De autoria própria.

Em relação ao seu erro médio absoluto, a média móvel com o período de seis meses obteve como resultado um valor de 376,42, ou seja, um valor de 31,41 melhor que o modelo anterior. Outro ponto importante é que essa segunda média móvel teve um impacto menor aos outliers, se mantendo mais próximo a um valor médio.

Analisando as duas médias móveis, é possível verificar que o erro apresentado pela média móvel com período é menor visto que há menos fragilidade aos outliers, estes que surgem muito por aplicar as técnicas em relação aos dados diários, onde ocorre picos de vendas em dias específicos, causados por feriados em dias seguintes.

Para a regressão linear o modelo foi introduzido seguindo o algoritmo do Anexo D, onde foi possível obter como intercepto um valor de 694,67 e sua representação gráfica é observada na Figura 11.

Figura 11: Regressão linear com os dados diários.



Fonte: De autoria própria.

Na Figura 11, é possível ver que há uma pequena tendência na elevação das vendas com o passar do tempo, ainda possível observar alguns outliers para quantidades superiores a reta, o que faz com que a maioria dos dados fique abaixo da reta de regressão. Após fazer o cálculo de r^2 foi obtido um valor de aproximadamente 0,02, sendo assim, uma correlação muito fraca positiva. Em relação ao intercepto dessa regressão linear foi obtido um valor de 694,66.

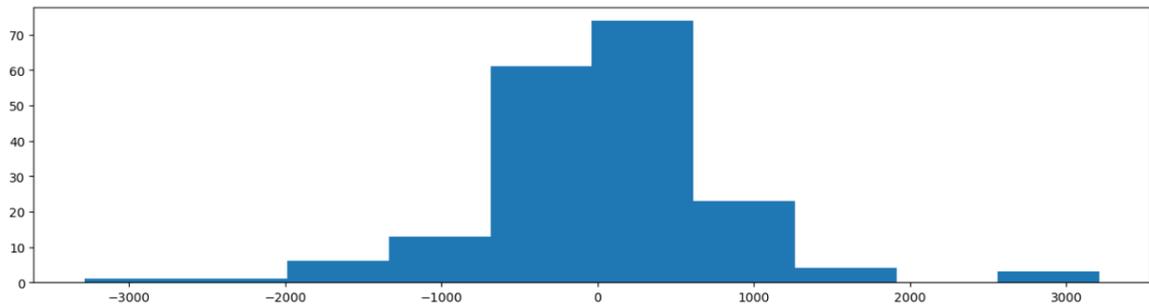
O erro médio absoluto desse modelo para a situação dos dados puros foi de 348,62, um valor que obteve melhores resultados que os de média móvel.

Para a árvore de decisão e a floresta aleatória, foi necessário fazer diferenciações para deixar os dados estacionários. Aplicando o algoritmo de árvore de decisão, com o nível de profundidade de 100, para evitar qualquer problema de sobreajuste. Segundo Grus (2021) deixar os modelos muito complexos podem causar sobreajuste, não generalizando quase nada além dos dados de treinamento e levando o modelo a uma complexidade excessiva.

Para Grus (2021) se o modelo possuir sobreajuste nos dados de treinamento, possivelmente terá um desempenho muito ruim nos dados de teste.

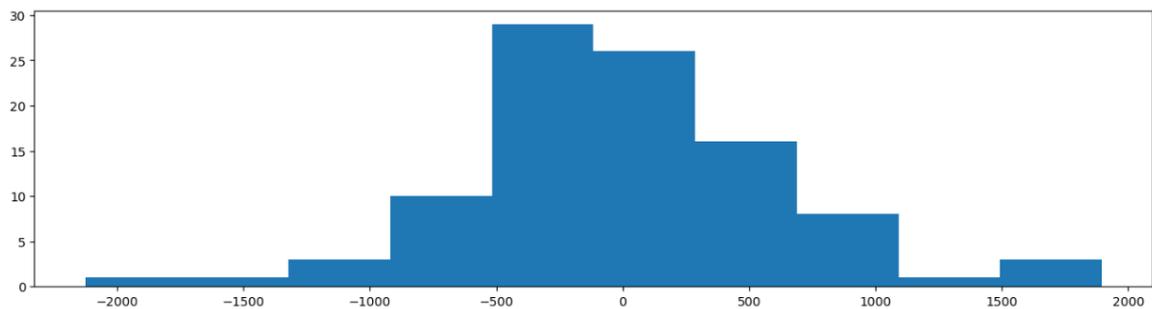
Fazendo a diferenciação é possível observar através da Figura 12 e Figura 13 que os dados obtiveram assim uma distribuição normal.

Figura 12: Distribuição dos dados de treino após a diferenciação.



Fonte: De autoria própria.

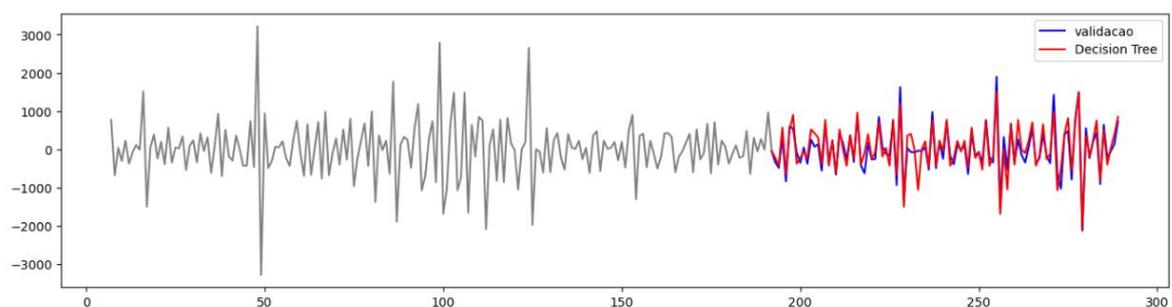
Figura 13: Distribuição dos dados de teste após a diferenciação.



Fonte: De autoria própria.

Após normalizar, foi realizado o algoritmo de árvore de decisão regressiva, onde foi dividido 33% dos dados para teste e 66% para treino, como apresentado pelo Anexo A, e tem como resultado o gráfico da Figura 14, onde os dados em cinza são os de treino, os dados de azul são os de teste e o resultado do algoritmo se apresenta na cor vermelha.

Figura 14: Aplicação do modelo de árvore de decisão.



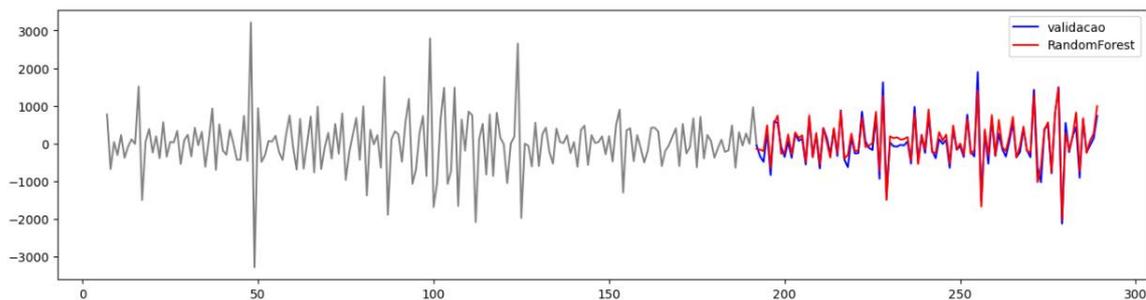
Fonte: De autoria própria.

Com isso, o algoritmo de árvore de decisão, idealizado através do algoritmo do Anexo C, possui um erro médio absoluto de 179,57, que até então é o que possui um melhor

resultado entre os demais demonstrados e o que há um resultado com valores muito próximos aos dados de validação, não sendo tão iguais e também não causando o sobreajuste.

Foram utilizados os mesmos processamentos dos dados da árvore de decisão, assim como está apresentado no Anexo B, para a floresta aleatória, afim de, comparar a eficiência dos dois processos para esse banco de dados. Na Figura 15 é possível observar a proximidade da previsão do modelo em relação aos dados de teste, sendo o algoritmo da floresta aleatória a linha vermelha e os dados de testes a linha azul.

Figura 15: Aplicação do modelo de floresta aleatória.



Fonte: De autoria própria.

Após utilizar a métrica de erro médio absoluto, o resultado obtido foi de um erro no valor de 125,85, o que mostra uma maior precisão e sendo bem próximo aos dados de validação.

Por conseguinte, todos os resultados foram registrados na Tabela 1 para comparação e análise final.

Tabela 1: Resultados dos métodos de predição utilizados.

Modelo	Erro absoluto médio	Primeiro valor previsto
Média Móvel 3	407,83	942,66
Média Móvel 6	376,42	986,33
Regressão Linear diário	348,62	897,4
Árvore de decisão	179,57	846
Floresta Aleatória	125,85	993,45

Fonte: De autoria própria.

O primeiro ponto a ser observado são os primeiros valores previstos com os métodos utilizados, onde estão todos entre valores de 846 até 993,45, mesmo sendo os que mais possuíam erros menores, a árvore de decisão e a floresta aleatória para o primeiro valor

previsto foram os que tiveram resultados mais distintos. Contudo, por ser uma previsão para uma quantidade maior de dias, a média geral dos erros se apresentaram com menores valores.

Em relação ao erro em si, a floresta aleatória obteve uma melhor precisão, chegando a valores muito próximos dos dados de teste com um valor de erro absoluto médio de 53,72 menor que o segundo modelo que obteve os melhores resultados, que foi a árvore de decisão.

Contudo, é possível observar alguns padrões, como por exemplo, o leve aumento no faturamento durante o período. Além disso, por ser uma empresa nova no setor, é normal que haja alguns outliers em seu faturamento, então o erro das médias móveis e da regressão linear diária também são bem justificáveis para esse problema.

6 CONCLUSÃO

O trabalho realizado teve como intuito encontrar o método preditivo mais adequado para prever o faturamento dessa empresa do setor de moto peças. Utilizando os quatro modelos escolhidos para esse projeto foi possível obter algumas conclusões.

Os modelos de média móvel obtiveram resultados bem semelhantes tanto com o período (n) de três meses quanto com seis meses, não diferente disso, o modelo de regressão linear obteve um erro também muito próximo ao de média móvel, se comparando apenas as três alternativas de previsão, qualquer uma que fosse escolhida teria desempenho semelhante.

Contudo, após utilizar os modelos de árvore de decisão e floresta aleatória, os resultados foram bem mais satisfatórios, diminuindo assim o erro pela metade e sendo os mais indicados para prever a demanda nesse caso em específico.

Entretanto, se a pesquisa fosse realizada com um período maior e com mais dados para identificar outros padrões que podem influenciar a demanda dessa empresa, é possível que o resultado pudesse ser um pouco diferente e com uma taxa de erro menor.

7 TRABALHOS FUTUROS

Sugere-se um maior aprofundamento dos dados para trabalhos futuros, com mais informações será possível encontrar uma maior padronização dos dados, diminuindo assim o erro de predição. Outro ponto que auxiliaria ainda mais a pesquisa seria utilizar outros tipos de modelos preditivos, como por exemplo, o ARIMA, SVM, suavização exponencial e redes neurais.

REFERÊNCIAS

- ACKERMANN, A. E. F.; SELMITTO, M. A. Métodos de Previsão de Demanda: Uma Revisão da Literatura. *Innovar: uma Revista de Ciências Administrativas e Sociais*, p. 83-100, 2022.
- BOX, George E. P.; JENKINS, Gwilym M. *Time Series Analysis: Forecasting and Control*. 5ª edição. Ano de publicação: 2016.
- CHASE, C. *Demand-Driven Forecasting: A Structured Approach to Forecasting*. Ano de publicação: 2009.
- CHOPRA, S.; MEINDL, P. *Supply Chain Management: Strategy, Planning and Operation*. 6ª edição. Ano de publicação: 2015.
- GRUS, J. *Data Science do Zero*. 2ª edição. O'Reilly, 2021.
- HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. Burlington, MA: Elsevier, 2012.
- HARRISON, M. *Machine Learning: Guia de Referência Rápida*. O'Reilly, 2020.
- HYNDMAN, R.J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2ª edição. Monash University Publishing, 2018.
- KNAFLIC, C. N. *Storytelling com Dados: Um Guia Sobre Visualização de Dados para Profissionais de Negócios*. Editora Alta Brooks, 2019.
- LUSTOSA, L.; MESQUITA, M. A.; QUELHAS, O.; OLIVEIRA, R. *Planejamento e Controle da Produção*. Rio de Janeiro: Elsevier Editora Ltda, 2008.
- PETERNELLI, Luiz Alexandre. *Regressão Linear e Correlação*. Disponível em: https://edisciplinas.usp.br/pluginfile.php/176798/mod_resource/content/1/Capítulo%20-%20Regressão%20Linear%20Simples.pdf. Acesso em: 20 de julho de 2023.

PROVOST, F.; FAWCETT, T. Data Science for Business: What You Need to Know About Data Mining and Data Analysis. Beijing, O'Reilly, 2013.

RUEDA, R. A. S.; ORTEGA, J. R.; RAMÍREZ, S. M. M.; ZAMORANO, C. A. Uso de los algoritmos Machine Learning para analizar Moodle y los teléfonos inteligentes em el processo educativo de la Física. Universidad Nacional Autónoma de México, Instituto de Ciencias Aplicadas y Tecnologia. Cidade do México, 2023.

SALES, R. N. B.; FERREIRA, T. R. O.; SOARES, T. F. A importância da previsão de demanda nos processos de decisão estratégica empresarial: Um estudo de caso em uma empresa de confecção em Caruaru-PE. ASCES. 2019

TUBINO, D. F. Manual de Planejamento e Controle da Produção. 2ª edição. São Paulo: Atlas, 2000.

WITTEN, I.H.; FRANK, E.; HALL, M.A. Data Mining: Practical Machine Learning Tools and Techniques. 3ª edição. San Francisco, Estados Unidos da América: Morgan Kaufmann, 2011.

WITTEN, I.H.; FRANK, E.; HALL, M.A. Data Mining: Practical Machine Learning Tools and Techniques. 4ª edição. San Francisco, Estados Unidos da América: Morgan Kaufmann, 2016.

ANEXO A – DIVISÃO DE DADOS PARA FLORESTA ALEATÓRIA E ÁRVORE DE DECISÃO

```
backtest_size = 0.66
train = df3.drop(['Data'], axis = 1).reset_index().loc[6:np.round(df3.shape[0]*backtest_size),:]
test = df3.drop(['Data'], axis=1).reset_index().loc[np.round(df3.shape[0]*backtest_size):,:]

x_cols = ['Valor_f', 'lag_diff_vendas_1', 'lag_diff_vendas_2',
          'lag_diff_vendas_3', 'lag_diff_vendas_4', 'lag_diff_vendas_5']
y_col = 'diff_vendas'

x_train, x_test = train[x_cols], test[x_cols]
y_train, y_test = train[y_col], test[y_col]
```

ANEXO B – ALGORITMO DA FLORESTA ALEATÓRIA

```
from sklearn.ensemble import RandomForestRegressor
rfr = RandomForestRegressor(max_depth=100, random_state=42, criterion = 'mae', verbose = 0)
rfr.fit(x_train, y_train)
print('R2 da Random Forest no teste: {:.2f}'.format(rfr.score(x_test, y_test)))
test['pred_RF'] = rfr.predict(x_test)
```

ANEXO C – ALGORITMO DA ÁRVORE DE DECISÃO

```
from sklearn.tree import DecisionTreeRegressor
dtr = DecisionTreeRegressor(max_depth=100, random_state=42, criterion = 'mae')
dtr.fit(x_train, y_train)
print('R2 da Decision Tree no teste: {:.2f}'.format(dtr.score(x_test, y_test)))
test['pred_DT'] = dtr.predict(x_test)
```

ANEXO D – ALGORITMO DA REGRESSÃO LINEAR

```
#x = df['dia']  
y = data.valor  
x = data[['dia', 'unique_id']]
```

```
# usando uma distribuição de 66% dos dados de treino e 33% dados de teste  
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.33, random_state=0)
```

```
lr = LinearRegression()
```

```
#treinando  
lr.fit(x_train, y_train)
```

```
LinearRegression()
```

```
#Verificando o resultado  
r_sq = lr.score(x, y)  
print('Coeficiente de Determinação (R²):', r_sq)
```

```
Coeficiente de Determinação (R²): 0.022284883384261045
```

```
#verificando o intercepto, valor de b0 da equação  
print('Intercepto:', lr.intercept_)
```

```
Intercepto: 694.6699374559096
```

ANEXO E – DEMANDA DO PRODUTO

Período	1	2	3	4	5	6	7	8	9	10	11	12	Média
Produto	260	250	220	258	222	267	205	240	270	235	214	287	244

ANEXO F - DEMANDA COM SUAS MÉDIAS MÓVEIS DE PERÍODOS IGUAIS A 3
E 6

PERÍODO	PRODUTO	MM(3)	MM(6)
1	260		
2	250		
3	220		
4	258	243,33	
5	222	242,67	
6	267	233,33	
7	205	249,00	246,17
8	240	231,33	237,00
9	270	237,33	235,33
10	235	238,33	243,67
11	214	248,33	239,83
12	287	239,67	238,50
13		245,33	241,83

ANEXO G – AMOSTRA COM VALORES DE X E Y

X	Y
220	137
220	137
220	137
220	136
220	135
225	135
225	133
225	132
225	133
225	133
230	128
230	124
230	126
230	129
230	126
235	122
235	122
235	122
235	119
235	122