

**Universidade Estadual Paulista**  
**Departamento de Computação**  
**Curso de Bacharelado em Sistemas de Informação**

**Mayara Canaver e Lopes**

**Análise de Séries Temporais e Previsão de Lucro e/ou  
Perda nas ações da empresa AMBEV S.A**

**Trabalho de Conclusão de Curso**

**Bauru**

**2022**

**Mayara Canaver e Lopes**

**Análise de Séries Temporais e Previsão de Dados da Ação  
da Ambev na Bolsa de Valores**

Proposta apresentada à disciplina de Trabalho de Conclusão de **Curso do Bacharelado em Sistemas de Informação**, da Universidade Estadual Paulista, como requisito parcial para obtenção do título de bacharel em Sistemas de Informação.

**Orientador:** Prof. Antonio Fernando Crepaldi

**Bauru**

**2022**

Lopes, Mayara Canaver e.  
Análise de Séries Temporais e Previsão de  
Lucro e/ou Perda nas ações da empresa AMBEV S.A  
/ Mayara Canaver e Lopes, 2023  
43 f. : il.

Orientador: Antonio Fernando Crepaldi

Monografia (Graduação)-Universidade Estadual  
Paulista (Unesp). Faculdade de Ciências, Bauru,  
2023

1. Análise de Séries Temporais. 2. Bolsa de  
Valores. 3. Análise de Dados. I. Universidade  
Estadual Paulista. Faculdade de Ciências. II.  
Título.

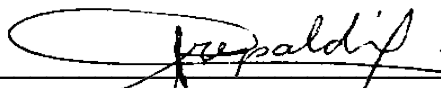
**Mayara Canaver e Lopes**

**Análise de Séries Temporais e Previsão de Dados da Bolsa  
de Valores**

Proposta apresentada à disciplina de Trabalho de Conclusão de **Curso do Bacharelado em Sistemas de Informação**, da Universidade Estadual Paulista, como requisito parcial para obtenção do título de bacharel em Sistemas de Informação.

**Orientador:** Prof. Antonio Fernando Crepaldi

**BANCA EXAMINADORA**



---

Prof. Antonio Fernando Crépaldi (Orientador)

---

Prof. José Remo Ferreira Brega

Digitizado com CamScanner

---

Profa. Enzo Barberio Mariano

**Bauru**

**2022**

*Dedico este trabalho aos meus pais e amigos próximos a quem pude ter o privilégio de gozar da vida e poder ter os melhores incentivos que poderia ter para a realização do mesmo.*

## AGRADECIMENTOS

*Agradeço primeiramente aos meus pais, pela oportunidade de estudo e incentivo para eu poder estar onde estou*

*Agradeço aos meus amigos que me ajudaram em todo o processo de desenvolvimento deste trabalho e pela companhia em diversos momentos*

*Agradeço ao meu orientador Antonio Fernando Crepaldi de trabalho pela ajuda e acompanhamento do mesmo e assim chegar a conclusão*

*Agradeço ao professor João Paulo Papa pela citação e engajamento sobre a área de ciência de dados, pois graças a isso pude entrar na área e acabar me aperfeiçoando*

Agradeço ao tempo também, pelo processo que levei para poder realizar esse trabalho e todo o entendimento e crescimento de carreira e principalmente pessoal que obtive durante todo esse caminho até aqui.

## RESUMO

A bolsa de valores é um mercado de grande magnitude, cujo impacto afeta diretamente milhões de pessoas em todo o mundo. Devido à sua natureza global e alta volatilidade, caracterizada pelas constantes variações nos valores das ações, torna-se extremamente desafiador prever o comportamento desses ativos. A volatilidade dos valores, eventos geopolíticos adversos, como conflitos e crises, e fatores socioeconômicos internos podem influenciar significativamente o valor das ações na bolsa. Muitos investidores assumem grandes riscos na tentativa de obter lucro ou uma renda extra mensal. Diante dessa realidade, surgem metodologias que buscam facilitar a tomada de decisões e aumentar as chances de lucratividade. A aplicação de técnicas de machine learning e análise de dados tem se tornado cada vez mais comum no cotidiano das pessoas. A partir dos dados, é possível extrair informações valiosas e gerar ideias lucrativas. Nesse contexto, surge o questionamento sobre como a análise de dados e modelos preditivos podem auxiliar no mercado de ações. Poderíamos formar conjuntos de dados históricos e analisar seus padrões ao longo do tempo? Seria possível formular e testar várias hipóteses nesse sentido. Com base nesse questionamento, este trabalho propõe a análise de dados do mercado de ações, com foco específico na AMBEV. Por meio da análise de séries temporais e aplicação de modelos preditivos, busca-se identificar a melhor estratégia de tomada de decisão e estimar os lucros potenciais. Para alcançar esse objetivo, a coleta de dados é realizada inicialmente, seguida pela observação desses dados utilizando modelos de séries temporais e, posteriormente, pela aplicação de modelos preditivos. Esses modelos podem fornecer insights valiosos, como a identificação das ações mais viáveis, a avaliação dos riscos envolvidos e a estimativa da margem de lucratividade associada a cada uma delas. Dessa forma, a análise de dados e modelos preditivos se mostram ferramentas promissoras para auxiliar os investidores na tomada de decisões mais informadas e estratégicas no mercado de ações.

**Palavras-chave:** AMBEV. Análise de Dados. Bolsa de Valores. Machine Learning. Modelos Preditivos. Séries Temporais

## ABSTRACT

The stock market is a trillion-dollar market that directly and indirectly affects millions of people worldwide. Being a global and highly volatile market, with stock values fluctuating from minute to minute, predicting the value of a particular stock is extremely challenging. Factors such as volatility, geopolitical events (e.g., wars and conflicts), and internal social issues can significantly impact the value of stocks. Many shareholders take high risks to generate profits or even earn extra income at the end of the month. Based on this, various methods can be considered to facilitate actions and increase the likelihood of profit. The use of machine learning combined with data analysis is increasingly present in people's daily lives. With data, there is information, and with information, new ideas for profit can be generated. The question arises: How can data analysis and predictive models help us with the stock market? What if we collect a dataset and observe its patterns over time? Multiple hypotheses can be formulated and tested. Based on this question, this work proposes the analysis of stock market data, specifically focusing on AMBEV, and using time series analysis along with predictive models to determine the best decision-making approach for estimating potential profits. To achieve this goal, data collection is initially performed, followed by data observation using a time series model, and subsequently, predictive models are applied. This approach can provide a range of final observations, such as identifying the most viable actions, assessing their associated risks, and estimating their profitability margins. In summary, data analysis and predictive models present promising tools to assist investors in making informed and strategic decisions in the stock market.

**Keywords:** AMBEV. Data Analysis. Machine Learning. Predictive Models. Stock Exchange. Time Series



## LISTA DE FIGURAS

Figura 1 – Demanda Mensal de Gasolina em Ontário.....	14
Figura 2 – Fluxo de Trabalho na Área de Dados.....	17
Figura 3 – Requisitos de Desempenho.....	19
Figura 4 – Dados da Ação ABEV3.SA.....	22
Figura 5 – Tipagem das Colunas do Conjunto de Dados.....	23
Figura 6 – Valores de Retorno Semanais Durante o Período Analisado.....	25
Figura 7 – Resultado Teste ADFuller.....	27
Figura 8 – Conjunto de Dados Após Aplicação da Diferenciação.....	28
Figura 9 – Resultado Teste ADFuller Após Aplicação da Diferenciação.....	28
Figura 10 – Autocorrelação.....	31
Figura 11 – Autocorrelação Parcial.....	31
Figura 12 – Janela Móvel.....	33
Figura 13 – Predição Geral da Série.....	35
Figura 14 – Predição Específica da Série.....	35
Figura 15 – Variância do Modelo GARCH.....	38

## LISTA DE ABREVIATURAS E SIGLAS

ACF.....	Função de Autocorrelação
ADF.....	Teste de Dickey-Fuller
AIC.....	Critério de informação de Akaike
ARIMA .....	Autorregressivo Integrado de Médias Móveis
MAE.....	Erro médio absoluto
MAPE .....	Erro Percentual Absoluto Médio
ML.....	Aprendizado de Máquina
MSE.....	Erro Quadrático Médio
PACF.....	Função de Autocorrelação Parcial
RMSE.....	Raiz quadrada do erro-médio

# SUMÁRIO

<b>RESUMO</b>	<b>7</b>
<b>ABSTRACT</b>	<b>8</b>
<b>LISTA DE FIGURAS</b>	<b>9</b>
<b>LISTA DE ABREVIATURAS E SIGLAS</b>	<b>10</b>
<b>SUMÁRIO</b>	<b>11</b>
<b>INTRODUÇÃO</b>	<b>12</b>
1.1 Objetivo	12
1.2 Justificativa	13
<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>13</b>
2.1 Séries Temporais	14
2.2 Estatística	15
2.3 Machine Learning	16
<b>MÉTODO</b>	<b>18</b>
3.1 Requisitos Funcionais	19
3.2 Requisitos Não-Funcionais	19
3.2.1 Desempenho	20
3.2.2 Interoperabilidade	20
3.3 Coleta de Dados	22
<b>DESENVOLVIMENTO</b>	<b>25</b>
4.1 Processamento dos Dados	25
4.1.1 Tratamento dos Dados	25
4.1.1.1 Etapas Estatísticas	27
4.1.1.2 Teste de Dickey Fuller (ADF)	27
4.1.1.3 Diferenciação dos Dados	28
4.1.1.4 Teste de Shapiro-Wilk	29
4.1.1.5 Transformação Box-Cox	30
4.1.1.6 ACF e PACF	31
4.1.1.7 Estudo de Médias Móveis	33
4.2 Modelagem dos Dados	34
4.2.1 ARIMA	34
4.1.2.5 AIC	37
4.2.2 GARCH	37
4.3 Métricas para Verificação do Modelo	40
4.3.1 Mean Absolute Percentage Error (MAPE)	40
4.3.2 Mean Absolute Error (MAE)	40
4.3.3 Mean Squared Error (MSE)	41
4.3.4 Root Mean Squared Error (RMSE)	41
4.4 Resultados	41
<b>CONCLUSÃO</b>	<b>43</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>44</b>

## INTRODUÇÃO

O mercado financeiro é um ambiente complexo e volátil, onde os preços das ações podem variar significativamente em curtos períodos de tempo. Isso dificulta para os investidores e acionistas tomarem decisões acertadas sobre onde alocar seus recursos. Assim é pensado na necessidade de solucionar o problema mencionado.

A relevância da resolução da proposta realizada é alta para os grupos de investidores e acionistas, além de entusiastas do estudo do meio financeiro investidor e comercial. Com a previsão das futuras ações é possível estipular quanto será a perda ou o ganho em cima de uma ação.

Para a realização desse projeto, foram utilizadas técnicas estatísticas, com ênfase em séries temporais, além da realização de testes de validação e métricas para averiguar qual o melhor resultado e desempenho obtido no geral.

A análise de séries temporais consiste em estudar um conjunto de dados em um determinado período de tempo e assim tomar uma decisão a partir do resultado obtido. Como há diversas ações no mercado financeiro brasileiro, optou-se por uma das mais conhecidas, a AMBEV, por ser líder no mercado brasileiro no setor de cervejaria sendo a 14º maior empresa do país em receita líquida, e realizou-se um estudo de séries temporais inicialmente. Após o estudo, os dados foram analisados e tratados para serem ajustados como parâmetros do modelo proposto, assim o modelo prediz qual a próxima possível valorização de determinada ação.

Os objetivos foram restritos, tais como realizar a predição da ação da bolsa para possíveis apostas e investimentos futuros, a fim de obter lucro no setor de investimento financeiro.

Os recursos foram obtidos com uma busca em sites específicos sobre a bolsa de valores e ações, e para montagem do conjunto dos dados temporais da ação foram utilizadas bibliotecas específicas estatísticas e de limpeza de dados, com a linguagem de programação Python, estudos e ajustes necessários foram feitos para a obtenção do melhor resultado.

### 1.1 Objetivo

O objetivo principal do trabalho de conclusão de curso é a predição da ação da bolsa de valores AMBEV, assim seu resultado é uma base de qual momento desta ação poderá beneficiar ou não o usuário que utilizará o resultado da predição.

## **1.2 Justificativa**

O trabalho em questão possui relevância significativa no âmbito da divulgação de informações sobre o mercado de ações e estratégias de investimento por meio da aplicação de modelagem estatística. Ele visa auxiliar indivíduos e profissionais do setor financeiro na tomada de decisões mais fundamentadas, contribuindo para evitar investimentos desnecessários que possam acarretar perdas financeiras.

Para o estudante universitário, a execução e os resultados deste trabalho representam a oportunidade de colocar em prática os conhecimentos adquiridos durante o curso de Sistemas de Informação, aplicando as teorias estudadas ao longo do programa acadêmico e explorando tópicos complementares por meio de projetos desenvolvidos durante períodos de tempo livre.

## FUNDAMENTAÇÃO TEÓRICA

### 2.1 Séries Temporais

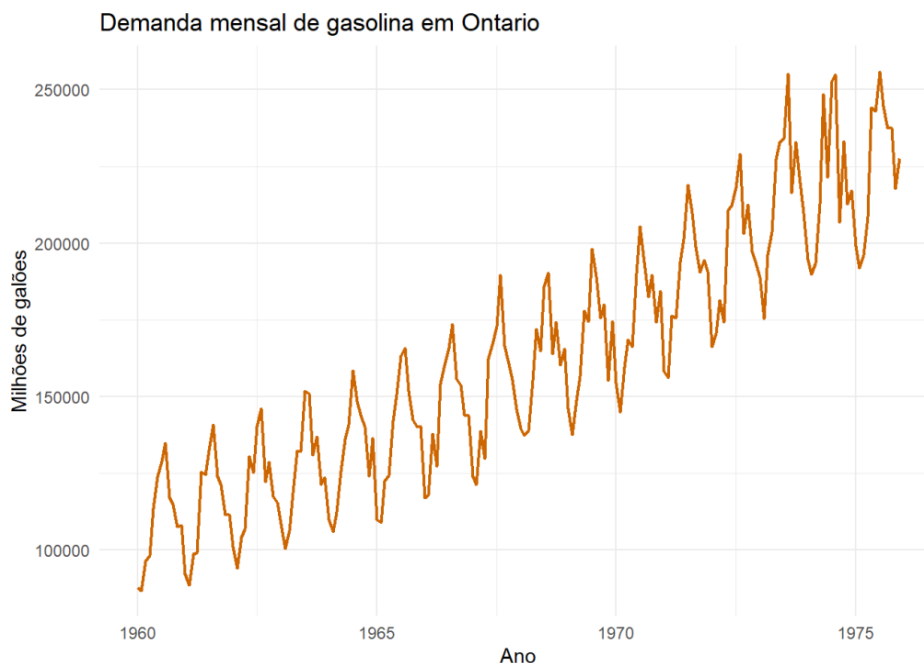
“As séries temporais são um conjunto de observações que podem ser feitas sequencialmente ao longo de um determinado tempo e frequência” dizia Brockwell. As observações são feitas de modo dependente uma da outra, o que faz essa ser a principal característica de uma série temporal, porém outras características igualmente relevantes também estão presentes.

Uma das primeiras etapas na análise de séries temporais é a decomposição da série, o que permite separar o comportamento da série em componentes, tais como tendência, sazonalidade, ciclos e ruído aleatório (Shumway & Stoffer, 2006).

Outro método comum utilizado na análise de séries temporais é o modelo ARIMA (AutoRegressive Integrated Moving Average), que é usado para prever valores futuros com base em valores anteriores da série (Hyndman & Athanasopoulos, 2018).

Para uma melhor descrição desses comportamentos, a figura 1 apresenta um exemplo de uma série temporal da demanda mensal de gasolina em Ontario (eixo Y) baseado nos anos (eixo X) e com ela é possível notar algumas características citadas acima presentes.

FIGURA 1 - DEMANDA MENSAL DE GASOLINA DE ONTÁRIO



FONTE: STATPLACE (2023)

É possível observar uma tendência crescente visualmente na série, com uma sazonalidade com o decorrer do período presente, o que podem ser confirmadas através de medidas estatísticas.

## 2.2 Estatística

A estatística é uma parte fundamental da análise de séries temporais, e sua importância é destacada por autores como Chatfield (2004) ao afirmar que "a estatística desempenha um papel central na análise de séries temporais". Ela é responsável por fornecer informações de formas diretas e indiretas.

Diretamente pode ser vista através de seus métodos para obter resultados e/ou manipular informações compostas nos dados presentes, e na utilização de métricas para resultados de equações com os dados disponíveis.

Indiretamente, a estatística também é importante por proporcionar informações sobre a distribuição dos dados e ajudar na identificação de padrões e tendências. Isso é fundamental para prever comportamentos futuros e tomar decisões informadas, especialmente em relação à bolsa de valores.

Os métodos estatísticos mais comuns utilizados em séries temporais incluem a decomposição de séries temporais, análise de tendências, modelos ARIMA, modelos de séries temporais estocásticas, entre outros. De acordo com Reinsel (1998), esses métodos são amplamente utilizados devido a sua eficácia e facilidade de uso. Além disso, técnicas de aprendizado de máquina também estão ganhando popularidade na previsão de séries temporais, como destacado por Hyndman et al. (2008).

Além da análise de tendências, a estatística também é usada para medir a volatilidade dos dados, o que pode ser útil para identificar pontos de entrada e saída no mercado de ações. A volatilidade é importante para os investidores, pois ajuda a entender o risco envolvido em uma determinada ação ou mercado, como destacado por Black (1976).

Em resumo, a estatística é uma parte vital da análise de séries temporais, tanto na bolsa de valores como em outras aplicações. Ela fornece informações sobre a distribuição dos dados, permite a identificação de padrões e tendências, e ajuda a prever comportamentos futuros.

### **2.3 Machine Learning**

A Fundamentação teórica sobre machine learning está estreitamente ligada com a estatística e as séries temporais. De acordo com Hastie, Tibshirani e Friedman (2009), machine learning é um ramo da estatística que se concentra em desenvolver modelos que são capazes de aprender a partir dos dados. Esse aprendizado é baseado em algoritmos que são treinados a partir de exemplos de dados e que, posteriormente, são capazes de fazer previsões sobre novos dados.

No que diz respeito às séries temporais, machine learning tem uma aplicação ampla e relevante. De acordo com Hyndman e Athanasopoulos (2018), as técnicas de machine learning são amplamente utilizadas para analisar séries temporais e para fazer previsões. Alguns exemplos incluem o uso de redes neurais, florestas aleatórias e modelos de séries temporais baseados em processos estocásticos.



Além disso, outros autores como Lutkepohl (2005) também destacam a importância da integração entre machine learning e séries temporais. Ele afirma que o uso de técnicas de machine learning em séries temporais permite não apenas a análise de tendências e comportamentos, mas também a identificação de padrões e previsões precisas.

## MÉTODO

A área de dados desempenha um papel fundamental em diversos setores, abrangendo desde pesquisas acadêmicas até a tomada de decisões estratégicas em empresas. No entanto, muitos projetos nessa área enfrentam desafios, como a falta de uma estrutura clara e consistente para orientar as diferentes etapas do processo. A Figura 2 representa as etapas de um fluxo de trabalho para a área de ciência de dados e como cada etapa é utilizada no projeto:

FIGURA 2 - FLUXO DE TRABALHO NA ÁREA DE DADOS



FONTE: CANAVER, MAYARA (2023)

- a) **Coleta de dados:** Realizada a coleta dos dados utilizando a api do Yahoo Finance;
- b) **Exploração dos dados:** Características dos dados analisados e observados quais dados possuem um peso maior para a tomada de decisão futura. Nesta etapa também se inicia a escolha do modelo preditor;
- c) **Preparação dos dados:** Dados transformados em um formato que o modelo possa compreender. Os dados são normalizados, removendo símbolos indesejados, espaços em excesso entre frases, dados vazios;
- d) **Criação, treinamento e validação do modelo:** Escolha do modelo ideal utilizado, com inserção dos dados já transformados e o resultado verificado em sequência;
- e) **Ajuste de hiperparâmetros:** Parâmetros do modelo ajustados para uma performance mais assertiva;
- f) **Implantação do modelo:** Implantado o modelo treinado em um ambiente de produção.

Esse fluxo de dados é aplicado ao projeto de conclusão de curso, o qual primeiramente mostra a incerteza dos investidores na aplicação da bolsa de valores.

A incerteza do resultado no momento de investir em uma determinada ação da bolsa influencia na decisão de se realizar o investimento ou não, essa incerteza surge devido à falta de conhecimento da área, cenários externos instáveis, como uma eleição ou uma catástrofe ambiental, ou também pelo perfil do investidor, como um perfil conservador.

O investidor é o principal cliente afetado na compra de títulos negociáveis no mercado, seja ele pessoa física ou jurídica, por conta de que há a confiança na empresa e caso a ação escolhida seja desvalorizada, ocorrerá o prejuízo financeiro.

Por ser uma área um pouco mais complexa e onde poucos brasileiros estão presentes, o assunto de investimento se torna um tabu entre a população. Isso faz com que as pessoas deixem de realizar um investimento por receio de não compreender o assunto, pois exige muito conhecimento na área e tempo disponível para acompanhar o pregão diário de negociações de ações entre compradores e vendedores, e também a perda financeira. Esse projeto viabiliza o ingresso e a utilização simplificada do mercado de ações, possibilitando que cada vez mais pessoas usufruam da bolsa de valores e, conseqüentemente, proporcione o aumento no volume de investimentos, fomentando a economia brasileira.

Alguns requisitos devem ser observados e preenchidos para que possa ser eficiente e útil este projeto:

### **3.1 Requisitos Funcionais**

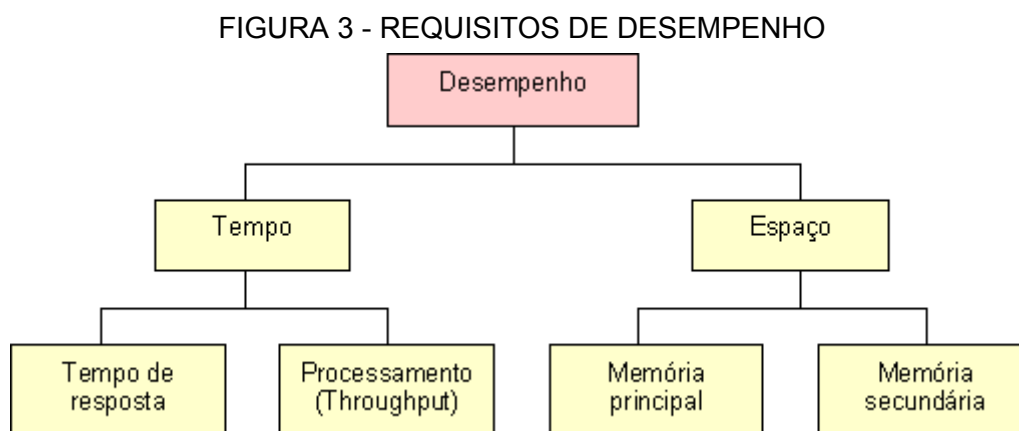
O modelo construído deverá ser capaz de, a partir das informações sobre as variações nos valores das ações de uma empresa a cada dia, gerar uma previsão da variação que tais ações irão sofrer no dia seguinte - auxiliando a tomada de decisão do usuário.

### **3.2 Requisitos Não-Funcionais**

A seguir serão demonstrados os requisitos não-funcionais de desempenho e de interoperabilidade.

### 3.2.1 Desempenho

A Figura 3 ilustra os requisitos de desempenhos que serão adotados:



FONTE: CANAVER, MAYARA (2023)

O modelo construído cumpre com requisitos máximos quanto ao tempo de inferência, tamanho do modelo e consumo de memória.

1. O tempo de inferência máxima aceitável será 1 minuto por entrada;
2. O tamanho máximo do modelo aceitável será 1 GB;
3. O consumo de memória máxima aceitável para carregar e realizar inferências com o modelo será de 6GB.

### 3.2.2 Interoperabilidade

1. O sistema deverá se comunicar com a *API* yfinance do Yahoo!.

Para que seja possível resolver e mitigar o problema apresentado, é utilizada a seguinte metodologia, assim como as tecnologias utilizadas no projeto e durante o seu desenvolvimento.

A metodologia empregada neste estudo é de natureza explicativa, na qual se aborda um tema específico, como a previsão do desempenho da ação da empresa AMBEV. Através da aplicação de práticas e experimentos, busca-se verificar a validade da hipótese formulada, desenvolvendo experimentos e comparando os resultados obtidos. Essa abordagem metodológica visa fornecer uma compreensão mais aprofundada do fenômeno em estudo, permitindo investigar e avaliar empiricamente a relação entre as variáveis envolvidas.

A implementação do modelo proposto no presente estudo fez uso de um conjunto de tecnologias amplamente utilizadas na área de Ciência de Dados. A seguir, serão apresentadas as principais tecnologias empregadas:

*Python* foi a linguagem de programação escolhida para desenvolver o modelo proposto. *Python* é uma linguagem de programação de alto nível, interpretada de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. A versão utilizada neste trabalho foi a versão 3.9. A escolha do *Python* deve-se à sua ampla adoção na comunidade de Ciência de Dados, bem como à sua rica coleção de bibliotecas e frameworks que facilitam a manipulação e análise de dados.

Para a edição do código-fonte, foi utilizado o *Visual Studio Code (VSCode)*. O *VSCode* é um editor de código-fonte desenvolvido pela Microsoft, amplamente adotado por sua interface intuitiva e recursos avançados. Ele oferece suporte para depuração, controle de versionamento *Git* incorporado, realce de sintaxe, complementação inteligente de código, snippets e refatoração de código. A escolha do *VSCode* foi motivada pela sua popularidade e eficiência no desenvolvimento de projetos de Ciência de Dados.

A plataforma de hospedagem de código-fonte e controle de versão *GitHub* foi utilizada para armazenar e compartilhar o código do projeto. O *GitHub* é amplamente reconhecido por sua facilidade de uso e integração com o *Git*, um sistema de controle de versão distribuído. Por meio do *GitHub*, os colaboradores podem contribuir para projetos privados e/ou *Open Source*, promovendo a colaboração e facilitando o compartilhamento do código entre membros da equipe.

A biblioteca *Seaborn* foi empregada para a visualização dos resultados obtidos. *Seaborn* é uma biblioteca de visualização *Python* baseada em *matplotlib*. Ela oferece uma interface de alto nível para desenhar gráficos estatísticos, facilitando a criação de visualizações informativas e esteticamente agradáveis. A escolha do *Seaborn* baseou-se em sua flexibilidade e recursos avançados de visualização, que permitem a representação clara e concisa dos dados analisados.

Por fim, a biblioteca *Pandas* desempenhou um papel fundamental na manipulação e análise dos dados. O *Pandas* é uma biblioteca de software criada para a linguagem *Python*, projetada especificamente para manipulação e análise eficiente de dados. Ela oferece estruturas e operações para manipular tabelas numéricas e séries temporais, tornando o processo de limpeza, transformação e análise dos dados mais eficiente e eficaz. O uso do *Pandas* neste estudo foi motivado pela sua popularidade na comunidade de Ciência de Dados e pela sua capacidade de lidar com dados de forma intuitiva e poderosa.

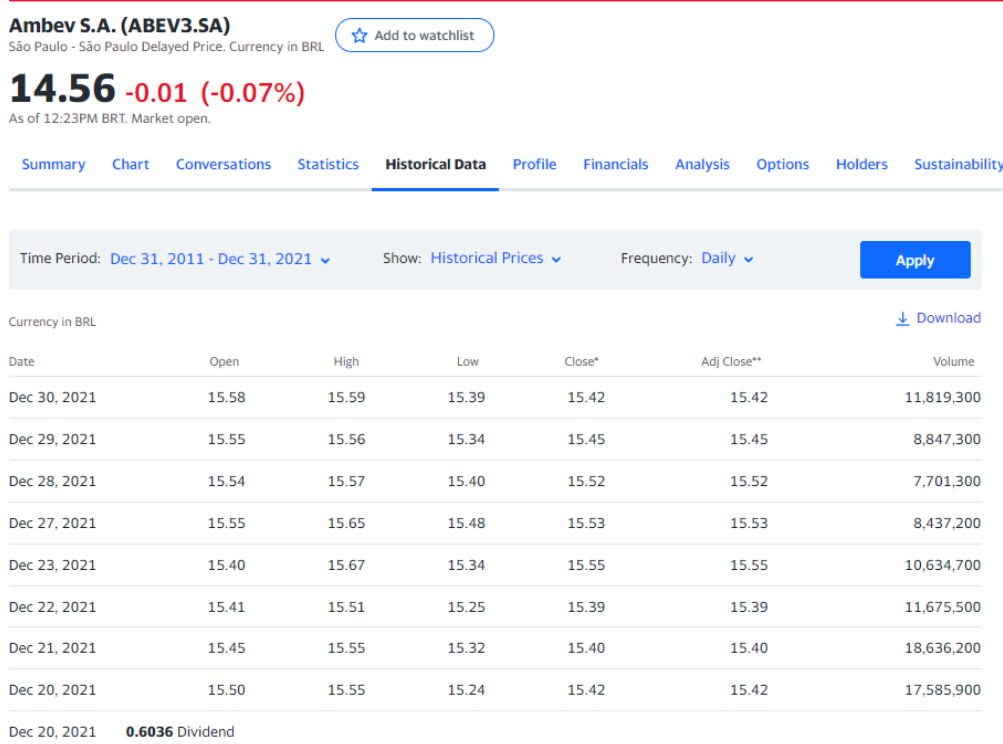
Em suma, a implementação do modelo proposto envolveu a utilização das tecnologias *Python*, *VSCode*, *GitHub*, *Seaborn* e *Pandas*. Essas tecnologias foram escolhidas com base em sua ampla adoção na área de Ciência de Dados, bem como em seus recursos e funcionalidades específicas que contribuíram para a eficiência e eficácia do desenvolvimento e análise dos dados.

### 3.3 Coleta de Dados

O Yahoo! Finance é um site que fornece notícias financeiras, dados e comentários, incluindo cotações de ações, relatórios financeiros e conteúdo original, além de ferramentas online para gerenciamento de finanças pessoais.

A partir da busca pela ação da AMBEV (ABEV3.SA), foi realizada uma primeira análise para reconhecimento dos dados em um determinado período de tempo (os 10 últimos dias de dezembro de 2021). O resultado da busca e a aparência dos dados podem ser vistos na Figura 4.

FIGURA 4 - DADOS DA AÇÃO ABEV3.SA



FONTE: YAHOO FINANCE (2023)

Foi utilizada a API open-source chamada yfinance, que permite realizar uma conexão direta com os dados apresentados acima por meio de código Python, e então armazenar esses dados em um formato de Data Frame (estrutura de dados bidimensional na qual os eixos são rotulados por linhas e colunas, semelhante à estrutura de uma planilha).

O período selecionado para o estudo foi entre o dia 01 de janeiro de 2023 até o dia 1 de maio de 2023, com uma granularidade semanal de informações. Ou seja, para cada semana foi obtido uma linha contendo as seguintes informações: data, preço de fechamento, preço de abertura, preço máximo alcançado no dia, preço mínimo alcançado no dia, volume de ações negociadas e mudanças percentuais do preço.

A Figura 5 apresenta as características de cada coluna e seu tipo (int, float, string, object etc...).

FIGURA 5 - TIPAGEM DAS COLUNAS DO CONJUNTO DE DADOS

Column	Dtype
Date	datetime64
Open	float64
High	float64
Low	float64
Close	float64
Volume	float64
Dividends	float64
Stock Splits	int64

FONTE: CANAVER, MAYARA (2023)

Não é necessário converter os tipos de dados, pois cada coluna possui a sua respectiva tipagem correta. Da mesma forma, não há necessidade de remover valores nulos, pois eles também são inexistentes no DataFrame.



## DESENVOLVIMENTO

Este capítulo é dedicado à realização de todos os processos de uma análise de séries temporais, incluindo coleta, limpeza e tratamento dos dados, modelagem e análise preditiva realizada, visualização, previsão dos dados e análise de resultados.

### 4.1 Processamento dos Dados

#### 4.1.1 Tratamento dos Dados

A variável alvo, ou seja, a que se deseja prever é o valor de retorno semanal da ação. O retorno semanal é uma medida que calcula a variação percentual do preço de uma ação entre dois pontos no tempo semanalmente. Ele fornece uma medida do desempenho da ação ao longo de cada semana.

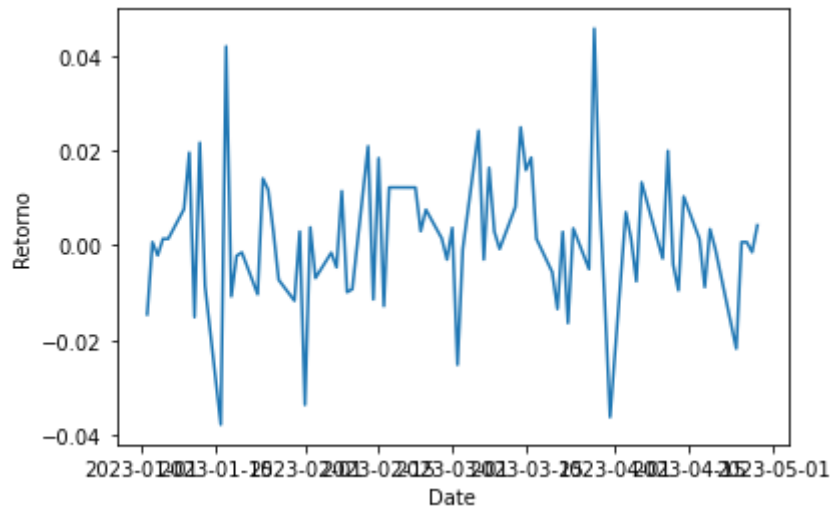
A equação para calcular o retorno de uma ação com base na variação do preço de fechamento e do preço de abertura é apresentada na equação (1) abaixo:

$$\text{Retorno} = (\text{Preço de Fechamento} - \text{Preço de Abertura}) / \text{Preço de Abertura} \quad (1)$$

Essa fórmula calcula a variação percentual entre o preço de abertura e o preço de fechamento da ação, ou seja, quanto o preço da ação mudou em relação ao preço inicial e será utilizada durante a previsão do retorno.

Ao realizar o cálculo do retorno, é observado na Figura 6 como essa variável reage conforme a variação do tempo analisado.

FIGURA 6 - VALORES DE RETORNO SEMANAIS DURANTE O PERÍODO ANALISADO



FONTE: CANAVER, MAYARA (2023)

Algumas características inicialmente observadas são as seguintes:

- Há necessidade de aplicar técnicas formais de séries temporais;
- Não é visualizada uma tendência aparente na figura, assim como não há presença de sazonalidade aparente.

Alguns testes formais são realizados e apresentam como base o *valor-p* (*p-value*), que representa a probabilidade de obter um resultado igual ou mais extremo do que o observado em uma amostra, assumindo-se que a hipótese nula seja verdadeira. Em outras palavras, o *valor-p* é a probabilidade de que a diferença entre os dados observados e os dados esperados sob a hipótese nula seja devida ao acaso.

O nível de significância é um valor pré-determinado utilizado em testes estatísticos para avaliar a força das evidências contra a hipótese nula. Ele representa a probabilidade máxima de erro que estamos dispostos a aceitar ao rejeitar a hipótese nula quando ela é realmente verdadeira. Nesse projeto o nível de significância é fixo em 5% (0,05), assim sendo, ao ter um *valor-p* abaixo desse nível, podemos rejeitar a hipótese nula proposta.

O nível de significância fixo em 5%, comumente representado por 0,05, é uma escolha convencional em muitas áreas de pesquisa e análise estatística. Esse

valor é amplamente utilizado como um limiar para tomada de decisão estatística. Ao definir um nível de significância de 5%, estamos estabelecendo um limite para a probabilidade máxima de erro que estamos dispostos a aceitar ao rejeitar a hipótese nula quando ela é realmente verdadeira. Em outras palavras, estamos dispostos a aceitar até 5% de chance de rejeitar incorretamente a hipótese nula, também conhecida como erro do tipo I.

#### 4.1.1.1 Etapas Estatísticas

Existem várias etapas utilizadas para determinar os valores dos parâmetros em nosso modelo, compreendendo a aplicação de métodos e testes estatísticos com propósitos distintos. A seguir, é apresentada a estrutura dos testes e os resultados esperados de cada um:

1. Verificação de Estacionariedade: Essa etapa visa identificar a presença de raiz unitária em uma série temporal, o que pode afetar a análise e interpretação dos dados.
2. Diferenciação: Caso a presença de raiz unitária seja identificada na etapa anterior, a diferenciação é empregada para eliminar esse efeito, tornando a série estacionária e facilitando as análises posteriores.
3. Teste de Normalidade: Nesta etapa, busca-se verificar se a série foi gerada por uma distribuição normal. Essa informação é relevante para a aplicação de métodos estatísticos que pressupõem a normalidade dos dados.
4. Transformação: Se o teste de normalidade indicar que a série não segue uma distribuição normal, é realizada uma transformação (Box-Cox) para estabilizá-la, permitindo uma melhor adequação aos pressupostos dos métodos estatísticos.
5. Testes de Autocorrelação: Esses testes avaliam a correlação entre as variáveis ao longo do tempo, fornecendo informações importantes para a determinação dos termos a serem incluídos no modelo de séries temporais.
6. Verificação de tendência e suavização da série: Essa etapa é fundamental para identificar tendências presentes na série temporal e realizar uma suavização adequada dos dados. Além disso, busca-se obter os parâmetros correspondentes necessários para o modelo de séries temporais.
7. Ao seguir essas etapas de forma metódica e criteriosa, é possível obter resultados mais confiáveis e robustos na modelagem de séries temporais, o que contribui para a precisão e validade das análises realizadas.

#### 4.1.1.2 Teste de *Dickey Fuller* (ADF)

Primeiramente é utilizado o teste de *Dickey Fuller*, que é utilizado para verificar a presença de raiz unitária em uma série temporal. A raiz unitária indica a

presença de tendência ou não estacionariedade na série, o que significa que os dados apresentam uma estrutura temporal que muda ao longo do tempo.

É definida a hipótese nula e então sua hipótese alternativa para o teste.

Hipótese Nula (H0): A série temporal possui uma raiz unitária e é não estacionária.

Hipótese Alternativa (H1): A série temporal não possui uma raiz unitária e é estacionária.

A Figura 7 apresenta o resultado do teste realizado:

FIGURA 7 - RESULTADO TESTE ADFULLER

ADF Estatístico	-10.577378
p-valor	0.000000

FONTE: CANAVER, MAYARA (2023)

A hipótese nula é rejeitada com base no valor-p resultante do teste realizado, sendo assim a série é estacionária e não há variação ao decorrer do tempo, assim como não há a presença de tendência.

#### 4.1.1.3 Diferenciação dos Dados

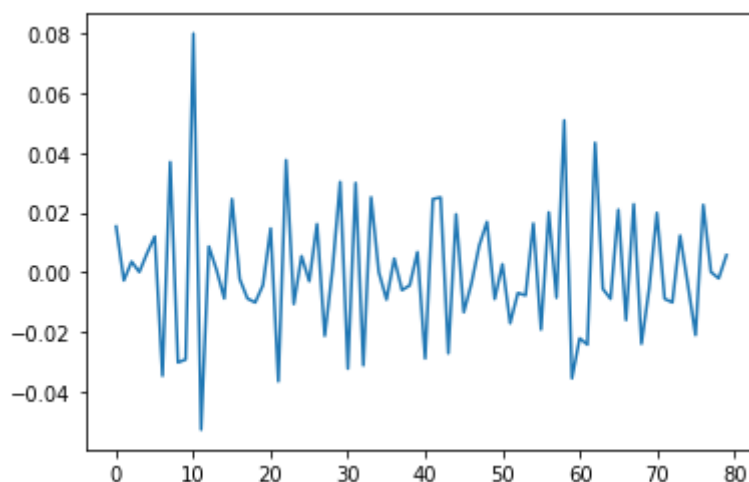
A diferenciação é uma técnica comumente aplicada em séries temporais para torná-las estacionárias quando há evidências de não estacionariedade, como a presença de tendência ou sazonalidade. A diferenciação envolve calcular a diferença entre os valores consecutivos da série ou a diferença em um determinado período de atraso.

Ao realizar a diferenciação, os valores originais da série são substituídos pelos valores das diferenças. O objetivo é remover ou reduzir a tendência e tornar a série estacionária, o que facilita a análise e modelagem.

Como a série temporal já é estacionária, não é necessário realizar a diferenciação. A diferenciação é aplicada quando há evidências de não estacionariedade na série, como a presença de tendência ou sazonalidade.

É aplicada a diferenciação para a confirmação da estacionariedade dessa série e conforme é visto na Figura 8 como o conjunto de dados permanece após a aplicação dessa diferenciação.

FIGURA 8 - CONJUNTO DE DADOS APÓS APLICAÇÃO DA DIFERENCIAÇÃO



FONTE: CANAVER, MAYARA (2023)

Assim é observado na Figura 9 novamente a validação com o teste ADF para confirmação da série estacionária.

FIGURA 9 - RESULTADO TESTE ADFULLER APÓS APLICAÇÃO DA DIFERENCIAÇÃO

ADF Estatístico	-11.553431
valor-p	0.000000

FONTE: CANAVER, MAYARA (2023)

#### 4.1.1.4 Teste de *Shapiro-Wilk*

O teste de normalidade é realizado por meio do teste de *Shapiro-Wilk*, que avalia uma amostra de dados e quantifica a probabilidade de esses dados terem sido extraídos de uma distribuição gaussiana (distribuição normal).

São estabelecidas a hipótese nula e a hipótese alternativa para o teste:

Hipótese Nula (H0): A amostra segue uma distribuição normal.

Hipótese Alternativa (H1): A amostra não segue uma distribuição normal.

O resultado preciso do teste estatístico é 0.9767, arredondado para quatro casas decimais, enquanto o valor-p é 0.1523, também arredondado para quatro casas decimais. Como o *valor-p* é maior que o nível de significância (0.05), não há evidências suficientes para rejeitar a hipótese nula de normalidade. Portanto, há indícios de que a amostra segue uma distribuição normal.

#### 4.1.1.5 Transformação Box-Cox

A transformação de Box-Cox é uma técnica amplamente empregada na análise de séries temporais com o intuito de estabilizar a variância e aproximar a distribuição dos dados à normalidade. É particularmente útil em contextos onde pressupõe-se uma distribuição normal dos dados ou quando a variância da série não se mantém constante ao longo do tempo.

Entretanto, é relevante ressaltar que a aplicação da transformação de Box-Cox em séries temporais nem sempre se mostra imprescindível. Quando a série em análise já exibe uma distribuição normal e uma variância estável ao longo do tempo, a referida transformação torna-se desnecessária, podendo inclusive distorcer as características originais dos dados.

No presente estudo, constatou-se que a série temporal em questão apresenta uma distribuição normal e uma variância relativamente constante, resultados confirmados mediante a realização de testes estatísticos apropriados. Dessa forma, deliberou-se pela não aplicação da transformação de Box-Cox aos dados, uma vez que tal procedimento não conferiria benefícios substanciais à análise. É fundamental adotar metodologias adequadas que preservem as peculiaridades estatísticas dos dados, a fim de propiciar uma análise coerente dos padrões temporais subjacentes à série.

Ao evitar a aplicação da transformação de Box-Cox, a presente investigação concentra-se diretamente na interpretação dos padrões temporais da série, proporcionando resultados mais robustos e passíveis de interpretação no contexto em que se insere.

#### 4.1.1.6 ACF e PACF

A autocorrelação e a autocorrelação parcial são duas medidas relacionadas usadas na análise de séries temporais para identificar a presença de dependência temporal nos dados.

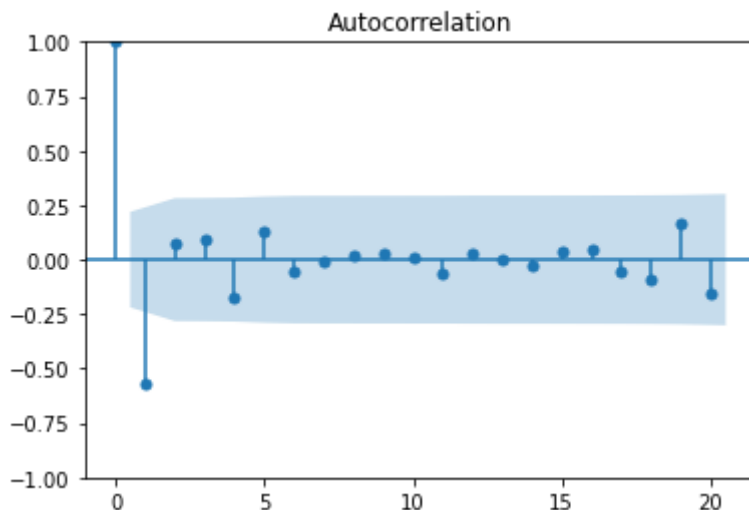
A autocorrelação (ACF - Autocorrelation Function) mede a correlação entre uma série temporal e suas versões defasadas (atrasadas) em diferentes intervalos de tempo. Em outras palavras, ela indica a relação entre os valores passados e futuros de uma série temporal. A ACF é usada para identificar a presença de padrões repetitivos e sazonalidade nos dados.

A autocorrelação parcial (PACF - Partial Autocorrelation Function) mede a correlação entre uma série temporal e sua versão defasada, controlando o efeito das defasagens intermediárias. Em outras palavras, ela mede a correlação direta entre dois pontos na série temporal, eliminando a influência das defasagens intermediárias. A PACF é usada para identificar a presença de dependência temporal direta e ajuda a determinar o número de termos de defasagem (lags) significativos em um modelo de séries temporais.

Ambas as medidas são úteis na análise de séries temporais e podem ajudar a identificar estruturas temporais relevantes nos dados. A ACF é amplamente utilizada para explorar a presença de sazonalidade e padrões repetitivos, enquanto a PACF é usada para determinar o número de termos de defasagem significativos em modelos de séries temporais, como o modelo AR (Autoregressive).

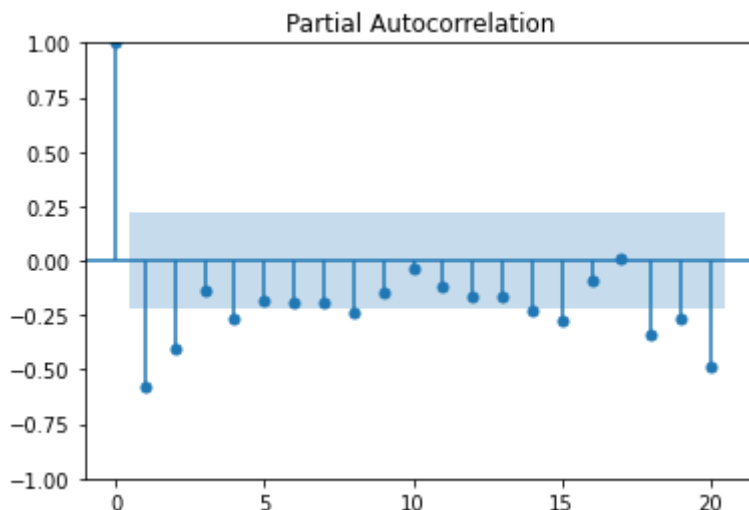
A Figura 10 é apresentada a autocorrelação realizada na série enquanto na Figura 11 é apresentada a autocorrelação parcial da série, e abaixo é explicado os resultados de cada figura.

FIGURA 10 - AUTOCORRELAÇÃO



FONTE: CANAVER, MAYARA (2023)

FIGURA 11 - AUTOCORRELAÇÃO PARCIAL



FONTE: CANAVER, MAYARA (2023)

É observado que na Figura 10, apenas o primeiro lag apresentou autocorrelação significativa, enquanto os demais lags estão dentro da área sombreada. Isso indica que há uma correlação forte entre os valores da série temporal apenas no primeiro lag, sugerindo a presença de uma dependência temporal imediata.

Na Figura 11, identificamos que os lags 1 e 2 mostram correlação parcial significativa, enquanto o lag 3 não apresenta relevância. Além disso, observamos



que o lag 4 é relevante, mas os lags subsequentes não são. No entanto, os três últimos lags mostram correlação parcial significativa.

Essas observações indicam que a série temporal pode ser influenciada por um efeito imediato (lag 1) e um efeito de curto prazo (lag 2) com relevância estatística. Os lags 3 e 4 também podem ter algum efeito, embora menos expressivo. Os demais lags não parecem ter influência significativa na série temporal.

Essas informações sugerem a possibilidade de usar um modelo AR (Autoregressivo) com termos de defasagem selecionados, como AR(1) ou AR(2), para modelar a estrutura temporal dos dados.

#### 4.1.1.7 Estudo de Médias Móveis

As médias móveis são uma ferramenta amplamente utilizada na análise de séries temporais, incluindo dados financeiros. Essa técnica permite suavizar as flutuações de curto prazo em um conjunto de dados, facilitando a identificação de tendências e padrões de longo prazo. A ideia central das médias móveis é calcular o valor médio de uma janela móvel de observações consecutivas ao longo do tempo.

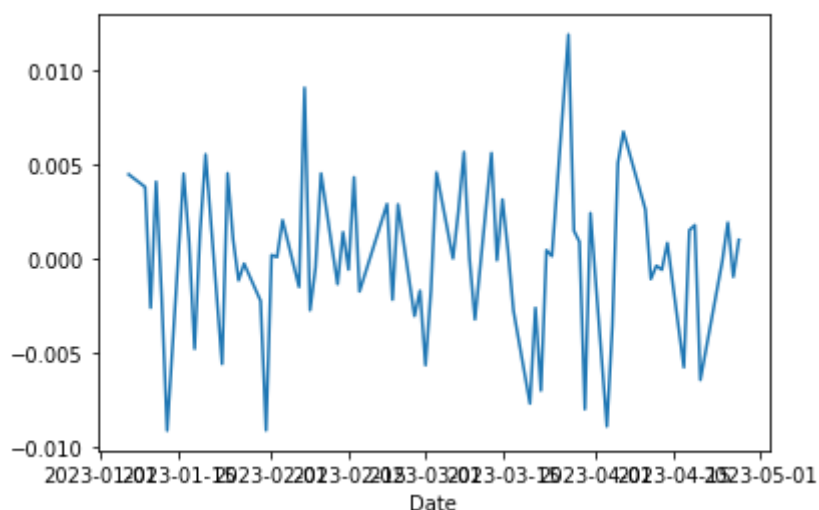
Neste estudo, optou-se por utilizar a média móvel com um tamanho de janela de 5 dias. Essa escolha foi baseada nas características específicas do conjunto de dados, que consiste nas cotações diárias da bolsa de valores. Dado o padrão de curto prazo e a variação diária observados nesse mercado, a média móvel com um window size de 5 dias foi considerada apropriada (pelo fato de que em uma semana há cinco dias úteis).

A janela de 5 dias permite suavizar as flutuações diárias e destacar as tendências de curto prazo de forma mais sensível. Ao calcular a média das últimas 5 observações, é possível reduzir o ruído do mercado e visualizar com mais clareza os movimentos de preço ao longo do tempo. Essa abordagem é particularmente útil para identificar padrões de curto prazo e ajustar estratégias de investimento de acordo com as variações diárias.

Vale ressaltar que a escolha do tamanho da janela é uma decisão empírica e depende das características do conjunto de dados e dos objetivos da análise. O window size de cinco dias foi adotado neste estudo após uma avaliação visual e considerando a frequência dos dados diários de cotações da bolsa de valores.

A Figura 12 mostra o estado da série ao aplicar a janela móvel com seu valor atribuído de cinco dias.

FIGURA 12 - JANELA MÓVEL



FONTE: CANAVER, MAYARA (2023)

## 4.2 Modelagem dos Dados

Após a realização de procedimentos de processamento e análise dos dados, incluindo a detecção e remoção de tendências, bem como a determinação de parâmetros relevantes, foram conduzidas etapas de modelagem e previsão dos dados. Para essa finalidade, foram empregados os modelos ARIMA e GARCH, com o intuito de obter e avaliar os resultados mais adequados. Com base nessa análise comparativa, foi possível identificar e retornar os resultados de cada modelo empregado no contexto deste estudo.

### 4.2.1 ARIMA

Portanto, quando se deseja realizar a análise de uma série temporal e é necessário diferenciá-la  $d$  vezes para torná-la estacionária, essa série é denominada ARIMA (Autoregressive Integrated Moving Average). O modelo ARIMA é uma abordagem estatística amplamente empregada na análise de séries temporais, sendo composto por três parâmetros:  $p$ ,  $d$  e  $q$ .

O parâmetro " $p$ " refere-se ao número de termos autorregressivos no modelo. Isso implica que o modelo utiliza os valores anteriores da série temporal para prever os valores futuros. Quanto maior o valor de  $p$ , mais termos autorregressivos são incorporados ao modelo, tornando-o mais complexo em termos de consideração do comportamento passado da série.

O parâmetro " $d$ " indica o número de diferenciações necessárias para tornar a série temporal estacionária. A estacionariedade é uma propriedade desejável em uma série temporal, pois indica que a média e a variância da série permanecem constantes ao longo do tempo. Caso a série não seja estacionária, ela pode ser diferenciada uma ou mais vezes até que se torne estacionária. Quanto maior o valor de  $d$ , maior a quantidade de diferenciações necessárias, resultando em um modelo mais complexo.

O parâmetro " $q$ " está relacionado ao número de termos da média móvel presentes no modelo. Isso significa que o modelo leva em consideração os erros passados das previsões para melhorar as previsões futuras. Um valor mais elevado de  $q$  implica a incorporação de um maior número de termos da média móvel, tornando o modelo mais complexo em termos de consideração dos erros passados.

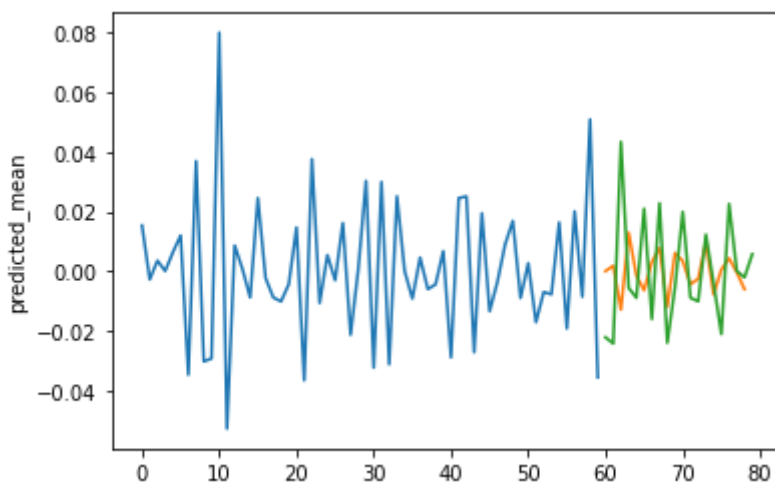
O motivo para utilizar o modelo ARIMA na análise de séries temporais reside em sua flexibilidade e capacidade de capturar diversas complexidades presentes nessas séries, tais como tendências, sazonalidade e correlação entre os erros. Além disso, o ARIMA é amplamente utilizado e bem compreendido, o que facilita sua aplicação e interpretação em contextos diversos de análise de séries temporais.

Com as verificações realizadas durante a decomposição da série é possível confirmar que os valores respectivos de  $p$ ,  $d$  e  $q$  são 4, 1, 5, onde existem quatro

termos auto-regressivos, apenas uma diferenciação foi realizada, e são necessários cinco termos para média móvel.

Na figura 13 é possível visualizar a predição resultante gerada pelo ARIMA, assim como na figura 14 é possível visualizar a mesma predição, porém apenas na região específica do resultado gerado.

FIGURA 13 - PREDIÇÃO GERAL DA SÉRIE

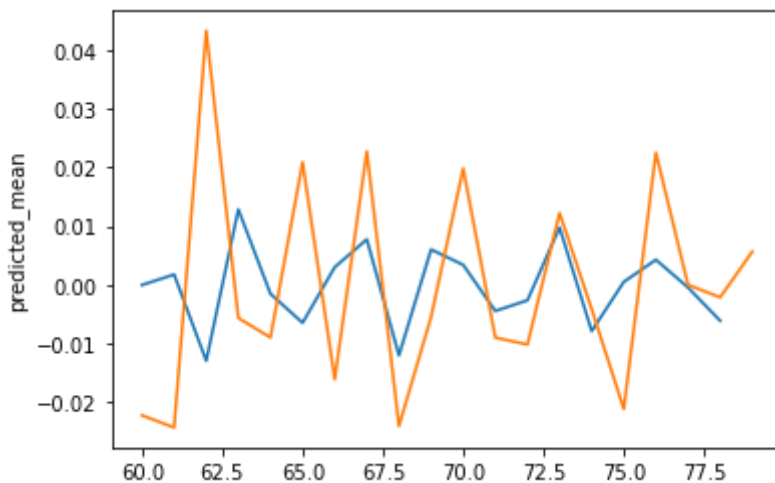


FONTE: CANAVER, MAYARA (2023)

A representação das cores do figura é feita da seguinte forma:

- Azul: plotagem dos dados de treino;
- Verde: plotagem dos dados de teste;
- Laranja: plotagem da previsão.

FIGURA 14 - PREDIÇÃO ESPECÍFICA DA SÉRIE



FONTE: CANAVER, MAYARA (2023)

#### 4.1.2.5 AIC

Em um modelo ARIMA, AIC significa Akaike Information Criterion. É um critério usado para a seleção do modelo que também equilibra a qualidade do ajuste do modelo com a complexidade do modelo. O AIC é baseado na verossimilhança logarítmica do modelo e penaliza modelos com mais parâmetros.

O AIC é calculado da seguinte forma como mostrado na equação (2):

$$\text{AIC} = -2 * \log\text{-verossimilhança} + 2 * \text{número de parâmetros} \quad (2)$$

Um valor AIC mais baixo indica um melhor ajuste do modelo, considerando tanto a qualidade do ajuste quanto a complexidade do modelo. Ao comparar vários modelos ARIMA, o modelo com o menor valor de AIC poderá ser escolhido como o modelo preferencial.

Neste caso, um valor AIC de -300.6834 (resultante do modelo) indica que o modelo ARIMA que está sendo avaliado alcançou um valor AIC relativamente baixo, sugerindo um bom equilíbrio entre ajuste e complexidade do modelo. Isso indica que esse modelo ARIMA específico é um forte candidato para prever os dados.

#### 4.2.2 GARCH

O modelo GARCH (Generalized Autoregressive Conditional Heteroscedasticity) é um modelo estatístico frequentemente utilizado na análise de séries temporais financeiras, especialmente para modelar a volatilidade de séries de retornos de ativos financeiros. O modelo é composto por dois parâmetros principais: p e q.

O parâmetro "p" refere-se ao número de termos autorregressivos do modelo que são utilizados para modelar a volatilidade. Esses termos autorregressivos permitem que o modelo leve em consideração a volatilidade passada ao prever a volatilidade futura.

O parâmetro "q" refere-se ao número de termos da média móvel do modelo que são utilizados para modelar a volatilidade.

Assim como no ARIMA, os parâmetros de volatilidade do modelo GARCH são respectivamente 4 e 5 por conta do número de termos autorregressivos (4) e seu valor de janelas móveis (5).

Ao contrário do modelo ARIMA, o modelo GARCH é capaz de capturar a heteroscedasticidade condicional, que é a propriedade de que a variância do erro não é constante ao longo do tempo. Em outras palavras, o modelo GARCH permite que a volatilidade de uma série temporal seja modelada de forma mais precisa e realista, levando em conta a variação na volatilidade ao longo do tempo.

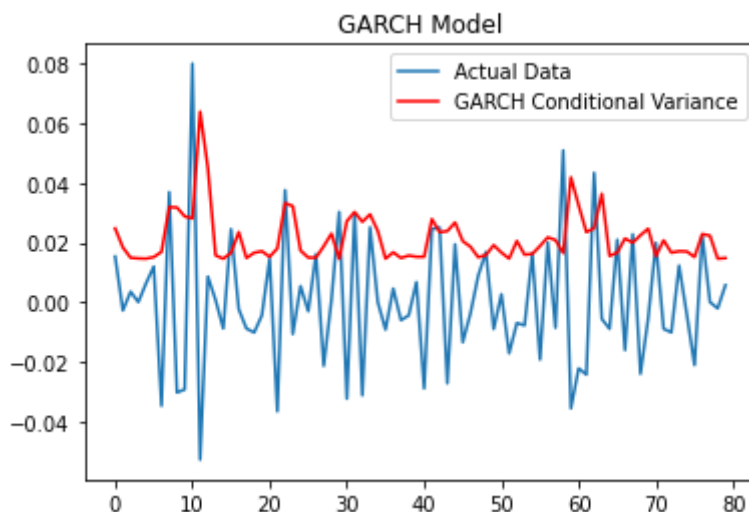
Os modelos GARCH são amplamente utilizados na econometria financeira e têm se mostrado eficazes na captura da natureza dinâmica da volatilidade dos mercados financeiros. Ao incorporar termos autorregressivos e de média móvel, os modelos GARCH podem capturar a persistência e o agrupamento de choques de volatilidade, fornecendo informações valiosas sobre o risco e a incerteza associados aos ativos financeiros.

A estimação e inferência dos modelos GARCH geralmente envolvem métodos de estimação de máxima verossimilhança, em que os parâmetros do modelo são estimados com base na função de verossimilhança, que mede o ajuste do modelo aos dados observados. Essas estimativas podem ser utilizadas para fazer previsões de volatilidade futura, avaliar níveis de risco e informar estratégias de investimento e gestão de risco.

Em resumo, o modelo GARCH é uma ferramenta poderosa para modelar a volatilidade de séries temporais financeiras, permitindo uma representação mais precisa da natureza mutável da volatilidade ao longo do tempo. Sua aplicação na análise financeira e na gestão de risco pode fornecer insights valiosos para investidores e profissionais financeiros.

A Figura 15 apresenta o resultado do modelo GARCH aplicado na série analisada e abaixo é possível verificar algumas informações observadas no conjunto de dados.

FIGURA 15 - VARIÂNCIA DO MODELO GARCH



FONTE: CANAVER, MAYARA (2023)

A incorporação do modelo GARCH em nossa análise ARIMA nos permitiu capturar a dinâmica de volatilidade presente nos dados. A série de volatilidade estimada apresentou padrões distintos ao longo do período investigado.

Inicialmente, no início da série temporal, observamos um período de alta volatilidade. Isso indica um nível significativo de flutuação e incerteza nos pontos de dados iniciais. Isso pode ser atribuído a vários fatores, como mudanças repentinas nas condições de mercado, choques externos ou outros eventos imprevistos.

À medida que avançamos na série temporal, notamos uma diminuição na volatilidade, com a série apresentando níveis relativamente baixos de flutuação. Este período de menor volatilidade indica um ambiente mais estável e previsível. No entanto, dentro dessa tendência geral de menor volatilidade, houve casos esporádicos de pequenos picos ou aumentos temporários na volatilidade. Esses "pequenos picos" sugerem surtos de incerteza de curta duração ou choques localizados que interromperam momentaneamente o período relativamente estável.

No final da série histórica, observamos um ligeiro aumento na volatilidade, embora ainda menor do que no período inicial de alta volatilidade. Essa “pequena grande alta” no final da série sugere um retorno a uma fase um pouco mais incerta ou volátil. No entanto, é importante observar que esse aumento da volatilidade não foi tão pronunciado quanto na fase inicial, indicando um período comparativamente mais estável no geral.

Essas descobertas destacam a natureza dinâmica dos dados, com a volatilidade exibindo variações em diferentes períodos de tempo. A presença de períodos de alta e baixa volatilidade, juntamente com picos intermitentes, ressalta a importância de considerar a volatilidade ao analisar e prever os dados.

### **4.3 Métricas para Verificação do Modelo**

É necessário realizar a medição do erro do modelo para saber se o mesmo está assertivo, as algumas métricas utilizadas após a previsão de uma série temporal são as seguintes:

#### **4.3.1 Mean Absolute Percentage Error (MAPE)**

O MAPE mede a diferença percentual média entre os valores previstos e reais. Um valor MAPE menor sugere que as previsões do modelo ARIMA estão mais próximas dos valores reais, indicando melhor precisão e desempenho.

Nesse caso, um MAPE de 6,9918% indica que, em média, as previsões do modelo desviam dos valores reais em aproximadamente 6,9918% em termos de diferença percentual. Isso significa que, em média, as previsões do modelo estão em torno de 6,9918% distantes dos valores verdadeiros.

#### **4.3.2 Mean Absolute Error (MAE)**

Em um modelo ARIMA, se o valor do erro médio absoluto (MAE) for 0,0155, isso indica que, em média, a diferença absoluta entre os valores previstos e os valores reais é 0,0155.



Um valor baixo de MAE sugere que as previsões do modelo ARIMA estão próximas dos valores reais. Um MAE menor indica melhor precisão e desempenho do modelo. Nesse caso, um MAE de 0,0155 indica que, em média, as previsões do modelo desviam dos valores reais em aproximadamente 0,0155 unidades.

#### 4.3.3 Mean Squared Error (MSE)

MSE é uma métrica comumente usada para medir a magnitude média dos erros nas previsões de um modelo. Ele amplifica erros maiores em comparação com o MAE (Erro Médio Absoluto), pois envolve o quadrado das diferenças.

Nesse caso, um valor MSE de 0,0003 sugere que, em média, as diferenças quadradas entre os valores previstos e os valores reais chegam a aproximadamente 0,0003. Ele fornece uma medida do tamanho típico dos erros quadrados entre as previsões do modelo e os valores reais.

#### 4.3.4 Root Mean Squared Error (RMSE)

RMSE é a raiz quadrada do erro quadrático médio (MSE) e geralmente é usado para medir a precisão das previsões de um modelo. Ele fornece uma medida do erro médio na mesma unidade que a variável de destino.

Nesse caso, um valor de RMSE de 0,0199 sugere que, em média, as previsões do modelo desviam dos valores reais em aproximadamente 0,0199 unidades. Indica o tamanho típico dos erros entre os valores previstos e os valores reais.

### 4.4 Resultados

A presente pesquisa apresentou uma análise de série temporal sobre a ação da AMBEV, utilizando um modelo ARIMA, e avaliou seu desempenho por meio de diversas métricas de avaliação. Os resultados das métricas obtidas revelaram um bom ajuste do modelo aos dados e uma precisão satisfatória nas previsões.

As métricas utilizadas, incluindo o Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) e Mean Absolute

Percentage Error (MAPE), demonstraram valores baixos, indicando que as previsões do modelo ARIMA se aproximaram significativamente dos valores reais das ações da AMBEV.

## CONCLUSÃO

As métricas baixas obtidas sugerem que o modelo ARIMA foi capaz de capturar os padrões e tendências presentes nos dados históricos da ação da AMBEV. Portanto, é possível afirmar que o modelo demonstrou uma capacidade adequada de previsão para a série temporal em questão.

Ao incorporar o modelo GARCH em nossa análise ARIMA, fomos capazes de explicar esses padrões de volatilidade variáveis e capturar a natureza variável no tempo da incerteza dos dados. Isso melhorou nossa compreensão da dinâmica dos dados e aumentou a precisão de nossas previsões.

Esses resultados têm implicações significativas para investidores, analistas financeiros e tomadores de decisão no mercado de capitais, fornecendo informações valiosas para a avaliação e previsão do comportamento futuro das ações da AMBEV. Além disso, esse estudo contribui para o avanço da literatura acadêmica relacionada à análise de séries temporais e modelagem de dados financeiros, ao demonstrar a eficácia do modelo ARIMA no contexto específico das ações da AMBEV.

No entanto, é importante destacar que a análise de séries temporais é um campo complexo e dinâmico, sujeito a múltiplos fatores que podem influenciar os resultados das previsões. Portanto, é recomendado que futuros estudos considerem a incorporação de um modelo multivariado.

Em suma, os resultados obtidos nesta pesquisa respaldam a utilização do modelo ARIMA na previsão das ações da AMBEV, proporcionando insights valiosos para investidores e auxiliando na tomada de decisões financeiras informadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

How to Difference a Time Series Dataset with Python. Machinelearningmastery. Disponível em: <<https://machinelearningmastery.com/difference-time-series-dataset-python/>>. Acesso em: 10 fev. 2016.

How to Perform a Shapiro-Wilk Test in Python. Statology. Disponível em: <<https://www.statology.org/shapiro-wilk-test-python/>>. Acesso em: 10 fev. 2016.

Princípios básicos para criar previsões de Séries Temporais. Ensina AI. Disponível em: <<https://medium.com/ensina-ai/princ%C3%ADpios-b%C3%A1sicos-para-criar-previs%C3%B5es-de-s%C3%A9ries-temporais-e58c451a25b>>. Acesso em: 11 fev. 2016.

Entenda o significado dos códigos das ações. XP Investimentos. Disponível em: <<https://conteudos.xpi.com.br/aprenda-a-investir/relatorios/codigos-das-acoes/>>. Acesso em: 15 fev. 2016.

Augmented Dickey-Fuller Test in Python. Statology. Disponível em: <<https://www.statology.org/dickey-fuller-test-python/>>. Acesso em: 16 fev. 2016.

Finding Seasonal Trends in Time-Series Data with Python. Towards Data Science. Disponível em: <<https://towardsdatascience.com/finding-seasonal-trends-in-time-series-data-with-python-ce10c37aa861>>. Acesso em: 19 fev. 2016.

How to Use and Remove Trend Information from Time Series Data in Python. Machinelearningmastery. Disponível em: <<https://machinelearningmastery.com/time-series-trends-in-python/>>. Acesso em: 23 fev. 2016.

Média móvel para NumPy Array em Python. Delft Stack. Disponível em: <<https://www.delftstack.com/pt/howto/python/moving-average-python/#:~:text=A%20m%C3%A9dia%20m%C3%B3vel%20%C3%A9%20freq%C3%BCentemente,tend%C3%AAncias%20dos%20pre%C3%A7os%20das%20a%C3%A7%C3%B5es.>>. Acesso em: 27 fev. 2016.

Mathematical statistics functions. Python. Disponível em: <<https://docs.python.org/3/library/statistics.html>>. Acesso em: 03 mar. 2016.

The Jupyter Notebook. Ipython. Disponível em: <<https://ipython.org/notebook.html>>. Acesso em: 05 mar. 2016.

MEDIDAS DE POSIÇÃO E DISPERSÃO. Projeto Livro Aberto de Matemática. Disponível em: <[https://impa.br/wp-content/uploads/2019/07/PAPMEM\\_julho\\_2019\\_Nocoas-de-Estatistica\\_2\\_parte\\_Flavia.pdf](https://impa.br/wp-content/uploads/2019/07/PAPMEM_julho_2019_Nocoas-de-Estatistica_2_parte_Flavia.pdf)>. Acesso em: 18 mar. 2016.

Conceito: Requisitos. UFPE. Disponível em:

<[https://www.cin.ufpe.br/~gta/rup-vc/core.base\\_rup/guidances/concepts/requirements\\_62E28784.html](https://www.cin.ufpe.br/~gta/rup-vc/core.base_rup/guidances/concepts/requirements_62E28784.html)>. Acesso em: 21 mar. 2016.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2). Springer Science & Business Media.

Artigo Engenharia de Software 3 - Requisitos Não Funcionais. DEVMEDIA. Disponível em: <<https://www.devmedia.com.br/artigo-engenharia-de-software-3-requisitos-nao-funcionais/9525#:~:text=O%20termo%20requisitos%20n%C3%A3o%20funcionais,arquitetural%20e%20forma%20de%20implementa%C3%A7%C3%A3o>>. Acesso em: 22 mar. 2016.

Apostila de Análise de Séries Temporais. Prof. Manoel Ivanildo Silvestre Bezerra. Disponível em:

<[https://www.academia.edu/35118481/Apostila\\_de\\_An%C3%A1lise\\_de\\_S%C3%A9ries\\_Temporais](https://www.academia.edu/35118481/Apostila_de_An%C3%A1lise_de_S%C3%A9ries_Temporais)>. Acesso em: 22 mar. 2016.

ANÁLISE DE SÉRIES TEMPORAIS. Laboratório de Estatística e Geoinformação.

Disponível em: <<http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>>. Acesso em: 28 mar. 2016.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

ANÁLISE DE SÉRIES TEMPORAIS. UFSC. Disponível em:

<<https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>>. Acesso em: 28 mar. 2016.

Dividendos: o que são? Dá para viver disso?. Nubank. Disponível em:

<<https://blog.nubank.com.br/dividendos/>>. Acesso em: 29 mar. 2016.

O que realmente significa o valor-p?. JBP. Disponível em:

<<https://www.scielo.br/j/jbpneu/a/SWk5XsCsXTW7GBZq8n7mVMJ/?format=pdf&lang=pt>>. Acesso em: 29 mar. 2016.

Lutkepohl, H. (2005). New introduction to multiple time series analysis. Springer Science & Business Media.