



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
CURSO DE SISTEMAS DE INFORMAÇÃO

Fernando Silva Poerner

Luiz Felipe Ribeiro Baroncello

**Sistema de análise comparativa fundamentalista utilizando inteligência
artificial**

FLORIANÓPOLIS - SC

2023

Fernando Silva Poerner

Luiz Felipe Ribeiro Baroncello

Sistema de análise comparativa fundamentalista utilizando inteligência artificial

Trabalho de Conclusão de Curso submetido ao curso de Sistemas de Informação do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador(a): Prof. Dr. Elder Rizzon Santos

FLORIANÓPOLIS - SC

2023

Sistema de análise comparativa fundamentalista utilizando inteligência artificial

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel e aprovado em sua forma final pelo Curso de Sistemas de Informação

Florianópolis, [dia] de [mês] de [ano].

Insira neste espaço
a assinatura

Coordenação do Curso

Banca examinadora

Insira neste espaço
a assinatura

Prof. Dr. Elder Rizzon Santos
Orientador

Insira neste espaço
a assinatura

Prof.(a) [nome do(a) professor(a)], Dr.(a)
Instituição [nome da instituição]

Insira neste espaço
a assinatura

Prof.(a) [nome do(a) professor(a)], Dr.(a)
Instituição [nome da instituição]

FLORIANÓPOLIS - SC - 2023

RESUMO

Este TCC tem o intuito de criar uma ferramenta para análise de dados de empresas considerando informações importantes como risco, liquidez, resultado operacional e inexistência de recuperação judicial. Os dados utilizados devem ser coletados em uma ou mais fontes de dados existentes do universo de ações do IBOVESPA. Fazendo tratamento em cima dos dados coletados e a utilização de séries temporais na análise do preço justo utilizando os parâmetros fundamentalistas comentados anteriormente.

Palavras-chave: Ações, IBOVESPA, Inteligência artificial, Machine Learning, Análise Fundamentalista, Preço Justo, Séries Temporais.

ABSTRACT

This TCC aims to create a tool for analyzing company data, considering important information such as risk, liquidity, operational results, and the absence of judicial recovery. The data used should be collected from one or more existing data sources within the IBOVESPA stock universe. Processing the collected data and employing time series in the fair price analysis using the fundamental parameters mentioned earlier.

Keywords: Stocks, IBOVESPA, Artificial Intelligence, Machine Learning, Fundamental Analysis, Fair Price, Time Series.

LISTA DE FIGURAS

Figura 1 – Ciclo de Data Science	19
Figura 2 – Exemplo de Redes Neurais Recorrentes	20
Figura 3 – Regressão polinomial	23
Figura 4 – Regressão Linear Simples	24
Figura 5 – Metodologia de previsão baseada em análise técnica	29
Figura 6 – Fluxo de dados para seleção de ações baseada em análise fundamentalista	34
Figura 7 – Criação dos DataFrames de teste e validação	40
Figura 8 – Criação da instância do Prophet, ajuste do modelo e previsão dos ROEs futuros através do modelo treinado	41
Figura 9 – Dados de saída da função predict	42
Figura 10 – Retorno da função rmse	43
Figura 11 – Gráfico de previsão para o modelo Prophet	44
Figura 12 – Gráfico de tendência e Sazonalidade	45
Figura 13 – Retorno da função rmse	47
Figura 14 – Gráfico da média móvel do modelo ARIMA	48
Figura 15 – Dados retornados pelo ADF	48
Figura 16 – Resultado informações modelo ARIMA	49
Figura 17 – Treinamento do modelo com ARIMA	49
Figura 18 – Gráfico com dados ajustados	50
Figura 19 – Valores preditos ARIMA	50
Figura 20 – Gráfico gerado ARIMA para previsão e validação do modelo	51
Figura 21 – Comparação de valores de previsão e valores de validação ARIMA	52

LISTA DE TABELAS

Tabela 1 – Comparação dos tipos, modelos de IA e métricas de avaliação utilizados dos trabalhos relacionados.	33
Tabela 2 – Principais dados do balanço patrimonial da Vale.	38
Tabela 3 – Principais dados da demonstração de resultados da Vale.	38
Tabela 4 – Cálculo do ROE para o período de 31/12/2022.	39
Tabela 5 – ROE calculado para todos os trimestres da empresa Vale.	40
Tabela 6 – Tabela dos dados preditos pelo modelo do Prophet.	47
Tabela 7 – Tabela com os dados para o cálculo do RMSE.	47
Tabela 8 – Tabela dos dados preditos pelo modelo do ARIMA.	53
Tabela 9 – Tabela com os dados para o cálculo do RMSE.	54
Tabela 10 – Valores relacionados ao cálculo do preço justo segundo modelo proposto.	57

SUMÁRIO

1 INTRODUÇÃO	11
2 OBJETIVOS	14
2.1 OBJETIVO GERAL	14
2.2 OBJETIVOS ESPECÍFICOS	14
3 FUNDAMENTAÇÃO TEÓRICA	15
3.1 INTRODUÇÃO AO MERCADO DE AÇÕES	15
3.2 ANÁLISE FUNDAMENTALISTA	15
Análise fundamentalista quantitativa	16
Análise fundamentalista qualitativa	16
Retorno sobre o Patrimônio Líquido (ROE)	16
Margem Líquida	17
Resultado Bruto	17
Margem Bruta	17
Dividend Yield	17
Preço Justo	17
3.3 MODELOS DE PREVISÃO	18
3.4 DATA SCIENCE	19
3.5 REDES NEURAIS	20
4 TRABALHOS RELACIONADOS	22
4.1 MACHINE LEARNING FOR QUANTITATIVE FINANCE APPLICATIONS: A SURVEY	22
4.2 SURVEY OF STOCK MARKET PREDICTION USING MACHINE LEARNING APPROACH	23
4.3 MACHINE LEARNING TECHNIQUES AND USE OF EVENT INFORMATION FOR STOCK MARKET PREDICTION: A SURVEY AND EVALUATION	26
4.4 STOCK MARKET FORECASTING TECHNIQUES: A SURVEY	28
4.5 USO DE REDES NEURAIS RECORRENTES PARA PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS	29
4.6 OUTROS TRABALHOS	30
4.7 COMPARAÇÃO/ANÁLISE DOS TRABALHOS	32
5 DESENVOLVIMENTO	35
5.1 FERRAMENTAS	36
5.2 COLETA DE DADOS DO FUNDAMENTUS	37
5.3 PRÉ-PROCESSAMENTO DOS DADOS	38
5.4 SÉRIES TEMPORAIS	40
5.4.1 PROPHET	40
5.4.2 ARIMA	48
5.4.3 DISCUSSÃO	54
5.5 CÁLCULO DO PREÇO JUSTO	55
5.6 ANÁLISE RESULTADOS	56
6 CONSIDERAÇÕES FINAIS	58
REFERÊNCIAS	60

7 ANEXOS
8 ARTIGO TCC

63
69

1 INTRODUÇÃO

No ano de 2020, a bolsa de valores brasileira atraiu 1,5 milhão de novos investidores, registrando um aumento significativo de 92% segundo Zonta d'Ávila, Mariana. Bolsa conquista 1,5 milhão de novos investidores em 2020, um aumento de 92% no ano. InfoMoney, 2021. Disponível em: <https://www.infomoney.com.br/onde-investir/bolsa-conquista-15-milhao-de-novos-investidores-em-2020-um-aumento-de-92-no-ano/>. Acesso em: 07 nov. 2023. Esse aumento tem relação com diversos fatores, como a queda da taxa de juros, a desvalorização brusca do mercado por conta da pandemia, entre outros. É de grande importância a criação de ferramentas que consigam levar análises mais técnicas e consolidadas para o público geral de maneira simplificada e acessível. Com a evolução tecnológica da Inteligência Artificial (IA) é possível criar estruturas para análise de padrões e comportamentos dos ativos ao longo dos anos considerando seus valores fundamentalistas e a variação dos mesmos. Através destas análises é possível criar um indicador único que facilite a visualização.

De acordo com Goldschmidt (2010) o aprendizado de máquina (AM) é uma área da inteligência artificial voltada para o desenvolvimento de algoritmos e técnicas computacionais que possibilitam que os computadores consigam aprender. Ainda segundo Goldschmidt (2010) o aprendizado só é possível através dos sistemas de aprendizado. Softwares conseguem tomar decisões baseados em um banco de dados com dados reais que já ocorreram, ou seja, são tomadas decisões por meio de padrões. Segundo Pereira (2013), padrões são conjuntos de características que determinam um objeto, sendo que essas características são determinadas buscando a separação das classes a fim simplificar as tarefas de um classificador. Existem duas categorias da AM sendo elas a: supervisionada e não supervisionada.

Entre os algoritmos de aprendizado podemos citar as Redes Neurais Artificiais (RNAs). Uma RNA procura simular o funcionamento do cérebro humano através de um conjunto de neurônios artificiais conectados entre si, sendo que um peso é atribuído a cada uma dessas conexões (BARRETO, 2002). Durante a fase de treinamento, estes pesos são ajustados a partir de exemplos informados como entrada para a RNA. Após o término da fase de treinamento, os pesos ajustados são

utilizados para a classificação de amostras desconhecidas pela rede (BARRETO, 2002).

No passado, as empresas poderiam contratar equipes de estatísticos, modeladores e analistas para explorar manualmente os conjuntos de dados, mas seu volume e variedade superam muito a capacidade da análise manual. Ao mesmo tempo, computadores se tornaram muito mais poderosos, a comunicação em rede é onipresente, desenvolvidos algoritmos que podem conectar conjunto de dados para permitir análises muito mais amplas e profundas do que antes. A convergência desses fenômenos deu origem à aplicação, cada vez mais difundida, de princípios de Data Science e de técnica de mineração de dados nos negócios. (PROVOST; FAWCETT, 2016).

Com a abundância de dados relacionados ao mercado financeiro, consumir e organizar tais informações se torna um trabalho complexo e desafiador. Segundo Emílio Ferrara, (2014), diversas abordagens para extração de dados da Web foram desenvolvidas para a resolução de problemas específicos. Por conta disso, é importante focarmos em um alto nível de automação para reduzir a necessidade de esforços humanos.

A arte de prever o mercado de ação é complexa devido a diversos fatores. Segundo Patel et al., (2015), uma das dificuldades são as incertezas que o mercado está envolvido, por esse motivo sugere que os investidores possuem duas formas principais de analisar uma ação. A primeira delas a análise fundamentalista, que permite ao investidor enxergar os valores intrínsecos das ações e seus indicadores de saúde econômica e política. A outra é a análise técnica através de estatísticas que envolvem a flutuação do mercado e seus gráficos, muitas vezes deixando de lado os valores que tornam uma empresa segura no longo prazo, dessa forma se tornam de extrema importância a utilização de Ferramentas para análise de dados através de AM e Data Science. Com estas ferramentas podemos utilizar dados históricos em períodos pré-definidos nos permitindo aplicar conceitos da análise fundamentalista e criar um modelo preditivo para análise de longo prazo.

Uma vez que os dados tenham sido coletados e analisados, busca-se desenvolver um modelo de predição que ajude a identificar flutuações nos valores de ações das empresas selecionadas a longo prazo, de forma que o modelo auxilie na realização de análises sobre as mesmas.

Para validar os modelos gerados, será definido um ou mais testes com premissas básicas para indicar se um modelo é útil ou não, para otimizar e selecionar o melhor modelo para o presente trabalho.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Esta monografia tem como objetivo geral implementar uma ferramenta para análise de dados históricos da bolsa de valores brasileira, considerando técnicas de análise fundamentalista, utilizando boas práticas de investimento a longo prazo, como o preço justo e buscando valorização e retorno sobre investimento em longos períodos

2.2 OBJETIVOS ESPECÍFICOS

A seguir, os objetivos específicos deste trabalho são definidos:

- Analisar o estado da arte em extração e transformação de dados, previsão do mercado de ações relativo à volatilidade, aprendizado de máquina e data science para embasamento teórico no assunto;
- A partir das análises realizadas, desenvolver um modelo de previsão para identificar flutuações nos valores de ações de empresas selecionadas;
- Utilizar dados históricos dos valores de análise fundamentalista da bolsa de valores para treinar e validar o modelo preditivo proposto;
- Definir um experimento para avaliar a capacidade de análise fundamentalista do modelo preditivo por meio de um protótipo;
- Comparar a funcionalidade do modelo criado com a valorização real do mercado;

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo pretende apresentar os conceitos fundamentais para a realização do trabalho sobre mercado financeiro e aprendizado de máquina, utilizados no modelo proposto.

3.1 INTRODUÇÃO AO MERCADO DE AÇÕES

Com o passar dos anos cada vez mais investidores brasileiros buscam aprender sobre finanças e investimentos, em busca da tão sonhada independência financeira, por esse motivo buscam entender melhor os conceitos por trás do mercado e como podem utilizar tais conhecimentos para alavancar seus investimentos. O mercado de ações é o local onde podem ser negociados, comprados ou vendidos, bens como valores mobiliários, mercadorias e câmbio. Os mercados também são considerados um indicador da saúde da economia de um país.

No Brasil temos a B3 que possui o índice IBOVESPA, responsável pela gestão e administração das negociações de títulos, valores mobiliários e contratos derivativos, realizando serviços de registro, compensação e liquidação das operações.

Segundo Nelson (2017) ,dentre os ativos que podem ser negociados, destacam-se as ações, títulos de empresas e correspondem à menor parcela de seu valor. As ações representam o direito à propriedade e lucros da empresa.

Os títulos são emitidos quando uma empresa abre seu capital para o público, o que permite ser negociado na bolsa de valores. O valor dos títulos é determinado pela relação de oferta e demanda, e pode ser considerado um reflexo das condições estruturais, de mercado e financeiras da empresa.

3.2 ANÁLISE FUNDAMENTALISTA

É uma das ferramentas utilizadas para se tomar decisões no mercado financeiro que auxilia na identificação de oportunidades de investimento. Nesse tipo de análise, estudamos indicadores econômicos da empresa em que tem interesse, seus demonstrativos financeiros e o cenário macroeconômico no qual ela está

inserida. Dessa forma é possível identificar o momento adequado à compra ou à venda da ação, avaliando se o preço do papel está justo ou não. A análise também fornece certo nível de segurança ao investidor, pois se baseia em indicadores que muitas vezes são utilizados para validação da saúde financeira de uma empresa.

No longo prazo, o alto retorno obtido com as escolhas certas compensa a atenção e o cuidado no acompanhamento e estudo dos indicadores fundamentalistas das empresas.

Existem tipos diferentes de análise fundamentalista, alguns deles são:

Análise fundamentalista quantitativa

A análise fundamentalista quantitativa é feita a partir dos números das empresas. Nela, são avaliados indicadores como a receita, a margem operacional e o endividamento da empresa, por exemplo.

Análise fundamentalista qualitativa

A análise qualitativa usa fundamentos que apontam para a qualidade da gestão, como o nível de governança, o potencial de crescimento e os riscos do segmento, por exemplo.

Alguns indicadores utilizados na análise fundamentalista retirados do site InvestSite. Glossário. Disponível em: <https://www.investsite.com.br/glossario.php>. Acesso em: 07 de novembro de 2023.

Retorno sobre o Patrimônio Líquido (ROE)

Lucro Líquido dividido pelo Patrimônio Líquido. Também conhecido como ROE ou *Return Over Equity*. Fornece uma visão do retorno para o acionista, já incluindo os efeitos de todos os custos e despesas da operação, inclusive alavancagem financeira e impostos. Por essa mesma razão é de difícil comparação

entre empresas com diferentes níveis de alavancagem financeira e/ou empresas atuando em indústrias com cargas diferentes de impostos.

Margem Líquida

Lucro Líquido dividido pela Receita Líquida. Também conhecido como *Net Margin*.

Resultado Bruto

Resultado Bruto é a Receita Líquida menos o custo dos bens vendidos e/ou dos serviços prestados.

Margem Bruta

Resultado Bruto dividido pela Receita Líquida. Também conhecido como *Gross Margin*.

Dividend Yield

Dividendos em dinheiro pagos por ação (incluindo juros sobre o capital próprio) divididos pelo preço da ação. É uma medida que pode ser facilmente comparável com algum investimento com características de "renda fixa", obviamente assumindo preço de ação e *dividend yields* constantes.

Conceitos retirados do glossário <https://www.investsite.com.br/glossario.php>

Preço Justo

Segundo Décio Bazin (1994) o preço justo de uma ação está relacionado à capacidade da empresa de fornecer dividendos consistentes ao longo do tempo. Isso envolve escolher ações de empresas sólidas e lucrativas que distribuem uma parte significativa de seus lucros aos acionistas. O cálculo do preço justo considera

a estabilidade e consistência dos dividendos, além dos indicadores tradicionais. Vale notar que diferentes abordagens podem ser usadas na avaliação de ações, e o conceito de preço justo pode variar entre investidores.

3.3 MODELOS DE PREVISÃO

Existem diversos modelos de previsão que podemos utilizar para a análise do mercado financeiro. No trabalho feito por Paul D. Yoo (2005) o autor traz quatro modelos utilizados para prever oscilações do mercado financeiro em diversas obras. Os quatro modelos citados são a Previsão de série temporal tradicional (Traditional time series prediction), Redes Neurais (Neural Network), máquina de vetores de suporte (Support vector machine), raciocínio baseado em casos (Case-based reasoning).

Os modelos de previsão de série temporal tradicional são amplamente utilizados em economia para previsão de séries temporais. Esses modelos conseguem modelar relações lineares entre fatores que influenciam o mercado e o valor do mercado. Em economia, existem dois tipos básicos de tempo previsão de séries: univariada (regressão simples) e multivariada (regressão multivariada) Maddala, G. S. (1992).

A máquina de vetores de suporte foi baseada na teoria da aprendizagem estatística. Se tornou um tema de estudo intensivo devido ao seu sucesso na aplicação em tarefas de classificação e regressão, especialmente na previsão de séries temporais e aplicativos relacionados.

O raciocínio baseado em casos é uma técnica de raciocínio que reutiliza casos passados para encontrar uma solução para o novo problema. Raciocínio baseado em casos não só captura o conhecimento da organização, mas também fornece explicações para as soluções derivadas. Por esta razão, raciocínio baseado em casos é normalmente utilizado em conjunto com outras técnicas como Algoritmo genético para previsibilidade de eventos e dados como no trabalho apresentado por Kim, K. (2004).

O modelo ARIMA (AutoRegressive Integrated Moving Average) destaca-se como uma abordagem amplamente adotada na previsão de séries temporais financeiras. Ao fundamentar-se na diferenciação e combinação de componentes

autorregressivos e médias móveis, o ARIMA demonstra eficácia na identificação de padrões temporais complexos presentes nos dados do mercado financeiro.

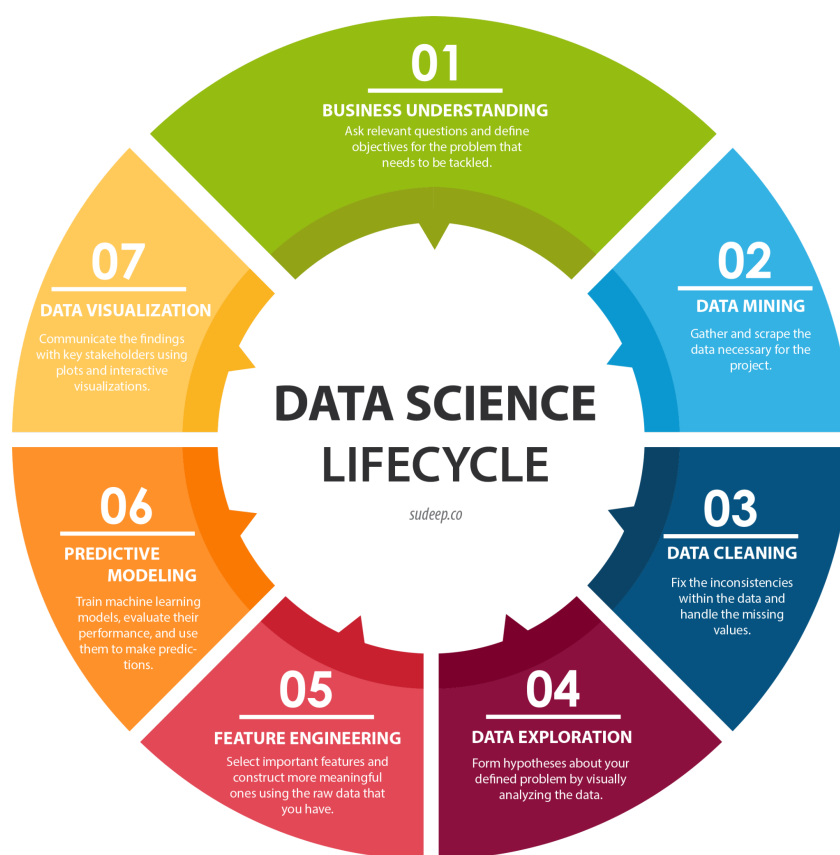
Outro modelo de grande relevância na previsão do mercado financeiro é o Prophet, desenvolvido pelo Facebook. Segundo a documentação em <https://facebook.github.io/prophet/>, acessado em 07 de Novembro de 2023. Projetado para enfrentar desafios associados a séries temporais que exibem sazonalidades anuais, efeitos de feriados e outros padrões temporais, o Prophet destaca-se pela sua utilidade em situações que envolvem dados ausentes ou irregularidades nos registros temporais. Ao oferecer flexibilidade e robustez, esse modelo se destaca na previsão de movimentos do mercado financeiro. Sua capacidade de capturar padrões sazonais e tendências de forma intuitiva o posiciona como uma ferramenta inestimável na análise preditiva para profissionais financeiros.

3.4 DATA SCIENCE

Data science ou ciência de dados, é uma forma de estatística que envolve elementos de matemática e ciências da computação para extrair percepções de dados quantitativos e qualitativos. As ferramentas e tecnologias utilizadas em data science incluem algoritmos e estruturas de machine learning, assim como linguagens de programação e bibliotecas de visualização. Um cientista de dados combina programação, matemática e conhecimentos de domínio para fazer estudos e chegar a importantes conclusões através dos dados. A Ciência de dados é conhecida por ser um processo cíclico tendo diversas etapas neste processo como, conhecimento do tópico sendo uma compreensão básica do problema e uma base de conhecimento sólida e normalmente especializada, a próxima etapa então é a aquisição de dados espalhados por diversos lugares e podem estar em formatos diferentes, em seguida vem o preparo dos dados, a qual é a etapa que leva mais tempo e a considerada mais importante e a qualidade deste processo depende diretamente da qualidade dos provenientes do processo anterior, vem então a etapa de a exploração de dados envolve a identificação e compreensão de padrões em um conjunto de informações, o próximo passo é fazer a modelagem e avaliação preditiva, onde é feito a validação dos modelos preditivos que os dados serão utilizados, após este processo é feita a interpretação e implantação dos dados

coletados, para identificar as possibilidades e utilidades do estudo feito e implantá-lo de forma efetiva para isso é necessária a etapa de monitoração para identificar se o modelo está se comportando corretamente como esperado ou se ajustes precisam ser feitos, alterações nos dados impactam diretamente o monitoramento visto que deve ser acompanhado para garantir o funcionamento do modelo e por fim vem a repetição das etapas, para garantir que a interpretação seja cada vez mais eficaz e garanta seu funcionamento no longo prazo.

Figura 1 – Ciclo de Data Science



Fonte:

[https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.ht](https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html)

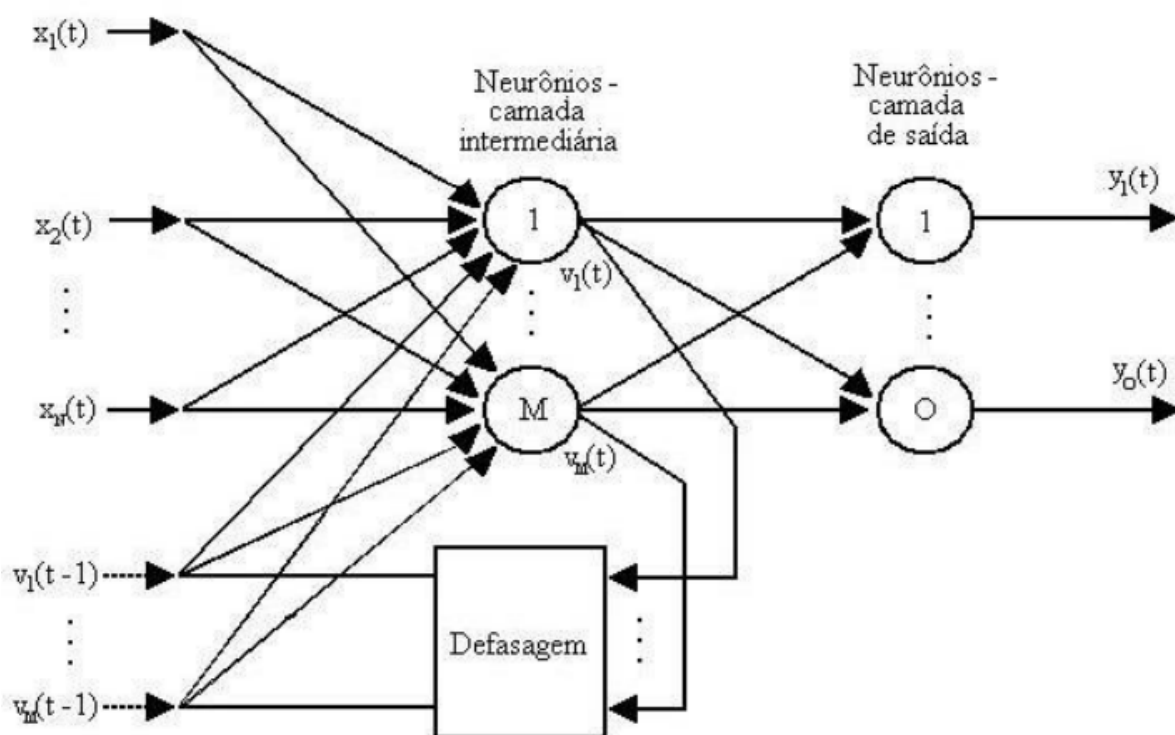
[ml](#) Acessado em 07 de novembro 2023

3.5 REDES NEURAIAS

Redes Neurais são geralmente divididas em dois tipos, sendo eles Redes Neurais Artificiais e Redes Neurais Recorrentes. Redes Neurais Artificiais são sistemas de aprendizado de máquina que simulam a estrutura de funcionamento dos neurônios no cérebro humano (Luger, 2013).

Já Redes Neurais Recorrentes constituem-se da adição de uma rede progressiva, sem retroalimentação, com algumas modificações, notadamente a introdução de realimentação, ampliando sua potencialidade de modelagem de dados temporais ou espaciais (Gomes, 2005). Neste caso, saídas de neurônios são usadas como entradas para outros neurônios, podendo esses serem de outra camada ou até mesmo eles mesmos, possibilitando a criação de diversas arquiteturas. Abaixo é possível visualizar um exemplo de Redes Neurais Recorrentes.

Figura 2 – Exemplo de Redes Neurais Recorrentes



Fonte: Gomes, Daniel T. (2005, p. 12)

Ambas buscam imitar como um conhecimento seria formado através da busca de padrões e tomada de decisões. Uma rede neural costuma receber vários exemplos como parte de sua fase de treinamento para poder tomar decisões baseadas neles. São muito usadas em aplicações de inteligência artificial e aprendizado de máquina.

4 TRABALHOS RELACIONADOS

Neste capítulo serão apresentados, discutidos e analisados trabalhos em que estejam relacionados ao tema deste. Estes trabalhos abordam diferentes técnicas relacionadas à inteligência artificial para realizar previsões no mercado financeiro.

4.1 MACHINE LEARNING FOR QUANTITATIVE FINANCE APPLICATIONS: A SURVEY

No survey de Rundo et al. (2019) os autores buscam, além de elencar artigos relacionados ao uso de técnicas de aprendizado de máquina no mercado financeiro, trazer uma comparação entre essas técnicas e técnicas tradicionais, com o intuito de entender qual seria a opção mais adequada para o cenário atual.

Desta forma, os autores diferenciam as duas abordagens comparadas, analisando inicialmente as literaturas disponíveis a respeito do de modelos autorregressivos, com essas análises eles chegaram a conclusão que os modelos poderiam ser divididos em dois principais métodos, sendo eles o estatístico e o baseado em aprendizado de máquina. Para Rundo et al (2019), através destas divisões definidas foi possível encontrar vários trabalhos onde modelos desse gênero eram usados para performar previsões no mercado financeiro.

Através destas análises, pode-se observar que esta técnica era usada principalmente para realizar previsões conhecidas como “um passo à frente”, pois quando eram usadas para fazer previsões de mais longo prazo, tendiam a falhar bastante. E apesar de apresentar um bom desempenho em relação a sua acurácia, o modelo apresenta uma grande falha quando lida com assimetrias, com tempos irregulares e também com alta volatilidade (Rundo et al, 2019).

Para a análise de modelos baseado em aprendizagem de máquina, foram analisadas três técnicas diferentes de modelos gerados por aprendizagem de máquina, sendo elas a de suporte de máquina vetor, aprendizagem profunda e redes neurais.

Em relação à comparação, vários trabalhos foram levantados comparando os dois modelos entre si, e vários trabalhos que comparam cada modelo individualmente, chegando a conclusão de que ambas as abordagens chegam a

resultados positivos em sua previsão, com o problema em comum de não serem ótimas em identificar volatilidades muito grandes.

Rundo et al. (2019) realizou uma pesquisa aprofundada dos modelos de Aprendizado de máquina usados no setor financeiro, concentrando-se principalmente na previsão de ações e alocação de portfólio. Os autores observaram que modelos tradicionais como o modelo autorregressivo integrado de médias móveis (e suas variações) e o modelo de suavização exponencial tendem a ter um desempenho ruim em comparação com os modelos de aprendizagem profunda para previsão de preço de ações e a volatilidade dos papéis devido a vários aspectos destes dados como: complexidade, alta dimensionalidade e dinamicidade causal. Eles também observaram que os modelos de Aprendizagem profunda, como a memória de curto prazo, tendem a ter um desempenho melhor do que a máquina de vetores de suporte e os modelos Perceptron multicamadas em diversos casos relacionados a tendências de ações e previsão de preços.

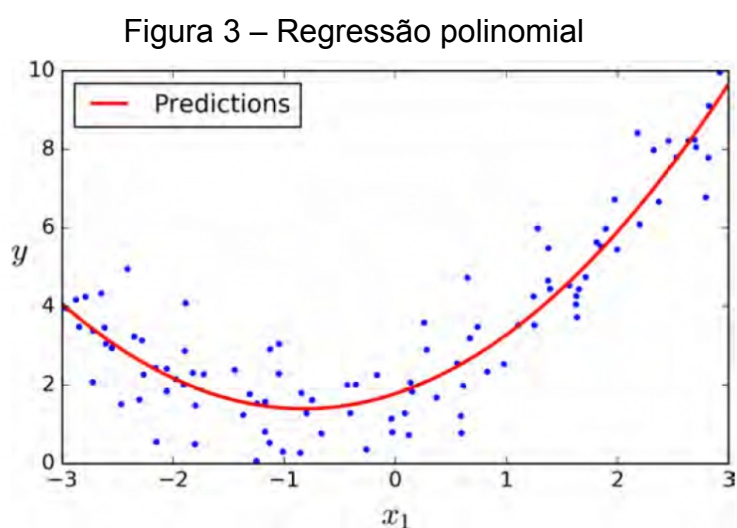
Entre os trabalhos avaliados, os métodos mais usados foram: Modelo Autorregressivo Integrado de Médias Móveis, Perceptron Multicamadas, Máquina de Vetores de Suporte e a Memória de Curto Prazo Longa. O método menos utilizado foi o Aprendizado por Reforço Profundo. Rundo et al. (2019) realizou uma exploração aprofundada de vários trabalhos que usam Aprendizado de Máquina, Aprendizagem Profunda e Aprendizado por Reforço Profundo no mercado financeiro. Os autores ainda comentam brevemente sobre análise sentimentalista e sobre como a mesma poderia ser usada em conjunto com as abordagens analisadas acima.

4.2 SURVEY OF STOCK MARKET PREDICTION USING MACHINE LEARNING APPROACH

Na pesquisa feita por Ashish Sharma e Dinesh Bhuriya (2017) o mercado de ações oscila por natureza e a pesquisa sobre o mercado de ações é uma das mais importantes questões nos últimos anos. As pessoas investem em ações com base em alguma previsão. Para prever, os preços de mercado, as pessoas pesquisam diferentes métodos e ferramentas para minimizar seus riscos e aumentar seus

lucros. A utilização de KDD (KNOWLEDGE DISCOVERY IN DATABASES) e Data Mining sendo utilizado como parte da técnica de Descoberta de conhecimento em banco de dados torna o processo de predição de preços mais concreto e traz mais confiança para os investidores que estão iniciando no mercado de ações. Ashish Sharma e Dinesh Bhuriya também mencionam em seus trabalhos correlatos algumas formas de alcançar a predição das ações como processo de desenvolvimento do modelo “fuzzy”. Através do estudo destes trabalhos, Ashis e Dinesh trazem o conceito da Análise de Regressão. Em estatística, a análise de regressão é um processo para estimar as relações entre variáveis. Isto inclui muitas técnicas para modelagem e análise de diferentes variáveis, quando o foco está na relação entre uma variável dependente e uma ou mais variáveis independentes. A análise de regressão ajuda a entender como o valor típico de uma variável dependente muda quando uma das variáveis independentes é alterada, enquanto as outras se mantêm no mesmo estado. Esta análise também é usada para entender qual das variáveis independentes estão relacionadas com a variável dependente, e explorar as relações das mesmas para entender melhor o modelo preditivo.

Trazem também o conceito de regressão polinomial, este procedimento ajusta até 10 modelos diferentes de regressão para duas variáveis, sendo uma dependente e uma independente. A variável independente é expandida num polinômio até o décimo grau com geração de novas variáveis.

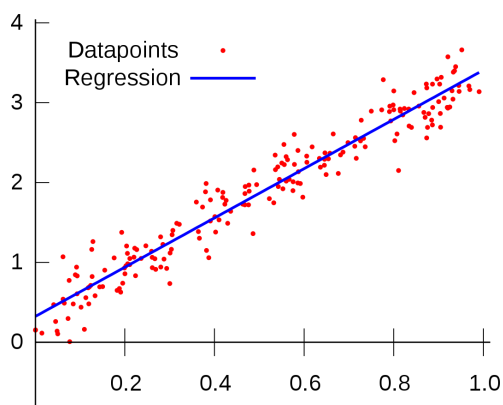


Além destas, Ashis e Dinesh trazem o conceito de Regressão RBF (Radial Basis Function), a função de base radial é caracterizada por apresentar uma resposta que decresce (ou cresce) monotonicamente com a distância em relação a um ponto central.

Outro conceito estudado é a Regressão Sigmoide, Uma função matemática que tem uma forma de "S" (curva sigmoide). E estuda dois valores possíveis entre 0 e 1.

Por último, o conceito de regressão linear é apresentado, uma abordagem para a relação entre um dependente escalar variável Y e uma ou mais explicações (ou variáveis independentes) denotado por X. No caso de uma única variável explicativa é chamada regressão linear simples. Para mais de uma variável explicativa, o processo é chamado de regressão linear múltipla. Este termo distingue de multivariado, onde variáveis correlacionadas e dependentes são previstas, em vez de uma única variável escalar. Na regressão linear, os dados são modelados usando funções de previsão, e os parâmetros do modelo são desconhecidos. Como todas as formas de análise de regressão, a regressão linear se concentra na distribuição de probabilidade de Y dado X, em vez da distribuição de probabilidade entre de Y e X.

Figura 4 – Regressão Linear Simples



Fonte: https://pt.wikipedia.org/wiki/Regress%C3%A3o_linear

Em resumo, os autores usaram modelagem de regressão para fazer previsões do mercado financeiro. O modelo de regressão utilizado foi redefinido para cada problema e seu grau específico de complexidade. Para os experimentos como especificado foram anteriormente realizadas as regressões Polinomiais, RBF, Sigmoide e Linear. Para fins de previsão, a regressão Linear se provou ser a mais adequada utilizando o método dos mínimos quadrados.

Concluindo seu estudo, Ashish Sharma e Dinesh Bhuriya buscam utilizar a previsão do desempenho das ações como um papel no importante mercado de ações que pode ser um processo desafiador devido à dinâmica natureza do mercado de ações.

4.3 MACHINE LEARNING TECHNIQUES AND USE OF EVENT INFORMATION FOR STOCK MARKET PREDICTION: A SURVEY AND EVALUATION

Neste artigo, Paul D. Yoo, Maria H. Kim e Tony Jan apresentam o desenvolvimento de modelos de previsão do mercado de ações, e levantam discussões sobre suas vantagens e desvantagens. Estudam a relação entre eventos globais e a previsão dos mercados de ações. A partir disso procuram incorporar informações destes eventos com modelos de previsão mais precisos. Assim buscando melhor desempenho na previsão de séries temporais financeiras.

Já em sua introdução este artigo levanta a importância de algumas tecnologias utilizadas na previsão do mercado de ações como Regressão Linear, Redes Neurais, Algoritmos Genéticos, Máquina de vetores de suporte e Raciocínio baseado em casos.

Entre estas a mais utilizada sendo as Redes Neurais por sua capacidade de performar em diversas abordagens de previsão do mercado de ativos. Com este conceito trazem um dos primeiros trabalhos feito por White, H (1988) com a utilização de Redes Neurais Direta para identificar algumas irregularidades do mercado não detectadas. Este movimento deu início a várias pesquisas em busca de formas mais precisas de prever o mercado. Um destes trabalhos foi feito por Phua et al. (2000) que utilizou Redes Neurais com Algoritmos genéticos para prever com a precisão de 81% o mercado de ações de Singapura.

Paul, Maria e Tony trazem em tópicos separados os conceitos de previsão de séries temporais utilizando mesclando os conceitos como a predição de séries

temporais de forma tradicional utilizando técnicas como Box-Jenkins que em diversas literaturas se mostra inferior ao compararmos com técnicas de aprendizado de máquina como redes neurais.

Os autores trazem diversas referências sobre a capacidade que as redes neurais oferecem para prever as oscilações do mercado com mais precisão que outras técnicas existentes por ser capaz de descobrir relações não lineares entre os pares de entrada e saída utilizados para treinar a rede tornando a ferramenta ideal para modelar sistemas dinâmicos como o mercado de ações, sendo algumas de suas vantagens são a capacidade de aprender relações entre dados por si só ao invés de tratá-los de forma funcional, são conhecidas por ser um aproximador universal onde podemos relacionar conjuntos de dados diferentes com graus diferentes de precisão quando possuímos dados suficientes para tal.

Fornecer também certa tolerância a dados inconclusivos e incompletos por essas características se adaptam bem a problemas complexos não lineares. Uma de suas desvantagens é sua caixa preta, pois não revela a importância de cada variável e como elas têm impacto em relação às demais variáveis independentes, por conta disso se impossibilita entender como a rede determina os preços futuros das ações. Outro problema é a sobrecarga de treinos, que pode ser evitada com testes e treinamentos com métodos de validação cruzada. Outra dificuldade é o aprendizado de padrões quando utilizamos conjuntos de dados com alta dimensionalidade.

Mais uma técnica de aprendizado de máquina estudada foi a máquina de vetores de suporte, um tipo muito específico de algoritmos de aprendizagem caracterizado pela capacidade de controle das funções de decisão, e a falta de ordem da solução. Se mostra mais resistente ao problema de sobrecarga de treinos, podendo alcançar mais desempenho ao fazer generalizações dos dados, Paul D. Yoo (2005). Nas comparações feitas por Kim, K. (2003) a máquina de vetores de suporte mostrou melhores resultados ao comparar os mesmos períodos e dados utilizando redes neurais e raciocínio baseado em casos, pois consegue generalizar melhor ao compararmos com as técnicas mais tradicionais.

Sobre Raciocínio baseado em casos, embora as Redes neurais ofereçam uma capacidade de aprendizado relativamente boa em comparação com outras técnicas, eles nem sempre podem explicar por que eles chegam a uma solução em particular. Além disso, elas nem sempre podem garantir uma solução

completamente correta, chegando à mesma solução repetidamente dada a mesma quantidade de dados para treinamento, e nem podem garantir que é a melhor solução. Ao contrário das Redes neurais, os sistemas especializados tendem a fornecer explicações para as soluções. Tais sistemas capturam principalmente o conhecimento de especialistas. As organizações têm conhecimento coletivo e conhecimentos que construíram ao longo dos anos. Raciocínio baseado em casos é uma técnica de raciocínio que reutiliza casos passados para encontrar uma solução para o novo problema. Por esta razão, raciocínio baseado em casos é popularmente aplicado a muitas aplicações, Paul D. Yoo (2005).

A capacidade das Redes neurais de aprender relacionamentos não lineares dos pares de entrada/saída de treinamento permite-lhes modelar sistemas dinâmicos não lineares como mercados de ações mais precisamente (Lawrence, R. 1997, 'Using Neural Networks to Forecast Stock Market Prices', University of Manitoba). Outros modelos, como Máquina de vetores de suporte e Raciocínio baseado em casos, também tornou-se popular na previsão do mercado de ações. A Máquina de vetores de suporte mostrou sua aplicação bem-sucedida em tarefas de classificação e tarefas de regressão, especialmente em séries temporais, predição e aplicações financeiras relacionadas (Yang, H. Chan, L. and King, I. 2002, 'Support Vector Machine Regression for Volatile Stock Market Prediction', Lecture Notes in Computer Science).

Além disso, a web é considerada a principal fonte de eventos para o mercado de ações, previsões contendo os eventos mais recentes e latentes em formação. Assim, para a previsão do mercado de ações, mineração na web é necessário para atingir maior precisão nas previsões em um curto espaço de tempo.

4.4 STOCK MARKET FORECASTING TECHNIQUES: A SURVEY

Neste survey realizado por G. PREETHI e B. SANTHI, foram destacadas as literaturas acerca de Redes Neurais, data mining, Markov Models e do modelo Fuzzy que foram usadas para prever a flutuação de valores no mercado financeiro. O objetivo é desenvolver um modelo que possa prever com acurácia os movimentos no mercado financeiro e ajudar investidores a fazerem decisões mais informadas e certas.

O estudo também busca abordar as limitações existentes nos modelos estudados através da previsão de alguns preços de ações individuais e considerando as características de cada empresa. Todas as técnicas propostas são comparadas e analisadas para determinar a efetividade delas na predição de movimentos no mercado de ações.

O survey começa elencando alguns trabalhos e indicando algumas conclusões, como a de R.K. and Pawar D.D. que indicam que uma rede neural consegue extrair informações úteis de uma grande base de dados e pode ser usada para prever séries temporais. Também aponta que Akinwale adio T, Arogundade O.T. e Adekoya Adebayo F examinaram a propagação de erros ao se usar regressão quando analisando os dados no mercado financeiro da Nigéria, onde dois métodos foram usados e comparados para que um algoritmo fosse destacado como melhor.

Entre todos os artigos levantados, os principais métodos encontrados foram média móvel, regressão e modal de ARIMA, cada um desses métodos é descrito no artigo referente.

Ao final da análise, é mostrado um algoritmo baseado nos artigos analisados que busca prever o preço de uma ação no próximo dia, apesar de nenhuma análise ter sido feita em cima do algoritmo em si.

4.5 USO DE REDES NEURAIAS RECORRENTES PARA PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS

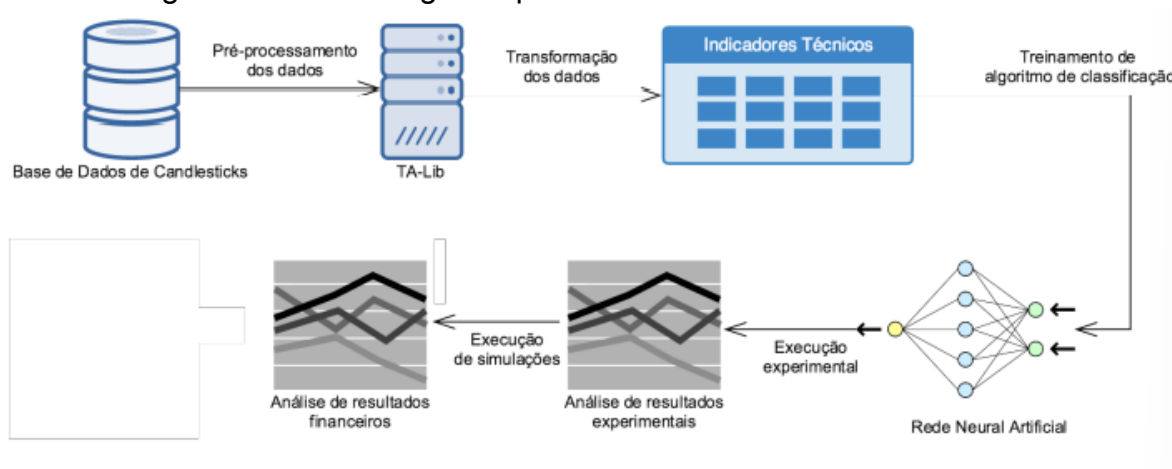
Na Dissertação realizada por David Michael Quirino Nelson (2017), o autor busca estudar a aplicação de redes Memória de Curto Prazo Longa no problema de previsão de tendências de preços de ações baseando-se no histórico de preços e outros indicadores.

O trabalho foca em analisar a previsão da variação de preço pelo ponto de vista da análise técnica, que segundo o próprio autor, se trata de “um conjunto de métodos matemáticos que se utilizam do histórico de preços e volume negociado para prever tendências futuras”, se baseando principalmente no preço histórico e no modelo de redes neurais gerado baseando-se nesses fatores, sendo o algoritmo para a mesma o de Memória de Curto Prazo Longa, que se baseia bastante nos dados a longo prazo, mas identifica tendências no curto prazo que levam bastante em consideração os acontecimentos recentes.

O intuito do projeto também foi comparar como o modelo se compara a outras abordagens muito exploradas na academia e geralmente baseadas em Aprendizado de Máquina.

O projeto foi executado conforme a imagem abaixo, onde uma base de dados passa pelo pré-processamento, é transformada para indicadores técnicos usados para o treinamento de um algoritmo de classificação, dando assim origem a uma rede neural artificial que ao ser executada gera resultados que podem ser analisados.

Figura 5 – Metodologia de previsão baseada em análise técnica



Fonte: <https://repositorio.ufmg.br/bitstream/1843/ESBF-AM2NTS/1/davidmichaelquirinonelson.pdf>

Cada etapa é mais aprofundada no trabalho citado e pode ser conferida no mesmo.

Após a criação do modelo, o projeto foca-se em analisar seu desempenho em diferentes ações, sendo elas BOVA11, BBDC4, CIEL3, ITUB4 e PETR4. Onde se conclui que o modelo considerou-se sim vantajoso para previsão financeira, constatando-se uma acurácia de 55,9% ao prever se o preço de uma determinada ação irá subir ou não no futuro imediato.

4.6 OUTROS TRABALHOS

No estudo *Fundamental Analysis via Machine Learning*, os autores Cao e You* pretendem esclarecer a utilidade dos itens de linha das demonstrações financeiras para os investidores, examinando a eficácia dos algoritmos de aprendizado de máquina na previsão de lucros corporativos usando essas

informações. Os algoritmos de aprendizado de máquina podem lidar com dados de alta dimensão e acomodar relacionamentos não lineares entre itens de linha de demonstrativos financeiros e ganhos futuros, potencialmente gerando previsões mais precisas e informativas. No entanto, esses modelos mais complexos também são mais suscetíveis ao *overfitting*, o que pode levar a um desempenho ruim fora da amostra. Para resolver isso, os autores propõem uma abordagem de aprendizado de máquina em dois estágios que usa regularização no primeiro estágio para selecionar itens de linha de demonstração financeira relevantes e, em seguida, emprega um modelo não linear no segundo estágio para prever ganhos. Os autores descobriram que sua abordagem produz previsões fora da amostra mais exatas e precisas do que os modelos de previsão de lucros existentes.

Em *Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques*, os autores propõem uma nova abordagem que usa o método de preparação de dados determinísticos de tendência (TDDP) para pré-processar dados de ações e, em seguida, aplica algoritmos ANN e SVM para prever o movimento de ações e índices de preços de ações. Os autores acham que sua abordagem produz previsões mais precisas do que os métodos existentes.

No trabalho *Web data extraction, applications and techniques*, é comentado sobre os sistemas de extração de dados da Web são aplicativos de software que extraem informações de fontes da Web. O trabalho foca no fato de que o projeto e a implementação de sistemas de extração de dados da Web apresentam vários desafios, incluindo a necessidade de experiência humana, a capacidade de processar grandes volumes de dados rapidamente, garantir a privacidade e lidar com fontes complexas e dinâmicas da Web. Neste artigo, os autores propõem uma nova abordagem que usa uma combinação de técnicas baseadas em regras e aprendizado de máquina para superar esses desafios e melhorar a precisão da extração de dados da web.

O trabalho *Uma introdução à Inteligência Computacional: Fundamentos, Ferramentas e Aplicações*, indica que a Inteligência Artificial (IA) tem sido objeto de estudo há milhares de anos, pois as pessoas buscam entender como uma pequena porção de matéria pode perceber, compreender, prever e manipular um mundo muito maior e mais complexo. AI, originalmente conhecida como *Computer Intelligence*, é uma das mais novas ciências, tendo surgido após a Segunda Guerra Mundial e

sendo cunhada em 1956. Este trabalho também explora o campo da AI, incluindo sua história e alguns dos principais desafios e aplicações. Também discute o conceito de inteligência e algumas das habilidades tipicamente associadas a ela, como a capacidade de raciocinar, aprender, perceber e se adaptar.

No trabalho *Introdução às redes neurais artificiais*. É comentado sobre como Redes neurais artificiais são uma abordagem para solucionar problemas de inteligência artificial, onde em vez de tentar programar um computador para imitar comportamentos inteligentes, se constrói um computador com circuitos que modelam os circuitos cerebrais e se espera que comportamentos inteligentes emerjam. Redes neurais também podem se auto-organizar em ambientes diferentes, criando suas próprias representações internas e apresentando comportamentos imprevisíveis.

Já no trabalho *Deep learning in the stock market—a systematic survey of practice, backtesting, and applications*, os autores indicam haver uma necessidade de desenvolvimento responsável de ML, incluindo explicabilidade e reprodutibilidade. Esta pesquisa examina a pesquisa de DL no mercado de ações, com foco em estudos que abordam explicabilidade, reprodutibilidade e praticidade. Backtesting e métricas de avaliação específicas de domínio são usadas como critérios primários, com foco no uso de redes neurais convolucionais (CNNs) e redes de memória de longo prazo (LSTM). Os resultados mostram que as CNNs são o método DL mais utilizado no mercado de ações, com foco na previsão e análise de preços de ações. Os LSTMs são menos usados, mas demonstraram ter melhor desempenho em determinadas tarefas. O estudo também destaca a necessidade de mais pesquisas sobre explicabilidade e reprodutibilidade em DL para o mercado de ações.

4.7 COMPARAÇÃO/ANÁLISE DOS TRABALHOS

A Tabela 1 faz uma comparação dos trabalhos citados no decorrer do capítulo, levando em conta os conteúdos citados em cada um.

Tabela 1 – Comparação dos tipos, modelos de IA e métricas de avaliação utilizados dos trabalhos relacionados

Título	Aprendizado de Máquina	Aplicação na Bolsa de Valores	Extração de dados
Fundamental Analysis via Machine Learning	X	X	
Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques	X	X	
Web data extraction, applications and techniques	X		X
Uma introdução à Inteligência Computacional: Fundamentos, Ferramentas e Aplicações	X		
Introdução às redes neurais artificiais.	X		

Deep learning in the stock market a systematic survey of practice, backtesting, and applications	X	X	
Sistema de análise comparativa fundamentalista utilizando inteligência artificial	X	X	X

Fonte: elaborado pelos autores

Com base na tabela acima, podemos concluir que apesar de diversos trabalhos abordarem assuntos relacionados ao presente trabalho, nenhum aborda todos juntos, sendo o mesmo ainda o único a abordar o uso de tais técnicas juntamente a uma análise fundamentalista.

5 DESENVOLVIMENTO

Neste capítulo aborda-se o desenvolvimento do modelo e propostas relacionadas à solução do problema ilustrada no fluxo de dados da Figura 6. Os modelos propostos serão validados utilizando diferentes conjuntos de dados da Bolsa de Valores em períodos diversos para comparação de padrões e possíveis generalizações que possam ocorrer através das análises.

O objetivo deste capítulo é descrever uma metodologia e estabelecer um processo estruturado para a determinação do preço justo de ações, indo desde a coleta inicial dos dados, tratamento dos mesmos, geração e comparação de séries temporais e por fim descrevendo detalhadamente o cálculo do preço justo.

O fluxo geral envolve a criação de uma ferramenta de análise de dados históricos da bolsa de valores brasileira com foco em técnicas de análise fundamentalista e boas práticas de investimento a longo prazo. O processo começa com a coleta de dados do Fundamentus Disponível em: <https://www.fundamentus.com.br/>. Acesso em: 07 nov. 2023, que é uma fonte de informações sobre empresas listadas na bolsa. Em seguida, os dados são submetidos a um pré-processamento para garantir a qualidade e consistência das informações.

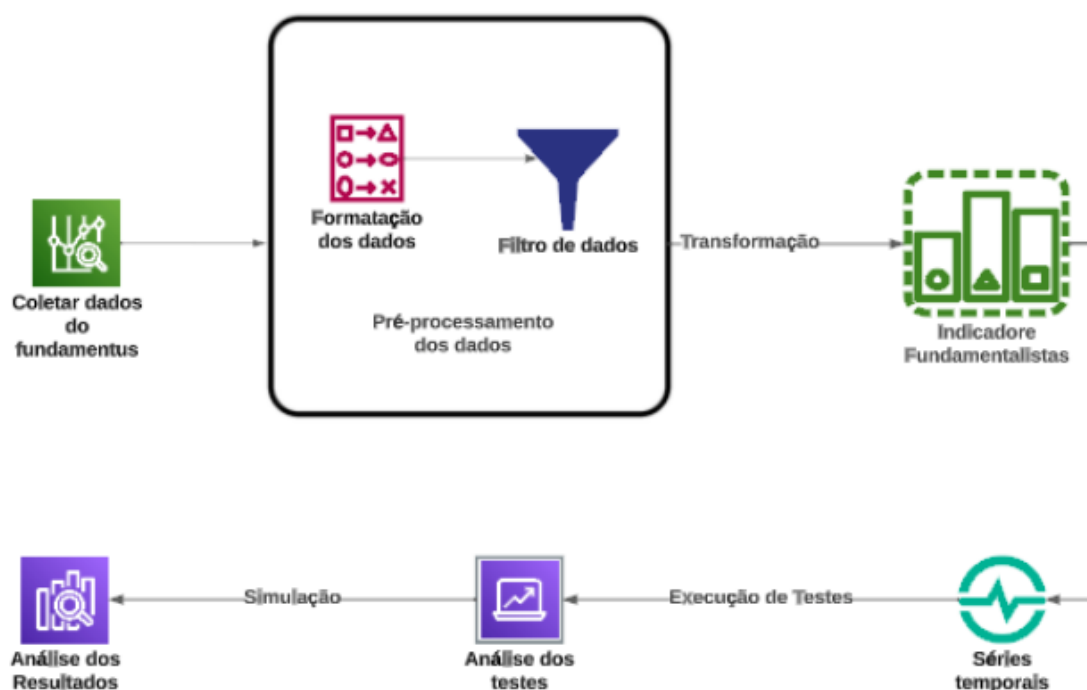
Após o pré-processamento, os dados são transformados em conjuntos de ROEs trimestrais representados no diagrama como indicadores fundamentalistas, que são métricas-chave usadas para avaliar o desempenho e a saúde financeira das empresas. Estes indicadores, juntamente com séries temporais, são utilizados para realizar testes e análises aprofundadas. Esses testes podem incluir a avaliação do desempenho passado das empresas e a identificação de oportunidades de investimento a longo prazo.

A etapa final envolve a simulação e a análise dos resultados, onde podemos tomar decisões com base nas informações obtidas. O objetivo geral desse processo é buscar valorização e retorno sobre investimento em longos períodos, visando a maximização dos ganhos no mercado de ações brasileiro. Essa ferramenta fornece uma base para tomar decisões de investimento mais informadas e alinhadas com estratégias de longo prazo. Pode-se verificar o fluxograma destas etapas na figura 6.

É importante frisar também que esta abordagem foi adotada principalmente pela dificuldade de se achar os indicadores fundamentalistas já prontos para serem

usados e com histórico na periodicidade necessária, sendo necessário toda a etapa de pré processamento de dados. Também foi escolhido seguir com séries temporais devido a quantidade de dados, que por serem poucos, se mostrou mais adequado a utilização desta técnica perante as outras estudadas no capítulo anterior.

Figura 6 – Fluxo de dados para seleção de ações baseada em análise fundamentalista



Fonte: elaborado pelos autores

5.1 FERRAMENTAS

Para a transformação dos dados foi usado o ambiente nodeJs, que é um ambiente de tempo de execução de código aberto construído sobre o motor JavaScript V8 da Google. Ele permite que os desenvolvedores executem código JavaScript no lado do servidor, o que é conhecido como JavaScript no lado do servidor (ou server-side JavaScript). Atualmente o nodeJs é utilizado em diversos tipos de aplicações.

Para infraestrutura de execução da etapa de pre-processamento de dados foi utilizado a AWS (Amazon Web Services) que é um serviço de computação em nuvem desenvolvido pela Amazon. Na AWS foram utilizados os serviços S3, que se trata de um serviço de armazenamento de arquivos, e o serviço lambda¹, que permite a execução de códigos sem a necessidade de se configurar servidores.

Para o treinamento e teste do modelo, foi utilizado o Python, que é uma linguagem de programação versátil e de alto nível amplamente adotado em diversos campos, devido à sua simplicidade e legibilidade. Sua vasta biblioteca padrão e uma comunidade ativa de desenvolvedores tornam-no uma escolha ideal para a análise de dados e aprendizado de máquina, disponibilizando várias opções para a criação de séries temporais e outras análises necessárias com facilidade.

5.2 COLETA DE DADOS DO FUNDAMENTUS

Para coleta de dados históricos de ações brasileiras, foi usado o site Fundamentus, o mesmo disponibiliza para download os dados históricos das ações em forma de planilha. Os dados coletados se referem a ações listadas na bolsa de valores e trazem informações sobre patrimônio líquido, ativos totais, passivos totais e demonstração de resultados como resultado bruto e lucro/prejuízo do período. Os dados obtidos através do site Fundamentus são divididos em trimestres, com um período de dados de mais de 10 anos. Dessa forma podemos enviar os dados para pré-processamento.

Para este trabalho, serão usados dados da empresa VALE3, que será a empresa analisada ao longo do trabalho. As informações obtidas se dividem em duas partes, sendo a primeira referente ao balanço patrimonial da empresa ao longo do tempo e a segunda referente a demonstração de resultados da mesma, também ao longo do tempo.

Abaixo é possível observar alguns dos dados principais do balanço patrimonial da empresa, na tabela obtida cada categoria de dados é subdividida em várias subcategorias, explicando cada parte da construção patrimonial da empresa, para o trabalho, como o foco será no cálculo do ROE, a principal informação

¹ Pela descrição no site Amazon, é uma forma de executar código sem necessidade de prover infraestrutura para a execução do mesmo

utilizada do balanço patrimonial será o valor de Patrimônio Líquido ao longo do tempo, dado que para a empresa VALE3, é registrado de 2008 até a atualidade.

Tabela 2 – Principais dados do balanço patrimonial da Vale

VALE3	30/09/2008	31/12/2008	31/03/2009
Ativo Total	R\$ 174.240.432,00	R\$ 185.779.470,00	R\$ 187.954.266,00
Passivo Total	R\$ 174.240.432,00	R\$ 185.779.470,00	R\$ 187.954.266,00
Patrimônio Líquido	R\$ 95.551.521,00	R\$ 96.274.637,00	R\$ 98.611.577,00

Fonte: elaborado pelos autores com dados obtidos em <https://www.fundamentus.com.br/detalhes.php?papel=VALE3>

A segunda parte, referente à demonstração de resultados da empresa, contém diversos dados referentes aos resultados obtidos pela empresa ao longo dos anos, abaixo é possível ver os principais resultados. Dentre eles, o utilizado para calcular o ROE será o Lucro/Prejuízo do período.

Tabela 3 – Principais dados da demonstração de resultados da Vale

VALE3	30/09/2008	31/12/2008	31/03/2009
Receita Líquida de Vendas e/ou Serviços	R\$ 20.697.999,00	R\$ 17.382.957,00	R\$ 12.915.840,00
Resultado Bruto	R\$ 12.056.218,00	R\$ 9.272.080,00	R\$ 6.049.073,00
Resultado Antes Tributação/Participações	R\$ 12.459.581,00	-R\$ 408.183,00	R\$ 3.801.033,00
Lucro/Prejuízo do Período	R\$ 12.433.386,00	R\$ 2.020.360,00	R\$ 3.150.903,00

Fonte: elaborado pelos autores com dados obtidos em <https://www.fundamentus.com.br/detalhes.php?papel=VALE3>

5.3 PRÉ-PROCESSAMENTO DOS DADOS

Ao chegar nesta etapa, os dados ainda precisam ser tratados e precisam dar origem a informações, por esse motivo é necessária uma formatação e filtragem.

Formatar dados para o escopo de análise fundamentalista ignorando outros valores possíveis não utilizados pela análise e filtrando os dados não necessários ou que não façam sentido para a análise.

Para filtragem dos dados, os arquivos baixados do site fundamentus serão colocados no S3 e processados, para calcular o ROE com base nas informações obtidas na seção anterior usando uma função lambda. Para calcular o ROE, a função lambda utiliza a fórmula abaixo, onde o lucro líquido do último ano, que é o somatório do lucro líquido dos últimos 4 trimestres, é dividido patrimônio líquido do trimestre analisado:

$$ROE = \frac{\text{Lucro Líquido (ano)}}{\text{Patrimônio Líquido}} \quad (1)$$

Na tabela a seguir, é possível observar um exemplo de como é feito o cálculo para o ROE de dezembro de 2022 para a empresa VALE3.

Tabela 4 – Cálculo do ROE para o período de 31/12/2022

	31/03/2022	30/06/2022	30/09/2022	31/12/2022
Lucro/Prejuízo do Período	R\$ 23.046.001,00	R\$ 30.033.000,00	R\$ 23.286.000,00	R\$ 19.559.002,00
Patrimônio Líquido				R\$ 187.112.004,00
ROE				51,27

Fonte: elaborado pelos autores

Ao calcular o ROE, a função escreve em um output os resultados de forma que eles possam ser usados na próxima etapa.

Será usado para análise o indicador fundamentalista ROE, assim como o cálculo para preço justo de Décio Bazin (1994). De forma que ambos estarão relacionados para indicar uma tendência de preço que faz sentido ser usado para comprar a ação. A seguir, apresentaremos uma pequena parte do ROE calculado para a empresa VALE3, no total foi calculado o ROE de 58 trimestres, os quais serão usados como entrada para as séries temporais.

Tabela 5 – ROE calculado para todos os trimestre da empresa Vale

Data	ROE
2009-03-31	17.85
2009-06-30	20.17
2009-09-30	9.92
2009-12-31	10.71
...	...
2023-06-30	30.45

Fonte: elaborado pelos autores

5.4 SÉRIES TEMPORAIS

Para realizar a previsão dos valores futuros dos Retornos sobre o Patrimônio Líquido (ROEs), foi empregado técnicas de análise de séries temporais com o suporte de ferramentas de Machine Learning, como Prophet, uma biblioteca Python de código aberto do Facebook, para prever valores futuros em séries temporais em áreas como finanças, saúde e meteorologia. Se destaca pela sua capacidade de lidar com sazonalidade, feriados e tendências, e ARIMA, com o auxílio do Jupyter Lab, um ambiente de desenvolvimento interativo que permite a criação de documentos que combinam código, texto explicativo e gráficos. Oferece uma interface versátil para análises de dados e programação. As séries temporais representam dados coletados ao longo do tempo, e a compreensão desses padrões temporais é crucial para as previsões. O ponto de partida envolve a importação e organização de dados trimestrais de ROEs, que previamente foram tratados e armazenados em arquivos CSV.

5.4.1 PROPHET

Para o modelo do Prophet foi utilizado a biblioteca Pandas, esses dados são estruturados em um DataFrame. Posteriormente, o DataFrame foi dividido em um conjunto de treinamento e um conjunto de validação, onde a maioria dos dados do Data Frame são utilizados para treinamento do modelo (no exemplo da figura são utilizados 56 trimestres armazenados na variável `training_df`) para treinamento e o restante para validação (2 trimestres restantes armazenados na variável

validation_df), fundamentais para o ajuste dos modelos, a decisão de utilizar apenas dois trimestres para o conjunto de validação se dá pela pouca quantidade de dados para um balanceamento maior entre os conjuntos. A criação destes DataFrames é demonstrada pela figura 7.

Figura 7 – Criação dos DataFrames de teste e validação.

```
def roe_forecast():  
  
    # Caminho do arquivo de dados da ação  
    stock_path = 'Data/VALE3.csv'  
    print('Time series forecast for ' + stock_path)  
  
    # Carrega os dados do arquivo CSV  
    df = pd.read_csv(stock_path)  
    df.head()  
  
    validation_size = 2  
  
    # Dados para treinamento = DataFrame - tamanho da validação  
    training_df = df.head(len(df) - validation_size)  
    print(*values: "Training:", training_df)  
  
    validation_df = df.tail(validation_size)
```

Fonte: elaborado pelos autores

A partir destes DataFrames foi realizado a predição como demonstrado na figura 8.

Figura 8 – Criação da instância do Prophet, ajuste do modelo e previsão dos ROEs futuros através do modelo treinado.

```
# Cria uma instância do modelo Prophet
m = Prophet()

# Ajusta o modelo aos dados
m.fit(training_df)

# Gera datas futuras para previsão
future = m.make_future_dataframe(periods=len(validation_df), freq='Q')
print(future.tail())

# Realiza a previsão
forecast = m.predict(future)
print(forecast.tail())

# Exibe os resultados da previsão
print(forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail())

predictions = forecast.tail(len(validation_df))
```

Fonte: elaborado pelos autores

Foi criada uma instância do modelo Prophet e armazenada na variável "m".

A função "Prophet()" possui parâmetros opcionais que permitem ajustar o modelo conforme necessário. Para realizar o ajuste, foi utilizada a função "fit()", que recebe como parâmetro o DataFrame de treinamento. Nesse processo, o Prophet analisa os dados históricos, aprendendo os padrões de tendência essenciais para as previsões. Em resumo, o Prophet realiza um ajuste de tendência com base nos dados de treinamento, representando o comportamento geral da série temporal ao longo do tempo, seguindo, por padrão, um modelo de tendência linear.

Além disso, o modelo identifica sazonalidades ao analisar padrões entre diferentes períodos, como semanais e anuais. Automaticamente, ajusta os parâmetros do modelo, como pontos de mudança na tendência, coeficientes de sazonalidade e forças de regularização, otimizando assim o ajuste aos dados de treinamento. Posteriormente, com os parâmetros ajustados, o Prophet realiza o treinamento do modelo de previsão de séries temporais com base no DataFrame fornecido.

Para gerar previsões futuras, foi utilizado a função "make_future_dataframe" do modelo "m", fornecendo dois parâmetros: a quantidade de períodos desejados para a previsão (neste caso, equivalente ao tamanho do conjunto de validação) e a frequência desejada, como "Q" para trimestre em inglês. Isso resulta em um Data Frame pronto para receber os valores da predição. A predição é então realizada a partir do modelo "m" usando a função "predict()", que recebe como parâmetro o Data Frame "future" criado anteriormente. A função retorna um Data Frame com 16 colunas, sendo as principais as datas (ds), a previsão (yhat) que representa o ROE no período, os limites inferior e superior dos intervalos de confiança (yhat_lower e yhat_upper), e outros componentes da previsão, como tendência e sazonalidade (tren e yearly).

Conforme o modelo utilizado, mesmo com uma tendência negativa, as previsões ('yhat') aumentam, indicando uma possível compensação dos efeitos sazonais anuais, como é possível ver pelos últimos registros dos dados previstos na figura 9.

Figura 9 – Dados de saída da função predict.

```
[5 rows x 16 columns]
```

	ds	yhat	yhat_lower	yhat_upper	trend	yearly
53	2022-06-30	24.886481	0.700401	49.994537	-17.441854	42.328335
54	2022-09-30	24.723360	-0.034455	50.817097	-17.067130	41.790490
55	2022-12-31	22.209709	-2.827499	47.304914	-16.692406	38.902116
56	2023-03-31	29.361914	5.163645	53.099609	-16.325828	45.687742
57	2023-06-30	29.965645	8.173585	53.807946	-15.955177	45.920822

Fonte: elaborado pelos autores

A partir destes dados foi realizada a validação RMSE, sigla do inglês root mean-square error que traduzido é raiz do erro quadrático médio, para medir o desempenho de modelos de regressão. Possui a seguinte fórmula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}$$

(2)

Onde:

- n é o número de amostras
- y_i é o valor observado para a amostra i
- p_i é o valor previsto pelo modelo para a amostra i

Informações obtidas em ([RMSE \(Raiz Do Erro Quadrático Médio\) Em Machine Learning | Mario Filho](#)) Acessado em: 07 de nov. 2023.

No código foi realizado o cálculo através da função "rmse()", que recebe como parâmetro os valores do ROEs preditos e como segundo parâmetro os ROEs do conjunto de validação retornando então o RMSE:

Figura 10 – Retorno da função rmse.

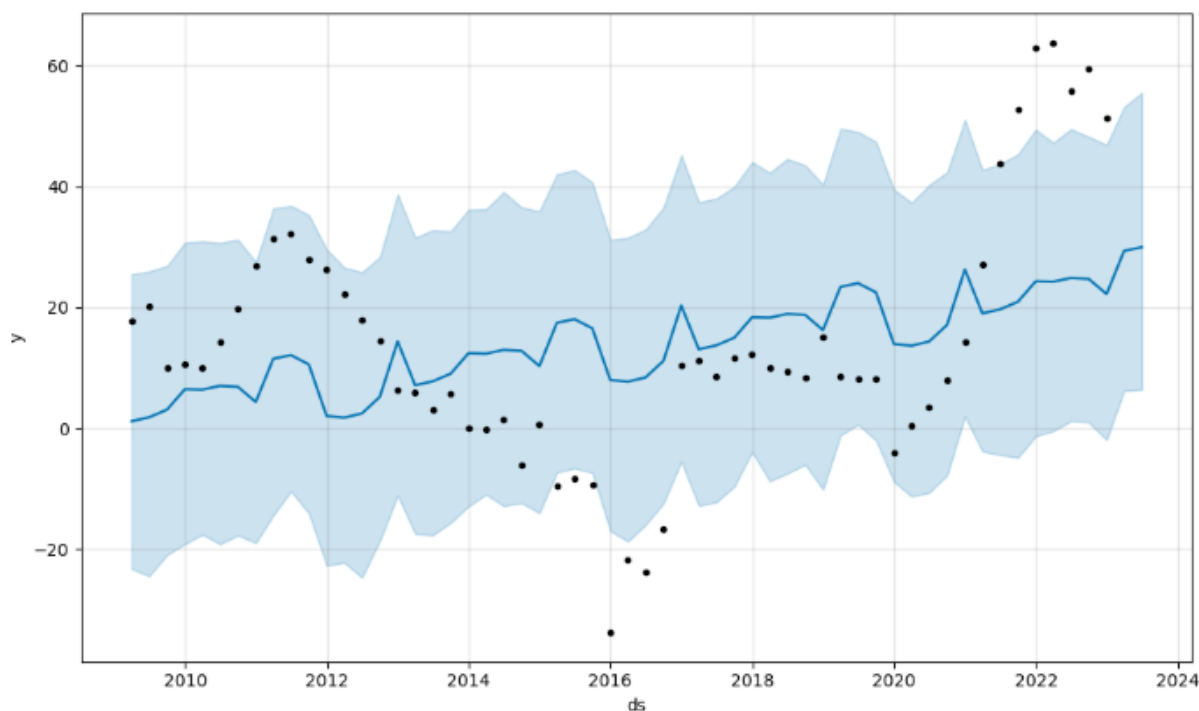
```
Validation:          ds          y
56  2023-03-31    43.44
57  2023-06-30    30.45
Dados da predição para o calculo do RMSE:
  56    29.361914
  57    29.965645
Name: yhat, dtype: float64
RMSE predito:  9.960600083933768
```

Fonte: elaborado pelos autores

Isso significa que o modelo do Prophet comete em média o erro de 9 pontos percentuais comparado com os valores reais, para o conjunto de 56 trimestres de

treino e 2 trimestres de validação. Por fim, foram construídos gráficos de previsão, tendência e sazonalidade para visualização dos dados.

Figura 11 – Gráfico de previsão para o modelo Prophet.

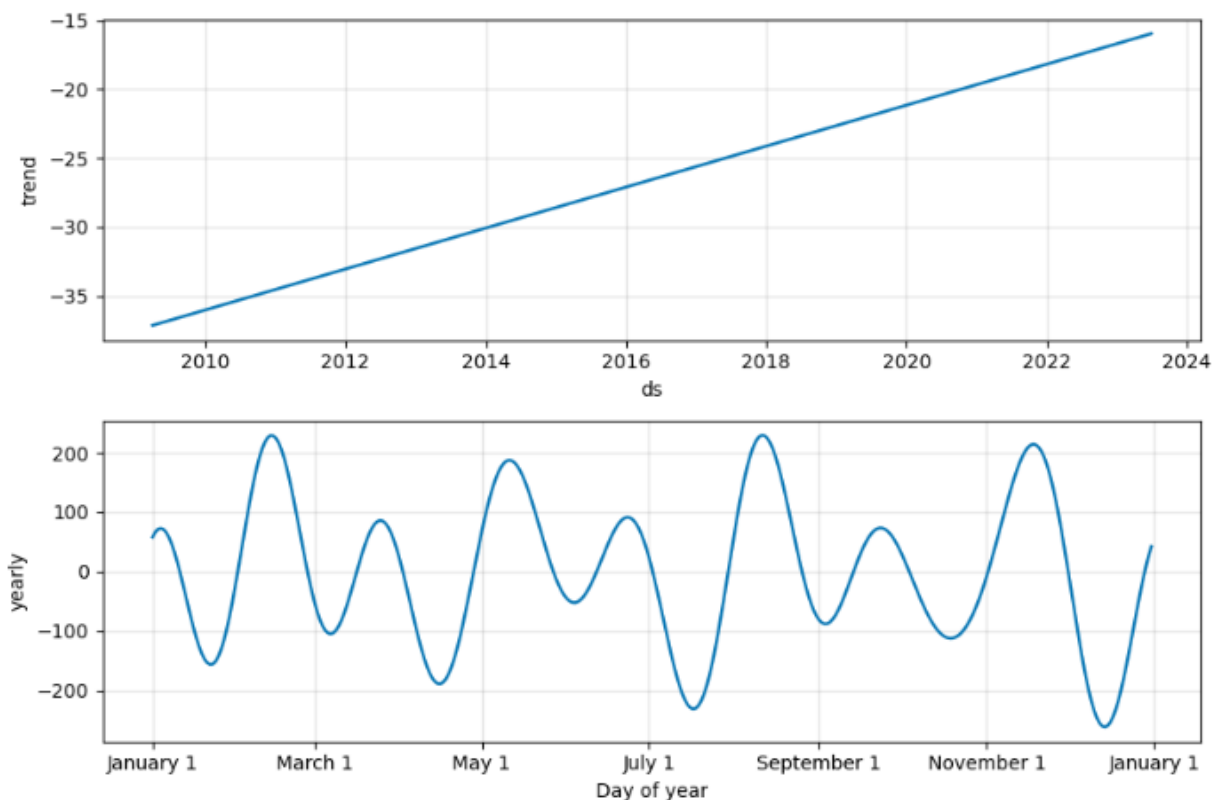


Fonte: elaborado pelos autores

Eixo y sendo o ROE e o eixo x sendo os trimestres, os pontos pretos representam os valores reais do conjunto de treinamento, já a linha azul representa os ROEs previstos pelo modelo. No gráfico também podemos ver uma zona em azul claro representada pelo limite inferior e o limite superior do intervalo de confiança para as previsões. Estes intervalos de confiança indicam a incerteza associada às previsões e são úteis para entender a variabilidade possível nos resultados.

Além deste gráfico de previsão foi feito também os gráficos de tendência e a sazonalidade respectivamente:

Figura 12 – Gráfico de tendência e Sazonalidade.



Fonte: elaborado pelos autores

Para entender melhor como interpretar estes gráficos é necessário entender mais a fundo como o Prophet encontra as previsões. O modelo é formulado como:

$$y(t)=g(t)+s(t)+h(t)+\epsilon t$$

onde:

- $g(t)$ representa a tendência.
- $s(t)$ representa a sazonalidade.
- $h(t)$ representa outros efeitos adicionais (feriados, por exemplo).
- ϵt é o termo de erro.

Portanto, para o modelo utilizando Prophet temos que a tendência negativa à medida que avançamos no tempo diminui, e a tendência se aproxima de zero. Isso pode indicar que, ao longo do tempo, a diminuição nas observações está se desacelerando como vistos na figura 10, o que indica que valores futuros serão maiores, pois o desconto da tendência será menor. Já a sazonalidade mostra

padrões ou variações que ocorrem regularmente a cada ano. Em outras palavras, ela captura a influência de eventos ou padrões que se repetem em ciclos anuais. Através da figura 10 é possível ver que períodos entre janeiro e março possuem alta influência positiva nos resultados, assim como julho e setembro, já no período entre novembro e janeiro possui uma variação muito grande com um pico muito alto de influência da sazonalidade e uma queda brusca para início do próximo ano, portanto os valores preditos nestes períodos devem ser analisados com mais cautela.

Tabela 6 – Tabela dos dados preditos pelo modelo do Prophet

Data (Trimestre)	ROE Previsto	Tendência	Sazonalidade anual
2009-03-31	1.172917	-37.151509	38.324426
2009-06-30	1.835088	-36.780859	38.615948
2009-09-30	3.093416	-36.406136	39.499551
2009-12-31	6.480837	-36.031413	42.512250
...
2023-06-30	29.965645	-15.955177	45.920822

Fonte: elaborado pelos autores

Obs. Esta tabela possui 58 trimestres preditos pelo modelo, como exemplo estamos utilizando os 4 primeiros registros e o último.

Tabela 7 – Tabela com os dados para o cálculo do RMSE.

Data (trimestre)	ROE previsto	ROE de validação
2023-03-31	29.361.914	43.44
2023-06-30	29.965.645	30.45
Total RMSE	9.9606000839 33760	

Fonte: elaborado pelos autores

5.4.2 ARIMA

O ARIMA também considera componentes como tendência e sazonalidade, visando capturar os padrões temporais presentes nos dados. Diferente do Prophet para o ARIMA precisamos passar alguns parâmetros para ajuste do modelo representados na função $ARIMA(p, d, q)$ onde (p) é a ordem do componente autorregressivo (número de lags anteriores), (d) Ordem de diferenciação (número de vezes que os dados são diferenciados para tornar a série estacionária) e (q) a ordem do componente de médias móveis (número de lags anteriores do termo de erro). Assim como no Prophet, foi utilizado a biblioteca Pandas para criação dos DataFrames a partir do arquivo .csv. Foram separados os dados em dois DataFrames um para treinamento e outro para validação. Em sequência, foi validado a estacionariedade da série através da função `stationary_check` como na figura 13. Esta validação é importante para definição do parâmetro (d), caso a série seja não-estacionária pode-se optar por configurar este parâmetro de diferenciação para estacionarizar a série.

Figura 13 – Retorno da função `rmse`.

```
[16]: def stationary_check(ts):
      # Média móvel
      roll_mean = ts.rolling(12).mean()
      plt.plot(ts, color='green', label='Original')
      plt.plot(roll_mean, color='blue', label='Média móvel')
      plt.legend(loc='best')
      plt.title('Média móvel')
      plt.show(block=False)
      # (ADF)Augmented Dickey-Fuller test:
      print('Augmented Dickey-Fuller test:')
      df_test = adfuller(ts)
      df_output = pd.Series(df_test[0:4], index=['Test Statistic', 'p-value', '#Lags', 'Número de observações'])
      print("df_output: \n", df_output)
      for key, value in df_test[4].items():
          df_output['Critical Value (%s)'%key] = value
      print(df_output)
```

Fonte: elaborado pelos autores

Retornando um gráfico com a média móvel representado pela figura 14, e os dados do ADF (Augmented Dickey-Fuller) é um teste de raiz unitária (unit root test) usado para verificar se uma série temporal é estacionária ou não estacionária representado na figura abaixo.

Figura 14 – Gráfico da média móvel do modelo ARIMA.



Fonte: elaborado pelos autores

Figura 15 – Dados retornados pelo teste ADF.

```

Augmented Dickey-Fuller test:
df_output:
  Test Statistic      -0.954107
  p-value             0.769633
  #Lags               4.000000
  Número de observações 51.000000
  dtype: float64
Test Statistic      -0.954107
p-value             0.769633
#Lags               4.000000
Número de observações 51.000000
Critical Value (1%) -3.565624
Critical Value (5%) -2.920142
Critical Value (10%) -2.598015
dtype: float64

```

Fonte: elaborado pelos autores

Dessa forma pode-se interpretar através do valor de p-value que a série é não estacionária, portanto possui tendências mais acentuadas. Em seguida utiliza-se o "auto_arima" para identificar qual o melhor ordem para modelo ARIMA:

Figura 16 – Resultado informações modelo ARIMA.

```

•[32]: from pmdarima import auto_arima

import warnings
warnings.filterwarnings("ignore")

•[34]: stepwise_fit = auto_arima(training_df['y'], trace=True, suppress_warnings=True)
stepwise_fit.summary()

Performing stepwise search to minimize aic
ARIMA(2,0,2)(0,0,0)[0] intercept : AIC=388.983, Time=0.04 sec
ARIMA(0,0,0)(0,0,0)[0] intercept : AIC=501.893, Time=0.00 sec
ARIMA(1,0,0)(0,0,0)[0] intercept : AIC=393.985, Time=0.01 sec
ARIMA(0,0,1)(0,0,0)[0] intercept : AIC=inf, Time=0.02 sec
ARIMA(0,0,0)(0,0,0)[0] : AIC=519.915, Time=0.00 sec
ARIMA(1,0,2)(0,0,0)[0] intercept : AIC=388.801, Time=0.02 sec
ARIMA(0,0,2)(0,0,0)[0] intercept : AIC=427.293, Time=0.03 sec
ARIMA(1,0,1)(0,0,0)[0] intercept : AIC=395.303, Time=0.01 sec
ARIMA(1,0,3)(0,0,0)[0] intercept : AIC=inf, Time=0.04 sec
ARIMA(0,0,3)(0,0,0)[0] intercept : AIC=inf, Time=0.04 sec
ARIMA(2,0,1)(0,0,0)[0] intercept : AIC=inf, Time=0.03 sec
ARIMA(2,0,3)(0,0,0)[0] intercept : AIC=inf, Time=0.05 sec
ARIMA(1,0,2)(0,0,0)[0] : AIC=389.045, Time=0.01 sec

Best model: ARIMA(1,0,2)(0,0,0)[0] intercept
Total fit time: 0.312 seconds

```

Fonte: elaborado pelos autores

Então é realizado o treinamento do modelo através da função ARIMA:

Figura 17 – Treinamento do modelo com ARIMA.

```

•[35]: model = sm.tsa.arima.ARIMA(training_df.y, order=(1,0,2))
results = model.fit()
plt.plot(training_df.y)
plt.plot(results.fittedvalues, color='red')
plt.title('Fitting data _ MSE: %.2f%' % ((results.fittedvalues-training_df.y)**2).mean())
plt.show()

```

Fonte: elaborado pelos autores

A função recebe por parâmetro o dataframe de treino e a ordem considerada pelo melhor modelo anteriormente, então foi plotado um gráfico com os dados ajustados pelo modelo em contraste com os dados de treino, calculando o MSE (Mean squared error):

Figura 18 – Gráfico com dados ajustados.



Fonte: elaborado pelos autores

Realizando a transformação para RMSE calculando $\sqrt{48.46}$ se tem o valor aproximado de 6.98 o que indica uma correlação maior que o modelo do Prophet de aproximadamente 9.96 se considerado apenas os dados ajustados do modelo. Para uma comparação mais justa é preciso pegar os valores preditos pelo modelo ARIMA:

Figura 19 – Valores preditos ARIMA.

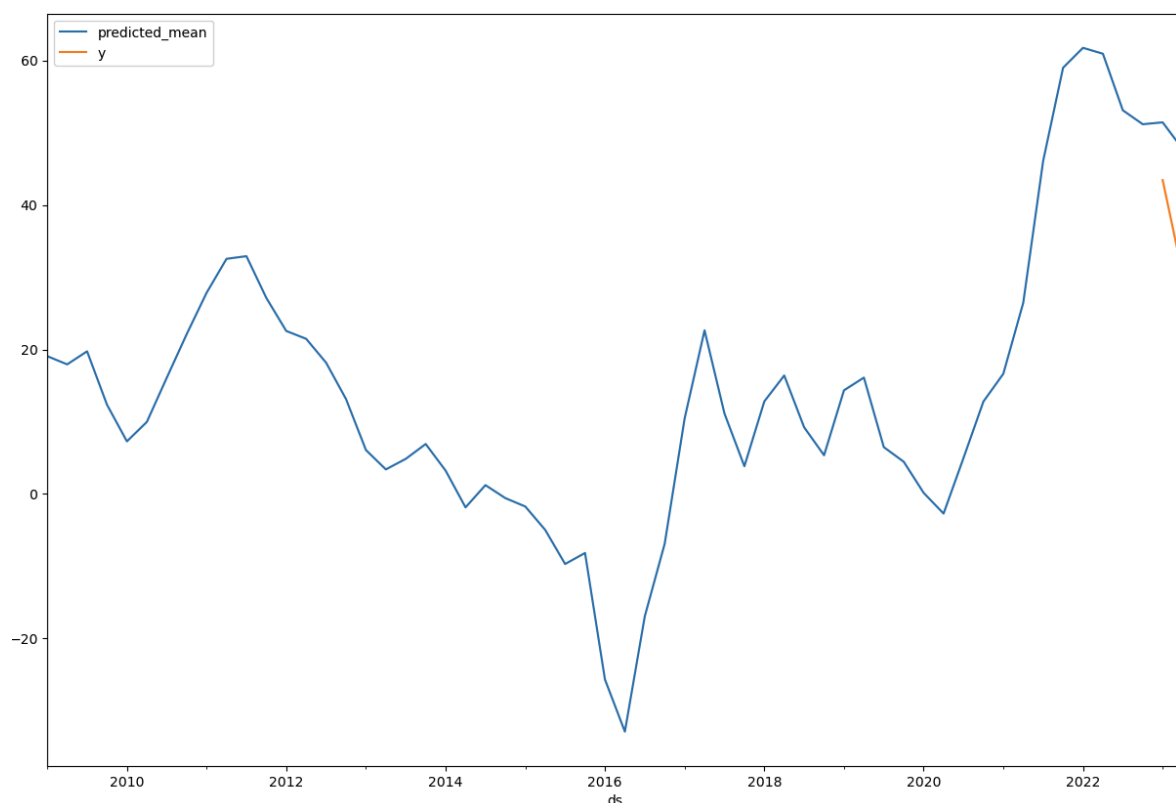
```
[22]: predictions = results.predict('2009-03-31', '2023-06-30')
[37]: predictions.tail()
[37]: 2022-06-30    65.068460
      2022-09-30    55.261356
      2022-12-31    59.726761
      2023-03-31    50.302692
      2023-06-30    49.729642
      Freq: Q-DEC, Name: predicted_mean, dtype: float64
```

Fonte: elaborado pelos autores

Para gerar os ROEs previstos pelo modelo utilizamos a função "predict()"

passando como parâmetro os índices de início e fim, como ajustamos o Data Frame para utilizar as datas como índices podemos passar a data do primeiro trimestre onde possuímos valor de ROE para VALE3 e a data de fim com o último trimestre do nosso conjunto de validação para compararmos seus valores.

Figura 20 – Gráfico gerado ARIMA para predição e validação do modelo.



Fonte: elaborado pelos autores

Em azul os valores preditos pelo modelo em conjunto com os valores de validação em laranja, com os últimos 2 trimestres. Comparando os valores preditos com os valores de validação temos que:

Figura 21 – Comparação de valores de predição e valores de validação ARIMA.

```
[36]: from sklearn.metrics import mean_squared_error
      from math import sqrt
      print("Trimestres preditos para validação: \n", predictions.tail(len(validation_df)))
      print("Conjunto para validação: \n", validation_df)
      rmse=sqrt(mean_squared_error(predictions.tail(len(validation_df)), validation_df['y']))
      print("RMSE: ", rmse)

Trimestres preditos para validação:
  2023-03-31    51.427026
  2023-06-30    47.746058
Freq: Q-DEC, Name: predicted_mean, dtype: float64
Conjunto para validação:
      y
ds
2023-03-31    43.44
2023-06-30    30.45
RMSE:  13.47119554829159
```

Fonte: elaborado pelos autores

RMSE para o modelo ARIMA utilizando os 2 semestres preditos fora do conjunto de treino em comparação com o conjunto de validação representa um erro de aproximadamente 13.47 pontos percentuais comparado com os 9.96 do Prophet. Essa etapa de validação é essencial para a avaliação da precisão do modelo preditivo, fornecendo insights valiosos sobre a capacidade do modelo em antecipar os valores futuros dos ROEs com base no histórico de dados disponíveis.

Tabela 8 – Tabela dos dados preditos pelo modelo do ARIMA.

Data (Trimestre)	ROE Previsto
2009-03-31	19.071905
2009-06-30	17.937828
2009-09-30	19.740456
2009-12-31	12.346083
...	...
2023-06-30	47.746058

Fonte: elaborado pelos autores

Obs. Esta tabela possui 58 trimestres preditos pelo modelo, como exemplo estamos utilizando os 4 primeiros registros e o último.

Tabela 9 –Tabela com os dados para o cálculo do RMSE.

Data (trimestre)	ROE previsto	ROE de validação
2023-03-31	51.427026	43.44
2023-06-30	47.746058	30.45
Total RMSE	13.47119554829150	

Fonte: elaborado pelos autores

5.4.3 DISCUSSÃO

Os modelos Prophet e ARIMA representam ferramentas robustas para a previsão de séries temporais, cada um com atributos distintos. O Prophet, desenvolvido pelo Facebook, se destaca pela flexibilidade na abordagem de padrões temporais complexos, incluindo tendências não lineares e múltiplas sazonalidades. A implementação de forma intuitiva e eficiência computacional tornam este modelo uma ferramenta acessível. No entanto, suas simplificações, podem limitar sua eficácia em cenários mais intrincados.

Por outro lado, o ARIMA é uma abordagem clássica com uma sólida fundamentação teórica, mostrando eficácia em séries temporais estacionárias. Seus componentes auto regressivos, integradas e de médias móveis são capazes de modelar adequadamente os padrões temporais em contextos mais tradicionais. Contudo, a aplicação tende a ser mais desafiadora em situações com tendências não lineares e sazonalidades complexas, exigindo um entendimento mais aprofundado na escolha dos parâmetros.

A decisão entre o Prophet e o ARIMA é pautada na natureza específica dos dados e na complexidade dos padrões temporais. O Prophet destaca-se em cenários mais flexíveis e automatizados, demonstrando um desempenho superior no conjunto de dados considerado neste trabalho. Em contrapartida, o ARIMA é mais adequado para contextos teóricos bem definidos e séries estacionárias, onde pode apresentar um desempenho melhor, ao contrário do que vimos nas nossas análises onde a utilização de um conjunto de dados não estacionário prejudica este modelo.

Ambos os modelos oferecem abordagens únicas para a modelagem de séries temporais, e a escolha deve ser direcionada pela aplicação específica e pelas características dos dados em questão. Para este cenário proposto, onde envolvemos séries não estacionárias, tendências e sazonalidades, o Prophet se demonstrou superior em suas previsões.

5.5 CÁLCULO DO PREÇO JUSTO

O conceito de preço justo de uma ação é fundamental para determinar se um investimento é atrativo ou não. Neste projeto, o cálculo do preço justo é abordado de maneira inovadora, combinando técnicas de séries temporais com a fórmula de preço justo proposta por Décio Bazin em 1994. Esta abordagem híbrida busca oferecer uma análise mais dinâmica e ajustada às tendências do mercado. A análise desta série temporal visa identificar tendências de crescimento ou declínio no desempenho da empresa. Com base nessa tendência, ajusta-se o valor calculado pelo método de Décio Bazin.

O método de Décio Bazin calcula o preço justo de uma ação com base em seu dividendo por ação e a taxa de retorno desejada. Neste projeto, propõe-se uma modificação neste cálculo: se a série temporal do ROE indica uma tendência de crescimento no próximo trimestre, sugere-se acrescentar 10% ao valor do preço justo calculado pela fórmula de Bazin. Esse acréscimo reflete a expectativa de que a empresa continue melhorando seu desempenho, justificando um investimento a um preço ligeiramente mais elevado.

Por outro lado, se a série temporal indica uma tendência de queda no ROE, recomenda-se reduzir o preço justo em 10%. Este ajuste para baixo reflete a precaução necessária em investir em uma empresa cujo desempenho financeiro está em declínio, indicando que o preço atual da ação pode não ser justificável.

Com isso, é possível gerar duas fórmulas, uma para quando a série temporal indicar crescimento e outra para quando a série temporal indicar declínio no valor do ROE:

- Fórmula para indicativa de crescimento do valor do ROE:

$$\text{Valor justo: } \text{Valor Justo (Décio Bazin)} + 10\% \quad (3)$$

- Fórmula para indicativa de declínio do valor do ROE:

$$\text{Valor justo: Valor Justo (Décio Bazin) - 10\%} \quad (4)$$

Sendo que o Valor Justo (Décio Bazin) é calculado levando em conta o valor pago em dividendos nos últimos 12 meses, dividido pela taxa de juros que se deseja obter de retorno. Para este trabalho, usamos uma taxa de juros baseada na SELIC, que é a taxa básica de juros da economia brasileira, estabelecida pelo Comitê de Política Monetária (COPOM) do Banco Central do Brasil. Esta taxa é usada como referência para as operações de empréstimos e investimentos no país, influenciando diretamente todas as outras taxas de juros, como os juros para empréstimos pessoais, financiamentos e rendimentos de investimentos em renda fixa. Obtendo o seguinte cálculo para Valor Justo (Décio Bazin) atual, levando em conta a SELIC em dezembro de 2023 a 12.25% ao ano:

$$\text{Valor justo (Décio Bazin): } \frac{\text{Dividendos(12 meses anteriores)}}{12,25 \%} \quad (5)$$

Essa abordagem integrada para o cálculo do preço justo oferece uma ferramenta mais robusta e adaptativa para a análise de investimentos, permitindo aos investidores tomar decisões mais informadas e alinhadas com as tendências de mercado e a saúde financeira das empresas.

5.6 ANÁLISE RESULTADOS

Para realizar as análises de resultados, é fundamental comparar os desempenhos reais da ação em relação às previsões do modelo proposto. A validação desse modelo será conduzida utilizando dados de junho de 2023 juntamente com o modelo Prophet, que como visto nas seções anteriores se demonstrou mais preciso para estes conjuntos de dados. Vamos realizar a comparação dos valores previstos pelo modelo com os valores reais para o papel da

VALE3. Utilizaremos o cálculo do preço justo para identificar o limite de preço ideal para investimentos de longo prazo.

Ao considerar a diferença negativa entre os valores previstos pelo Prophet para o ROE no trimestre terminado em 30 de junho de 2023 (28,157302) e o trimestre terminado em 30 de setembro de 2023 (26,626063), observamos uma variação negativa de 1,531239 pontos percentuais. Diante desse resultado, aplicamos uma correção de -10% sobre o preço justo, conforme o cálculo realizado utilizando o método de Décio Bazin (1994) explicado na seção anterior.

Devido ao fato de a taxa Selic em junho de 2023 ser 13,75% ao ano, ajustamos o cálculo do valor justo de Décio Bazin, para refletir um ganho esperado com base nesta taxa, podendo ser observado o cálculo na fórmula abaixo:

$$\text{Valor justo (Décio Bazin): } \frac{\text{Dividendos(12 meses anteriores)}}{13,75 \%} \quad (6)$$

Tabela 10 – Valores relacionados ao cálculo do preço justo segundo modelo proposto.

Data	ROE previsto	Delta ROE do próximo trimestre e atual	Percentual a considerar na fórmula do preço justo	Preço real da Ação	Preço justo calculado	Dividendos (12 meses anteriores)
2023-06-30	28,157302	-1,531239	-10%	64,22	37,25	5,69171539
2023-09-30	26,626063	67,58

Fonte: elaborado pelos autores

Considerando o valor da ação VALE3 em 30 de junho de 2023 como 64,22. Temos que o preço justo calculado é inferior ao valor da ação neste momento, sendo assim segundo o modelo a ação não está em uma faixa de preço ideal para investimentos de longo prazo, onde a faixa seria valores inferiores ou iguais a 37,25 reais por ação de VALE3 para uma rentabilidade esperada de 13,75% ao ano, onde o investimento se ajusta aos títulos de renda fixa no período utilizado.

6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo analisar o estado da arte em extração e transformação de dados, desenvolver um modelo de predição para identificar flutuações nos valores de ações de empresas selecionadas, treinar o modelo com base em dados históricos de análise fundamentalista, definir e realizar um experimento para avaliar a capacidade de análise do modelo gerado e comparar os resultados com a valorização real do mercado, isso foi feito com base em boas práticas de investimento a longo prazo e para buscar valorização e retorno sobre o investimento analisado.

Ao longo do desenvolvimento deste TCC, várias etapas foram cumpridas, desde a coleta inicial de dados, a definição de um modelo preditivo e a validação do mesmo em comparação aos valores reais do mercado. A metodologia adotada envolveu a coleta de dados, o pré-processamento para extração das informações necessárias, a utilização destas informações para a criação de séries temporais e a realização de análises aprofundadas com base nas séries temporais em conjunto com o cálculo do preço justo para avaliar o desempenho da empresa VALE3 e identificar o preço de compra justo para a mesma. Desta forma, os objetivos gerais e específicos propostos no início do projeto foram atingidos. Apesar do preço justo não satisfazer o preço da ação e da mesma ter apresentado uma certa valorização durante o período analisado, vale ressaltar que o modelo visa encontrar um preço justo para o longo prazo, e variações em um curto período de tempo podem ter diversos motivos que não podem ser medidos pela fórmula proposta.

Durante a elaboração deste trabalho, alguns desafios surgiram. A ausência de um conjunto de dados contendo as informações essenciais para a aplicação do modelo demandou a criação do mesmo, utilizando a extração manual de dados de sites relacionados a indicadores fundamentalistas, de forma que foi extraído dados abrangendo um período considerável de em média 14 anos, sendo estes de significativa importância para viabilizar uma análise do modelo proposto e os cálculos necessários para atingir o indicador fundamentalista escolhido, o ROE. Porém, uma análise mais precisa exigiria mais dados trimestrais, os quais não estão disponíveis diretamente em fontes de dados públicas.

O método criado demonstra um passo a passo de como chegar a um preço justo de compra, o qual pode ser levado em conta no momento de se investir em

empresas às quais já se analisam como interessantes a longo prazo, vale ressaltar que apenas o cálculo de preço justo gerado neste trabalho, não deve ser usado de forma isolada para se analisar uma ação, sendo necessário também um estudo mais aprofundado sobre a mesma e todo o seu contexto em que ela está inserida.

Apesar do trabalho ter atingido todos os objetivos propostos, ainda há muito espaço para melhoria e evolução do modelo gerado, abaixo estão listados alguns próximos trabalhos possíveis:

- Automatização dos passos gerados no modelo, de forma que o input seja apenas uma ação e todo o processo aconteça de forma automática, podendo ser inclusive agendado para acontecer em diferentes datas
- Melhoria do Cálculo de Preço Justo, podendo-se utilizar outros métodos descritos na literatura, para comparar e possivelmente atingir um resultado mais satisfatório.
- Disponibilização dos resultados em um Site ou Aplicativo, de forma a disponibilizar os resultados para o público geral, não só tornando a ferramenta acessível, mas também permitindo personalização e análises conforme a necessidade e preferência individual.
- Análise de mais ações e outros setores, embora o trabalho atual tenha focado apenas em uma ação e um setor, a análise poderia ser expandida para outras ações do mesmo setor e de setores diferentes, trazendo características de cada um para à análise e possivelmente gerando um resultado mais abrangente.

REFERÊNCIAS

1. **BARRETO, J. M. Introdução às Redes Neurais Artificiais. V Escola Regional de Informática. Sociedade Brasileira de Computação, Regional Sul, Santa Maria, Florianópolis, Maringá, p. 5–10, 2002.**
2. **BAZIN, Décio. Faça Fortuna com Ações, Antes que Seja Tarde. 2ª edição, 1994, JMJ, São Paulo-SP.**
3. **Brazilian Business Review. Disponível em: <https://www.scielo.br/j/bbr/a/4grKtNqy5fWjy39NychnHmL/?format=pdf&lang=pt> Acessado em 07 de novembro 2023**
4. **Emilio Ferrara, Pasquale De Meo, Giacomo Fiumara, Robert Baumgartner. (2014) Web Data Extraction, Applications and Techniques: A Survey, Knowledge-Based Systems, Volume 70, Pages 301-323, ISSN 0950-7051. Disponível em: <https://doi.org/10.1016/j.knosys.2014.07.007>.**
5. **GOLDSCHMIDT, Ronaldo Ribeiro. Uma Introdução à Inteligência Computacional: Fundamentos, Ferramentas e Aplicações. 1. ed. Rio de Janeiro: Instituto Superior de Tecnologia do Rio de Janeiro, 2010.**
6. **Gomes, Daniel T. (2005) Redes Neurais Recorrentes Para Previsão de Séries Temporais de Memórias Curta e Longa. São Paulo, Brasil. UNICAMP.**
7. **Kim, K. (2003) 'Financial Time Series Forecasting Using Support Vector Machines', Neurocomputing, vol. 55, pp. 307-319.**
8. **Kim, K. (2004) 'Toward Global Optimization of Case-Based Reasoning Systems for Financial Forecasting', Applied Intelligence, vol. 21, no. 3, pp. 239-249.**
9. **Luger, George F. (2013) Inteligência Artificial. 6 ed. São Paulo, Brasil. Pearson Education.**

10. Maddala, G. S. (1992) Introduction to Economics. Macmillan Publishing Company, New York, Toronto.
11. O que é o AWS Lambda? Amazon, 2023. Disponível em: https://docs.aws.amazon.com/pt_br/lambda/latest/dg/welcome.html. Acesso em: 07, dez. de 2023.
12. P. D. Yoo, M. H. Kim and T. Jan (2005) "Financial Forecasting: Advanced Machine Learning Techniques in Stock Market Analysis," 2005 Pakistan Section Multitopic Conference, Karachi, Pakistan, pp. 1-7, doi: 10.1109/INMIC.2005.334420.
13. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015) "Predicting Stock and Stock Price Index Movement using Trend Deterministic Data Preparation and Machine Learning Techniques", Expert Systems with Applications, 42, 259-268. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417414004473>.
14. PEREIRA, Luciano de Santana. Metodologia para Avaliar Técnicas de Redução de Protótipos: Protótipos Gerados versus Protótipos Selecionados. Recife, 2013. Dissertação (Pós-graduação em Ciência da Computação) - UNIVERSIDADE FEDERAL DE PERNAMBUCO, 2013. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/12402/1/Disserta%C3%A7ao%20Luciano%20Pereira.pdf>. Acesso em: 07 nov. 2023.
15. Preethi, G., & Santhi, B. (2005). "Stock Market Forecasting Techniques: A Survey." International Journal of Theoretical and Applied Information Technology.
16. PROVOST, FOSTER; FAWCETT, TOM. Data Science para Negócios: O que Você Precisa Saber Sobre Mineração de Dados e Pensamento Analítico de Dados. Rio de Janeiro: Alta Books, 2016.

17. RMSE: Raiz do Erro Quadrático Médio em Machine Learning. Disponível em:

<https://mariofilho.com/rmse-raiz-do-erro-quadratico-medio-em-machine-learning/> Acessado em 07 de novembro 2023

18. Rundo, F. et al. Machine Learning for Quantitative Finance Applications: A Survey. Applied Sciences, v. 9, n. 5574, p. 1–20, 2019.

19. You, Haifeng; Cao, Kai (2021) Fundamental Analysis via Machine Learning. The 4th Annual SQA-CFA Society NY Joint Conference Data Science in Finance: Learning from Machine Learning, Virtual; New York, USA, 12 - 13 January 2021.

20. Zonta d'Ávila, Mariana. Bolsa conquista 1,5 milhão de novos investidores em 2020, um aumento de 92% no ano. InfoMoney, 2021. Disponível em: <https://www.infomoney.com.br/onde-investir/bolsa-conquista-15-milhao-de-novos-investidores-em-2020-um-aumento-de-92-no-ano/>. Acesso em: 07 nov. 2023.

7 ANEXOS

7.1 CÓDIGO-FONTE

Prophet.py

```
import pandas as pd
from prophet import Prophet
import matplotlib.pyplot as plt
from statsmodels.tools.eval_measures import rmse
def roe_forecast():

    # Caminho do arquivo de dados da ação
    stock_path = 'Data/VALE3.csv'
    print('Time series forecast for ' + stock_path)

    # Carrega os dados do arquivo CSV
    df = pd.read_csv(stock_path)
    df.head()

    validation_size = 2

    # Dados para treinamento = DataFrame - tamanho da validação
    training_df = df.head(len(df) - validation_size)
    print("Training: ", training_df)

    validation_df = df.tail(validation_size)

    # Cria uma instância do modelo Prophet
    m = Prophet()

    # Ajusta o modelo aos dados
    m.fit(training_df)

    # Gera datas futuras para previsão
    future = m.make_future_dataframe(periods=2, freq='Q')
    print("DataFrame futuro \n", future.tail())

    # Realiza a previsão
    forecast = m.predict(future)
    print(forecast.tail())
```

```

# Exibe os resultados da previsão
print(forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper', 'trend',
'yearly']])

predictions = forecast.tail(len(validation_df))

print("Predictions: ", predictions)

print("Validation: ", validation_df)

# # rmse entre o conjunto de dados previstos e o conjunto completo
# print("RMSE completo: ", rmse(forecast['yhat'], df['y']))

print("Dados da predição para o calculo do RMSE: \n",
predictions['yhat'])

#rmse entre os dados de validação e e os dados previstos
print("RMSE predito: ", rmse(predictions['yhat'],validation_df['y']))

# Plota o gráfico da previsão
fig1 = m.plot(forecast)

# Plota componentes da previsão
fig2 = m.plot_components(forecast)

# Exibe os gráficos
plt.show()

```

ARIMA.ipynb

```

import itertools
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import statsmodels.api as sm
import warnings
from pylab import rcParams
from statsmodels.graphics.tsaplots import plot_pacf,plot_acf
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.stattools import adfuller

```



```
%matplotlib inline
rcParams['figure.figsize'] = 15, 10
warnings.filterwarnings('ignore')

file_path = '../Data/VALE3.csv'

df = pd.read_csv(file_path)

df.head()

validation_size = 2

training_df = df.head(len(df) - validation_size)
training_df.head()

validation_df = df.tail(validation_size)
validation_df.set_index(pd.to_datetime(validation_df.ds,
format="%Y-%m-%d"), drop=True, inplace=True)
validation_df.drop('ds', axis=1, inplace=True)
validation_df.head()

training_df.set_index(pd.to_datetime(training_df.ds, format="%Y-%m-%d"),
drop=True, inplace=True)
training_df.drop('ds', axis=1, inplace=True)

training_df.head()

training_df.shape

training_df = training_df.sort_index()

training_df.describe()

training_df['Ticks'] = range(0, len(training_df.index.values))

training_df.head(10)

training_df.tail(10)

plt.plot(training_df.y)
```

```

def stationarity_check(ts):
    # Média móvel
    roll_mean = ts.rolling(12).mean()
    plt.plot(ts, color='green', label='Original')
    plt.plot(roll_mean, color='blue', label='Média móvel')
    plt.legend(loc='best')
    plt.title('Média móvel')
    plt.show(block=False)
    # (ADF) Augmented Dickey-Fuller test:
    print('Augmented Dickey-Fuller test:')
    df_test = adfuller(ts)
    df_output = pd.Series(df_test[0:4], index=['Test
Statistic', 'p-value', '#Lags', 'Número de observações'])
    print("df_output: \n", df_output)
    for key, value in df_test[4].items():
        df_output['Critical Value (%s) %key] = value
    print(df_output)

stationarity_check(training_df.y)

from pmdarima import auto_arima

import warnings
warnings.filterwarnings("ignore")

stepwise_fit = auto_arima(training_df['y'], trace=True,
suppress_warnings=True)
stepwise_fit.summary()

model = sm.tsa.arima.ARIMA(training_df.y, order=(1,0,2))
results = model.fit()
plt.plot(training_df.y)
plt.plot(results.fittedvalues, color='red')
plt.title('Fitting data _ MSE: %.2f'%
(((results.fittedvalues-training_df.y)**2).mean()))
plt.show()

predictions = results.predict('2009-03-31', '2023-06-30')

predictions.tail()

```

```

predictions.plot(legend=True)
validation_df['y'].plot(legend=True)

from sklearn.metrics import mean_squared_error
from math import sqrt
print("Trimestres peditos para validação: \n",
predictions.tail(len(validation_df)))
print("Conjunto para validação: \n", validation_df)
rmse=sqrt(mean_squared_error(predictions.tail(len(validation_df)),
validation_df['y']))
print("RMSE: ", rmse)

```

pre-processamento.js

```

const AWS = require('aws-sdk');
const s3 = new AWS.S3();
const csv = require('csv-parser');

exports.handler = async (event) => {
  const bucketName = 'dadosfundamentals';
  const fileName = 'dados.csv';

  try {
    const s3Stream = s3.getObject({ Bucket: bucketName, Key: fileName
}).createReadStream();
    const results = await processCSV(s3Stream);

    // Processar resultados para calcular ROE
    const roeMap = calculateROE(results);

    console.log(roeMap);
    return roeMap;
  } catch (error) {
    console.error(error);
    throw error;
  }
};

```

```
function processCSV(s3Stream) {
  return new Promise((resolve, reject) => {
    const results = [];
    s3Stream.pipe(csv())
      .on('data', (data) => results.push(data))
      .on('end', () => resolve(results))
      .on('error', reject);
  });
}

function calculateROE(data) {
  const roeMap = {};
  for (let i = 0; i < data.length; i++) {
    if (i % 2 === 0 && data[i + 1]) { // Linha par e a próxima linha
      existe
      const sum = parseFloat(data[i].Coluna1) +
        parseFloat(data[i].Coluna2) + parseFloat(data[i].Coluna3) +
        parseFloat(data[i].Coluna4);
      const divisor = parseFloat(data[i + 1].Coluna4);
      const date = data[i].Data;

      if (divisor !== 0) {
        roeMap[date] = sum / divisor;
      }
    }
  }
  return roeMap;
}
```

8 ARTIGO TCC

Sistema de análise comparativa fundamentalista utilizando inteligência artificial

Fernando Silva Poerner¹ , Luiz Felipe Ribeiro Baroncello¹ , Elder Rizzon Santos¹

1 Departamento de Informática e Estatística – Universidade Federal de Santa Catarina
(UFSC)

Caixa Postal 476 – 88040-370 – Florianópolis – SC – Brazil

fernandospoerner@gmail.com, luizbaroncello@gmail.com, elder.santos@ufsc.br

Abstract. *This study presents a novel approach to stock market analysis by integrating fundamentalist comparative analysis with artificial intelligence. The focus is on developing a tool that automates the evaluation of company data such as liquidity, and operational results to form the fair market value. Utilizing data of the analyzed company, the tool employs advanced machine learning techniques to process and analyze time-series data. The goal is to accurately predict the fair market value of stocks based on fundamentalist principles. This project not only demonstrates the practical application of artificial intelligence in financial analysis but also provides insights into the evolving landscape of the stock market influenced by technological advancements. The system is tested and evaluated for accuracy and applicability, offering a practical and advanced perspective for investors in the stock market*

Resumo. *Este estudo apresenta uma abordagem inovadora para a análise do mercado de ações, integrando análise comparativa fundamentalista com inteligência artificial. O foco é desenvolver uma ferramenta que automatiza a avaliação de dados empresariais, como liquidez e resultados operacionais, para formar o preço justo. Utilizando dados da empresa analisada, a ferramenta emprega técnicas avançadas de aprendizado de máquina para processar e analisar dados em séries temporais. O objetivo é prever com precisão o preço justo de uma ação com base em princípios fundamentalistas. Este projeto demonstra a aplicação prática da inteligência artificial na análise financeira e oferece insights sobre a evolução do mercado de ações influenciada pelos avanços tecnológicos. O sistema é testado e avaliado quanto à precisão e aplicabilidade, oferecendo uma perspectiva prática e avançada para investidores no mercado de ações.*

1. Introdução

Em 2020, a bolsa de valores brasileira experimentou um notável aumento de 92%, atraindo 1,5 milhão de novos investidores Zonta d'Ávila, Mariana. Bolsa conquista 1,5 milhão de novos investidores em 2020, um aumento de 92% no ano. InfoMoney, 2021. Disponível em:

<https://www.infomoney.com.br/onde-investir/bolsa-conquista-15-milhao-de-novos-investidores-em-2020-um-aumento-de-92-no-ano/>. Acesso em: 07 nov. 2023. Esse crescimento significativo pode ser atribuído a diversos fatores, como a queda da taxa de juros e a volatilidade do mercado decorrente da pandemia. Diante desse cenário, torna-se crucial desenvolver ferramentas acessíveis e simplificadas para realizar análises técnicas e consolidadas.

A evolução tecnológica da Inteligência Artificial (IA) abre caminho para a criação de estruturas capazes de analisar padrões e comportamentos ativos ao longo do tempo, considerando seus valores fundamentalistas e variações. De acordo com Goldschmidt (2010), o aprendizado de máquina (AM) desempenha um papel essencial, permitindo que computadores aprendam com dados reais e tomem decisões com base em padrões identificados durante o treinamento.

Os algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs) mencionadas por Barreto (2002), simulam o funcionamento do cérebro humano, ajustando pesos durante o treinamento para classificar amostras desconhecidas. A convergência de princípios de Data Science e mineração de dados nos negócios destaca a importância dessas técnicas diante do volume e variedade de dados no mercado financeiro (Provost; Fawcett, 2016).

Com a abundância de dados financeiros, a automação torna-se essencial para reduzir a necessidade de esforços humanos na coleta e organização das informações (Ferrara, 2014). A complexidade da previsão do mercado de ações, devido a incertezas e diversos fatores, ressalta a importância de ferramentas de AM e Data Science. A combinação da análise fundamentalista e técnica, utilizando dados históricos, permite a criação de modelos preditivos para análises de longo prazo.

Após a coleta e análise de dados, busca-se desenvolver modelos de predição que identifiquem flutuações nos valores das ações a longo prazo. A validação desses modelos será realizada por meio de testes com premissas básicas, visando

otimizar e selecionar o modelo mais eficaz para aplicação no contexto do mercado financeiro através do cálculo do preço justo de Décio Bazin (1994).

2. Fundamentação teórica

A análise fundamentalista é uma ferramenta crucial no mercado financeiro, envolvendo a avaliação de indicadores econômicos, demonstrativos financeiros e cenário macroeconômico para identificar oportunidades de investimento. Ela oferece segurança ao investidor ao se basear em indicadores que validam a saúde financeira das empresas, utilizando abordagens quantitativas e qualitativas.

Diversos indicadores, como ROE, Margem Líquida, Resultado Bruto, Margem Bruta e Dividend Yield, são utilizados nesse processo, conforme definidos no glossário do InvestSite. O conceito de preço justo, segundo Décio Bazin (1994), relaciona-se à capacidade da empresa de fornecer dividendos consistentes ao longo do tempo, considerando diferentes abordagens na avaliação de ações.

Na previsão do mercado financeiro, modelos como séries temporais tradicionais, redes neurais, máquina de vetores de suporte, raciocínio baseado em casos tais modelos apresentados por Paul D. Yoo (2005), ARIMA e Prophet são aplicados. Esses modelos destacam-se pela capacidade de antecipar oscilações e padrões temporais complexos nos dados do mercado financeiro.

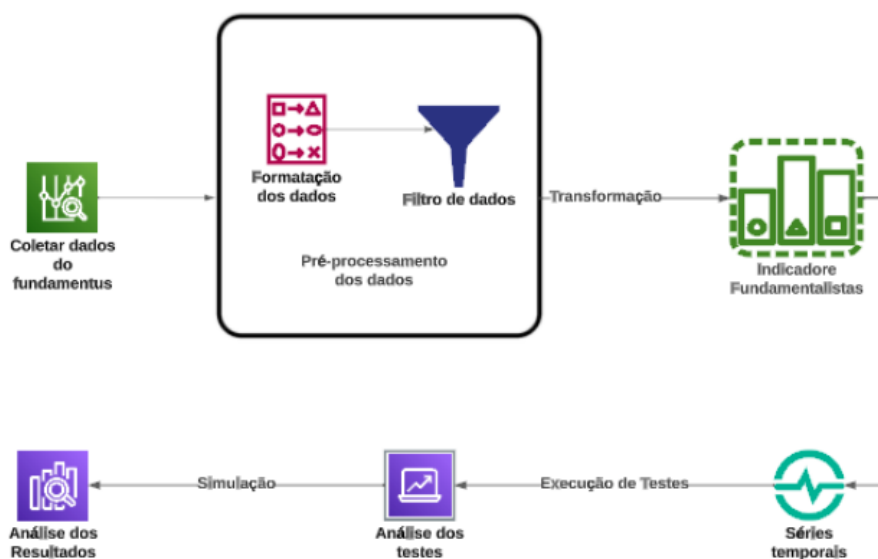
A ciência de dados, uma forma de estatística, utiliza elementos de matemática e ciências da computação para extrair percepções de dados quantitativos e qualitativos. Esse processo cíclico envolve etapas como conhecimento do tópico, aquisição e preparo de dados, exploração, modelagem, interpretação, implantação, monitoramento e repetição, garantindo a eficácia e funcionamento contínuo no longo prazo.

3. Desenvolvimento

O objetivo principal é descrever uma metodologia estruturada para determinar o preço justo de ações, abrangendo desde a coleta inicial dos dados até o cálculo detalhado do preço justo.

Para que este objetivo seja alcançado foi criado um modelo que utilizando séries temporais para identificar boas ações através de dados fundamentalistas coletados a partir do site fundamentus que pode ser acessado em:

<https://www.fundamentus.com.br/>. Acesso em: 09 dez. 2023. Estes dados passam por uma etapa de pré-processamento onde são transformados para um indicador fundamentalista, neste trabalho foi utilizado o ROE trimestral como entrada para o treinamento e validação do modelo proposto.



Fonte: elaborado pelos autores

3.1 Ferramentas

Para a transformação dos dados, foi utilizado Node.js. Este ambiente possibilita aos desenvolvedores a execução de código JavaScript no lado do servidor, conhecido como JavaScript no lado do servidor (server-side JavaScript).

Na etapa de pre-processamento de dados, AWS (Amazon Web Services), um serviço de computação em nuvem desenvolvido pela Amazon, foi utilizado Assim como os serviços S3 para armazenamento de arquivos e o serviço Lambda, que permite a execução de códigos sem a necessidade de configurar servidores.

Para o treinamento e teste do modelo, Python foi utilizado. Reconhecido por sua versatilidade e alto nível, o Python é amplamente adotado em diversos campos devido à sua simplicidade e legibilidade. Sua vasta biblioteca padrão e uma comunidade ativa de desenvolvedores fazem dele uma escolha ideal para análise de

dados e aprendizado de máquina, oferecendo diversas opções para a criação de séries temporais e outras análises de maneira facilitada.

3.2 Coleta de dados

Para coleta de dados históricos de ações brasileiras, foi usado o site Fundamentus, que disponibiliza para download os dados históricos das ações em forma de planilha e divididos em trimestres, com um período de mais de 10 anos.

Para este trabalho, serão usados dados da empresa VALE3. As informações obtidas se dividem em duas partes, sendo a primeira referente ao balanço patrimonial da empresa ao longo do tempo e a segunda referente a demonstração de resultados da mesma, também ao longo do tempo.

Abaixo é possível observar alguns dos dados principais do balanço patrimonial da empresa, na tabela obtida cada categoria de dados é subdividida em várias subcategorias, explicando cada parte da construção patrimonial da empresa, para o trabalho, como o foco será no cálculo do ROE, a principal informação utilizada do balanço patrimonial será o valor de Patrimônio Líquido ao longo do tempo, dado que para a empresa VALE3, é registrado de 2008 até a atualidade.

Tabela 1 – Principais dados do balanço patrimonial da Vale

VALE3	30/09/2008	31/12/2008	31/03/2009
Ativo Total	R\$ 174.240.432,00	R\$ 185.779.470,00	R\$ 187.954.266,00
Passivo Total	R\$ 174.240.432,00	R\$ 185.779.470,00	R\$ 187.954.266,00
Patrimônio Líquido	R\$ 95.551.521,00	R\$ 96.274.637,00	R\$ 98.611.577,00

Fonte: elaborado pelos autores com dados obtidos em <https://www.fundamentus.com.br/detalhes.php?papel=VALE3> Acessado em: 07 nov.2023

A segunda parte, referente à demonstração de resultados da empresa, contém diversos dados referentes aos resultados obtidos pela empresa ao longo dos anos, abaixo é possível ver os principais resultados. Dentre eles, o utilizado para calcular o ROE será o Lucro/Prejuízo do período.

Tabela 2 – Principais dados da demonstração de resultados da Vale

VALE3	30/09/2008	31/12/2008	31/03/2009
Receita Líquida de Vendas e/ou Serviços	R\$ 20.697.999,00	R\$ 17.382.957,00	R\$ 12.915.840,00
Resultado Bruto	R\$ 12.056.218,00	R\$ 9.272.080,00	R\$ 6.049.073,00
Resultado Antes Tributação/Participações	R\$ 12.459.581,00	-R\$ 408.183,00	R\$ 3.801.033,00
Lucro/Prejuízo do Período	R\$ 12.433.386,00	R\$ 2.020.360,00	R\$ 3.150.903,00

Fonte: elaborado pelos autores com dados obtidos em <https://www.fundamentus.com.br/detalhes.php?papel=VALE3>

3.3 Pré-processamento dos dados

A etapa de pré-processamento de dados é onde os dados que são coletados na seção anterior são filtrados e transformados, dando origem ao ROE ao longo do tempo. Para isso os dados brutos são colocados no S3 e um script é executado no lambda, de forma a calcular o ROE e dar origem a um output que será usado nas seções seguintes. Abaixo é possível observar o cálculo do ROE e um exemplo do mesmo para o quarto trimestre de 2022.

$$ROE = \frac{\text{Lucro Líquido (ano)}}{\text{Patrimônio Líquido}} \quad (1)$$

Tabela 3 – Cálculo do ROE para o quarto trimestre de 2022

	31/03/2022	30/06/2022	30/09/2022	31/12/2022
Lucro/Prejuízo do Período	R\$ 23.046.001,00	R\$ 30.033.000,00	R\$ 23.286.000,00	R\$ 19.559.002,00
Patrimônio Líquido				R\$ 187.112.004,00
ROE				51,27

Fonte: elaborado pelos autores

Ao calcular o ROE, a função escreve em um output os resultados de forma que eles possam ser usados na próxima etapa. Relacionando o trimestre e o ROE encontrado para o mesmo.

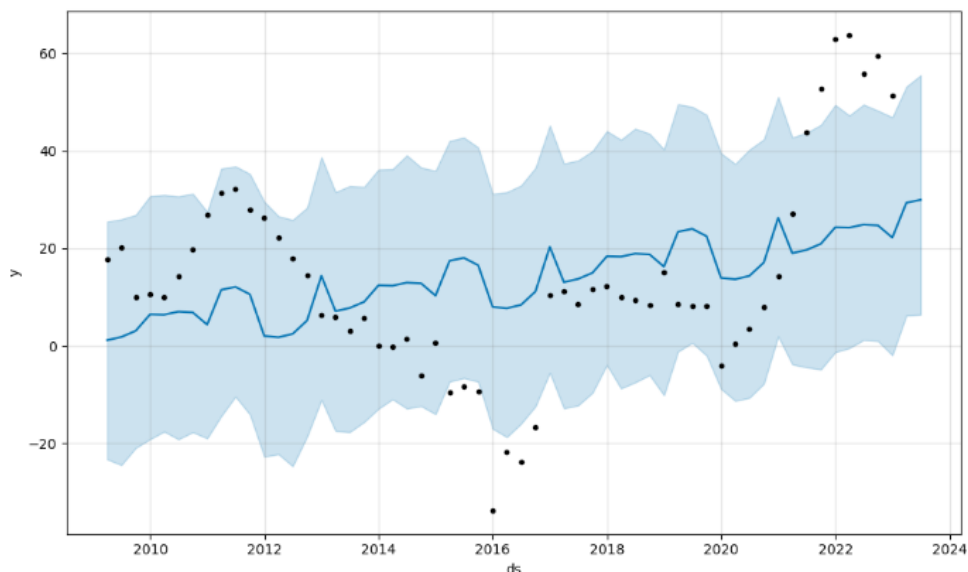
3.4 Séries temporais

Para prever os futuros valores dos Retornos sobre o Patrimônio Líquido (ROEs), utilizamos técnicas de análise de séries temporais com o suporte de ferramentas de Machine Learning, como a biblioteca Python Prophet do Facebook, conhecida por sua capacidade de lidar com sazonalidade, feriados e tendências. Além disso, empregamos o método ARIMA com o auxílio do ambiente de desenvolvimento interativo Jupyter Lab. As séries temporais, que representam dados coletados ao longo do tempo, foram importadas e organizadas a partir de dados trimestrais de ROEs previamente tratados e armazenados em arquivos CSV.

3.4.1 Prophet

Para o modelo do Prophet utilizamos um conjunto de dados de ROEs trimestrais da VALE3 com 58 trimestres de 2009 à 2023 sendo 56 trimestres utilizados para treinamento do modelo e 2 trimestres para validação, esse tamanho reduzido de dados para validação se dá pela pequena quantidade de dados trimestrais disponíveis para o escopo de trabalho nesse caso o modelo se comporta melhor quando utilizamos mais dados para treinamento do que a proporção comumente utilizada de 70/30.

Ao realizar a função "predict()" do Prophet temos um conjunto de dados preditos a partir do conjunto de dados ajustados ao modelo com o conjunto de datas futuras configurados através da função "make_future_dataframe()" então quando apresentamos este conjunto de dados em um gráfico temos o seguinte resultado:



Fonte: elaborado pelos autores

O gráfico mostra a evolução do Retorno sobre o Patrimônio Líquido (ROE) ao longo dos trimestres. Os pontos pretos representam os valores reais do conjunto de treinamento, enquanto a linha azul indica as previsões do modelo. Uma zona em azul claro destaca o intervalo de confiança, revelando a incerteza nas previsões e a variabilidade possível nos resultados.

O Prophet encontra as previsões para o conjunto de dados seguindo algumas variáveis, as principais delas são tendência e sazonalidade. O modelo é formulado como:

$$y(t)=g(t)+s(t)+h(t)+\epsilon t$$

onde:

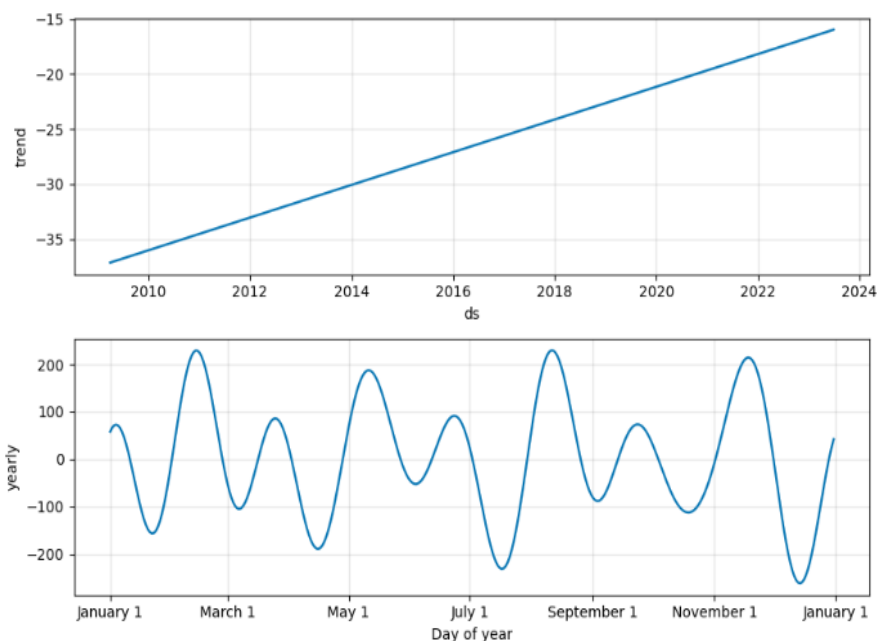
$g(t)$ representa a tendência.

$s(t)$ representa a sazonalidade.

$h(t)$ representa outros efeitos adicionais (feriados, por exemplo).

ϵt é o termo de erro.

Para o Modelo proposto do Prophet temos as seguintes variáveis representadas através de gráficos tendência e sazonalidade respectivamente:



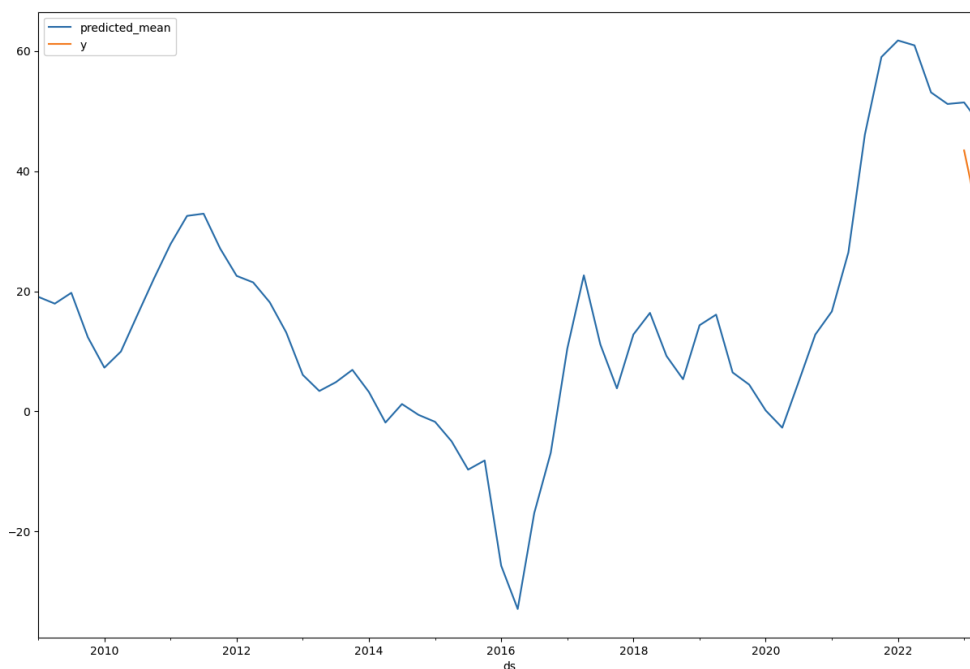
Fonte: elaborado pelos autores

3.4.2 ARIMA

O ARIMA aborda componentes como tendência e sazonalidade para capturar padrões temporais nos dados. Ao contrário do Prophet, o ARIMA requer a definição de parâmetros na função $ARIMA(p, d, q)$, onde (p) é a ordem do componente autorregressivo (número de lags anteriores), (d) é a ordem de diferenciação (número de vezes que os dados são diferenciados para tornar a série estacionária), e (q) é a ordem do componente de médias móveis (número de lags anteriores do termo de erro). A estacionariedade da série foi verificada usando a função `stationary_check`, crucial para definir o parâmetro (d) .

A mesma proporção de dados utilizada para o modelo Prophet foi utilizada no ARIMA com 56 trimestres de treinamento e 2 de validação.

Com isso temos a predição de dados do ARIMA através do conjunto de dados formado a partir do resultado dos dados ajustados pelo modelo de treinamento utilizando função "predict()" que recebe o período desejado de predição, e para visualização destes dados temos o gráfico a seguir:



Fonte: elaborado pelos autores

Em azul os valores preditos pelo modelo em conjunto com os valores de validação em laranja, com os últimos 2 trimestres.

3.4.3 ARIMA VS Prophet

Para validação e comparação dos modelos foi utilizado o RMSE, sigla do inglês root mean-square error que traduzido é raiz do erro quadrático médio, para medir o desempenho de modelos de regressão. A partir dos dados de predição foram realizadas a validação do RMSE que possui a seguinte fórmula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}$$

Onde:

- n é o número de amostras
- y_i é o valor observado para a amostra i
- p_i é o valor previsto pelo modelo para a amostra i

Informações obtidas em ([RMSE \(Raiz Do Erro Quadrático Médio\) Em Machine Learning | Mario Filho](#)) Acessado em: 07 de nov. 2023.

RMSE para o Prophet:

Data (trimestre)	ROE previsto	ROE de validação
2023-03-31	29.361.914	43.44
2023-06-30	29.965.645	30.45
Total RMSE	9.9606000839 33760	

Fonte: elaborado pelos autores

RMSE para o ARIMA:

Data (trimestre)	ROE previsto	ROE de validação
2023-03-31	51.427026	43.44
2023-06-30	47.746058	30.45
Total RMSE	13.47119554829150	

Fonte: elaborado pelos autores

Sendo assim temos que a escolha entre Prophet e ARIMA depende da natureza dos dados e da complexidade dos padrões temporais. O Prophet destaca-se em cenários flexíveis e automatizados, demonstrando desempenho superior em conjuntos de dados não estacionários. Por outro lado, o ARIMA é mais adequado para contextos teóricos e séries estacionárias, sofrendo desvantagens em conjuntos não estacionários. Ambos oferecem abordagens únicas para a modelagem de séries temporais, sendo a escolha guiada pela aplicação específica e pelas características dos dados. No cenário abordado, envolvendo séries não estacionárias, tendências e sazonalidades, o Prophet

mostrou-se superior em suas previsões.

3.5 Preço justo

No presente projeto, propõe-se uma modificação do cálculo de preço justo de Décio Bazin, onde se a série temporal do ROE indica uma tendência de crescimento no próximo trimestre, sugere-se acrescentar 10% ao valor do preço justo calculado pela fórmula. Por outro lado, se a série temporal indica uma tendência de queda no ROE, recomenda-se reduzir o preço justo em 10%. Estes ajustes, refletem uma expectativa de que o preço possa acompanhar o fato de a empresa estar apresentando um ROE futuro maior ou menor. Abaixo é possível observar a fórmula proposta:

$$\text{Valor justo: Valor Justo (Décio Bazin)} + / - 10\% \quad (3)$$

3.6 Análise de Resultados

É crucial comparar o desempenho real da ação com as previsões do modelo proposto. A validação será feita em junho de 2023 usando o modelo Prophet, considerado mais preciso nos conjuntos de dados anteriores. A análise comparativa dos valores previstos para a ação VALE3 com os valores reais será realizada, utilizando o cálculo do preço justo para determinar o limite ideal de preço para investimentos de longo prazo.

Data	ROE previsto	Delta ROE do próximo trimestre e atual	Percentual a considerar na fórmula do preço justo	Preço real da Ação	Preço justo calculado	Dividendos (12 meses anteriores)
2023-06-30	28,157302	-1,531239	-10%	64,22	37,25	5,69171539
2023-09-30	26,626063	67,58

Fonte: elaborado pelos autores

O preço da ação VALE3 em 30 de junho de 2023 é de 64,22. Segundo o cálculo do preço justo, que é inferior a esse valor, o modelo indica que a ação não está em uma faixa de preço ideal para investimentos de longo prazo. A faixa considerada ideal para uma rentabilidade esperada de 13,75% ao ano seria de

valores iguais ou inferiores a 37,25 reais por ação de VALE3, alinhando-se a investimentos em títulos de renda fixa no período considerado.

4. Considerações finais

O presente trabalho teve como objetivo analisar o estado da arte em extração e transformação de dados, desenvolver um modelo de predição para identificar flutuações nos valores de ações de empresas selecionadas, treinar o modelo com base em dados históricos de análise fundamentalista, definir e realizar um experimento para avaliar a capacidade de análise do modelo gerado e comparar os resultados com a valorização real do mercado, isso foi feito com base em boas práticas de investimento a longo prazo e para buscar valorização e retorno sobre o investimento analisado. Desta forma, os objetivos gerais e específicos propostos no início do projeto foram atingidos. Apesar do preço justo não satisfazer o preço da ação e da mesma ter apresentado uma certa valorização durante o período analisado.

Durante a elaboração deste trabalho, alguns desafios surgiram. A ausência de um conjunto de dados contendo as informações essenciais para a aplicação do modelo demandou a criação do mesmo, utilizando a extração manual de dados de sites relacionados a indicadores fundamentalistas, de forma que foi extraído dados abrangendo um período considerável de em média 14 anos, sendo estes de significativa importância para viabilizar uma análise do modelo proposto e os cálculos necessários para atingir o indicador fundamentalista escolhido, o ROE. Porém, uma análise mais precisa exigiria mais dados trimestrais, os quais não estão disponíveis diretamente em fontes de dados públicas.

Apesar do trabalho ter atingido todos os objetivos propostos, ainda há muito espaço para melhoria e evolução do modelo gerado, abaixo estão listados alguns próximos trabalhos possíveis: Automatização dos passos gerados no modelo, melhoria do cálculo do preço justo, disponibilização dos resultados em um site ou aplicativo e análise de mais ações e outros setores.

Referências

1. **BAZIN, Décio. Faça Fortuna com Ações, Antes que Seja Tarde. 2ª edição, 1994, JMJ, São Paulo-SP.**

2. **BARRETO, J. M. Introdução às Redes Neurais Artificiais. V Escola Regional de Informática. Sociedade Brasileira de Computação, Regional Sul, Santa Maria, Florianópolis, Maringá, p. 5–10, 2002.**

3. **Fundamentus. Disponível em: <https://www.fundamentus.com.br/>. Acesso em: 07 nov. 2023.**

4. **GOLDSCHMIDT, Ronaldo Ribeiro. Uma Introdução à Inteligência Computacional: Fundamentos, Ferramentas e Aplicações. 1. ed. Rio de Janeiro: Instituto Superior de Tecnologia do Rio de Janeiro, 2010.**

5. **P. D. Yoo, M. H. Kim and T. Jan (2005) "Financial Forecasting: Advanced Machine Learning Techniques in Stock Market Analysis," 2005 Pakistan Section Multitopic Conference, Karachi, Pakistan, pp. 1-7, doi: 10.1109/INMIC.2005.334420.**

6. **PROVOST, FOSTER; FAWCETT, TOM. Data Science para Negócios: O que Você Precisa Saber Sobre Mineração de Dados e Pensamento Analítico de Dados. Rio de Janeiro: Alta Books, 2016.**

7. **Zonta d'Ávila, Mariana. Bolsa conquista 1,5 milhão de novos investidores em 2020, um aumento de 92% no ano. InfoMoney, 2021. Disponível em: <https://www.infomoney.com.br/onde-investir/bolsa-conquista-15-milhao-de-novos-investidores-em-2020-um-aumento-de-92-no-ano/>. Acesso em: 07 nov. 2023.**