

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE GESTÃO E NEGÓCIOS

GABRIEL ALEXANDRE GONÇALVES

**COMPARAÇÃO DE MODELOS DE MACHINE LEARNING PARA PREVISÃO DE  
PREÇO DE FECHAMENTO DE UMA AÇÃO DO SETOR BANCÁRIO LISTADA NA  
B3**

UBERLÂNDIA

2022

GABRIEL ALEXANDRE GONÇALVES

**COMPARAÇÃO DE MODELOS DE MACHINE LEARNING PARA PREVISÃO DE  
PREÇO DE FECHAMENTO DE UMA AÇÃO DO SETOR BANCÁRIO LISTADA NA  
B3**

Trabalho de Conclusão de Curso apresentado à Faculdade de Gestão e Negócios (FAGEN) da Universidade Federal de Uberlândia (UFU) para obtenção do Grau de Bacharel no curso de Graduação em Gestão da Informação da Universidade Federal de Uberlândia.

Orientação: Prof. Dr. José Eduardo Ferreira Lopes

UBERLÂNDIA

2022

GABRIEL ALEXANDRE GONÇALVES

**COMPARAÇÃO DE MODELOS DE MACHINE LEARNING PARA PREVISÃO DE  
PREÇO DE FECHAMENTO DE UMA AÇÃO DO SETOR BANCÁRIO LISTADA NA  
B3**

Trabalho de Conclusão de Curso aprovado para obtenção do título de Bacharel no Curso de Graduação em Gestão da Informação da Universidade Federal de Uberlândia (UFU) pela banca examinadora formada pelos seguintes professores:

---

Prof. Dr. José Eduardo Ferreira Lopes  
Professor Orientador

---

Prof. Dr. Peterson Elizandro Gandolfi  
Membro

---

Prof. Dr. Antônio Sérgio Torres Penedo  
Membro

UBERLÂNDIA

2022

## RESUMO

Este estudo aborda investimentos e aprendizado de máquina, realizando uma comparação entre os algoritmos *Árvore de Decisão*, *Random Forest* e *Regressão Linear*. O Objetivo deste trabalho é prever o preço do ativo ITUB4, e comparar as previsões dos três algoritmos utilizados. Para esse fim, foram coletadas informações financeiras do ativo, e feita a engenharia de dados para criar os insumos finais a serem utilizados pelos algoritmos de previsão. Os resultados mostram que todos os modelos conseguiram ter boas previsões para o período em que foi proposto.

Palavras-chave: Investimento; Previsão; Aprendizado de Máquina

## **ABSTRACT**

This study addresses investments and machine learning, making a comparison between the Decision Tree, Random Forest and Linear Regression algorithms. The objective of this work is to predict the price of the ITUB4 asset and compare the predictions of the three algorithms used. To this end, the financial information of the asset was collected, and the data engineering was done to create the final inputs to be used by the forecasting algorithms. The results show that all models were able to have good predictions for the period in which it was proposed.

Keywords: Investment; Forecast; Machine Learning

## LISTA DE ABREVIATURAS E DE SIGLAS

AD	Árvore de Decisão
RF	Random Forest
RL	Regressão Linear
B3	Bolsa de Valores do Brasil

# Sumário

<b>1. Introdução</b> .....	1
<b>2. Revisão Bibliográfica</b> .....	1
<b>2.1 Investimento</b> .....	1
<b>2.2 Mercado de Capitais</b> .....	2
<b>2.3 Análise Fundamentalista</b> .....	2
<b>2.4 Métodos Estatísticos e Aprendizagem de Máquina</b> .....	2
<b>3. Metodologia</b> .....	4
<b>3.1 Etapas da Metodologia</b> .....	4
<b>3.2 Base de Dados</b> .....	5
<b>3.3 Extração dos Atributos</b> .....	6
<b>3.3.1 P/VP</b> .....	6
<b>3.3.2 ROE</b> .....	6
<b>3.3.3 LPA</b> .....	7
<b>3.3.4 P/L</b> .....	7
<b>3.2 Método para Avaliação das Predições</b> .....	7
<b>3.2.1 Coeficiente de Determinação - <math>R^2</math></b> .....	7
<b>3.2.2 Erro Médio Absoluto – MAE</b> .....	8
<b>3.3.3 Erro Percentual Absoluto Médio – MAPE</b> .....	8
<b>3.3.4 Erro Quadrático Médio – MSE</b> .....	8
<b>3.3.5 Raiz do Erro Quadrático Médio – RMSE</b> .....	9
<b>3.4 Ferramentas Utilizadas</b> .....	9
<b>3.4.1 Python</b> .....	9
<b>3.4.2 Pandas</b> .....	9
<b>3.4.3 NumPy</b> .....	10
<b>3.4.4 Scikit Learning</b> .....	10
<b>3.4.5 Matplotlib</b> .....	10
<b>3.4.6 Google Colab</b> .....	10

<b>4. Resultados e discussões .....</b>	<b>10</b>
<b>5. Considerações finais .....</b>	<b>12</b>
<b>6. Referências Bibliográficas .....</b>	<b>13</b>



## **1. Introdução**

O mercado de capitais é um ambiente onde são negociados ativos de empresas e do estado. No Brasil esse ambiente é representado pela B3. Algumas empresas utilizam este mercado para atrair investimentos, bem como algumas pessoas utilizam como uma forma de investimento em aposentadoria, visando um investimento a longo prazo e independência financeira. E neste mesmo mercado existem pessoas que tentam prever as cotações, buscando lucro a curto e médio prazo.

Mas a previsão dos preços é uma tarefa complicada, pois o mercado é volátil, sofrendo influência de diversos fatores, não só das próprias empresas, mas de eventos políticos, ambientais e do mercado global. E para tentar solucionar esse problema existem algumas ferramentas para tentar prever essas cotações, através da utilização de técnicas de aprendizado de máquina.

O problema a ser tratado é: Como posso prever o preço de uma ação utilizando dados diários de negociação e indicadores de mercado?

Portanto, esse artigo, tem como objetivo desenvolver e testar modelos para prever o preço de ações do setor bancário, em específico o Banco Itaú (ITUB4), utilizando dados históricos de preços diários, e indicadores de mercado, como P/VP (Preço sobre valor patrimonial), LPA (Lucro por ação), P/L (Preço por lucro) e ROE (Retorno sobre patrimônio). Para tanto, são apresentados os conceitos básicos relativos a investimentos e *machine learning*.

Este artigo está composto por cinco capítulos, sendo o primeiro deles a introdução, seguidamente do referencial teórico. No terceiro capítulo é apresentado a metodologia utilizada para a elaboração das informações apresentadas no capítulo quatro, em que é realizada a análise dos resultados. E por fim, o capítulo cinco apresenta a conclusão dos resultados.

## **2. Revisão Bibliográfica**

### **2.1 Investimento**

Segundo Marques (2014), investimento significa acumulação de projetos diretos e indiretos de produção, que estimulam o desenvolvimento econômico, gerando emprego, salto de produtividade e com isso melhorando os índices sociais.

Ainda segundo Giudicce e Estender (2017), investimento está relacionado à compra de ativos financeiros, como ativos de renda fixa e variável, com o objetivo principal de recuperar o valor monetário que foi desvalorizado pela inflação.

## **2.2 Mercado de Capitais**

Segundo Toro (2021), o mercado de capitais é uma parte do sistema financeiro que intermedia negociações entre quem busca a captação de recursos e quem procura investir, por meio da negociação de ativos entre empresas e investidores.

A movimentação do mercado de capitais é viabilizada pelas bolsas de valores que no Brasil se chama B3, foi criada após a união das antigas Bovespa (Bolsa de Valores de São Paulo), BM&F (Bolsa de Mercadorias e Futuros de São Paulo) e Cetip (Central de Custódia e de Liquidação Financeira de Títulos), e hoje é uma das maiores bolsas de valores do mundo (NUBANK, 2021).

## **2.3 Análise Fundamentalista**

Uma das formas de se analisar a possibilidade de variação de preços dos ativos no mercado de capitais é por meio da análise fundamentalista e, segundo a definição feita pelo IG (2022, tradução nossa), a análise fundamentalista é um método de avaliação do valor intrínseco de um ativo e análise dos fatores que podem influenciar seu preço no futuro. Essa forma de análise baseia-se em eventos e influências externas, bem como demonstrações financeiras e tendências do setor.

Essa análise tem uma abordagem extremamente abrangente que requer um profundo conhecimento de contabilidade, finanças e economia. Por exemplo, a análise fundamentalista requer a capacidade de ler demonstrações financeiras, a compreensão dos fatores macroeconômicos e o conhecimento das técnicas de avaliação. Ele se baseia principalmente em dados públicos, como os ganhos históricos de uma empresa e margens de lucro, para projetar crescimento futuro (CFI, 2022, tradução nossa).

## **2.4 Métodos Estatísticos e Aprendizagem de Máquina**

De uma forma geral, ao longo dos tempos, busca-se mecanismos que sejam capazes de prever os preços dos ativos em períodos futuros com a menor margem de erro possível. Dentre estes mecanismos, destacam-se as técnicas estatísticas e de inteligência artificial ou aprendizagem de máquina (*machine learning*), sendo que alguns destes métodos, os utilizados nesta pesquisa, são definidos a seguir.

Segundo Finkler (2017), *Machine Learning* é definida de forma geral como uma área que abrange o desenvolvimento, análise e aplicação de procedimentos que detectam de forma automática padrões em agrupamentos de dados.

Brown (2021, tradução nossa), diz que o aprendizado de máquina está por trás de *chatbots* e texto preditivo, aplicativos de tradução de idiomas, os programas que a Netflix sugere para você e como seus feeds de mídia social são apresentados. Os princípios de *Machine Learning* alimentam veículos e máquinas autônomas que podem diagnosticar condições médicas com base em imagens.

E para complementar, o site IBM (2020, tradução nossa), diz que o aprendizado de máquina é um componente importante do crescente campo da ciência de dados. Através do uso de métodos estatísticos, os algoritmos são treinados para fazer classificações ou previsões, e para descobrir *insights* chave em projetos de mineração de dados. Esses *insights* posteriormente impulsionam a tomada de decisões dentro de aplicativos e empresas, impactando idealmente as principais métricas de crescimento.

Segundo Lemos e outros (2005), “a árvore de Decisão é um método adequado quando o objetivo da Mineração de Dados é a classificação de dados ou predição de saídas.”

Uma árvore tem muitas analogias na vida real e acaba influenciando uma ampla área de aprendizado de máquina, abrangendo tanto a classificação quanto a regressão. Embora seja uma ferramenta comumente usada na mineração de dados para derivar uma estratégia para alcançar um objetivo específico, também é amplamente utilizada no aprendizado de máquina (GUPTA, 2017, tradução nossa).

A floresta aleatória ou *Random Forest*, como o próprio nome indica, consiste em um grande número de árvores de decisão individuais que operam como um conjunto. Cada árvore individual na floresta aleatória devolve uma previsão de classe e a classe com mais votos se torna a previsão do nosso modelo (YIU, 2019, tradução nossa).

A principal diferença entre o algoritmo de árvore de decisão e o algoritmo de floresta aleatória é que o estabelecimento de nós raiz e a segregação de nós é feito aleatoriamente no

último. A floresta aleatória emprega o método de ensacamento para gerar a previsão necessária (MBAABU, 2020, tradução nossa).

Além disso, Khaidem, Saha e Dey (2016, tradução nossa), afirmam que um leve ruído nos dados pode fazer com que a árvore cresça de uma maneira completamente diferente. Isso se deve ao fato de que as árvores de decisão têm tendência muito baixa e variância alta. A *Random Forest* supera esse problema treinando várias árvores de decisão em diferentes subespaços do espaço de recursos ao custo de um viés ligeiramente maior.

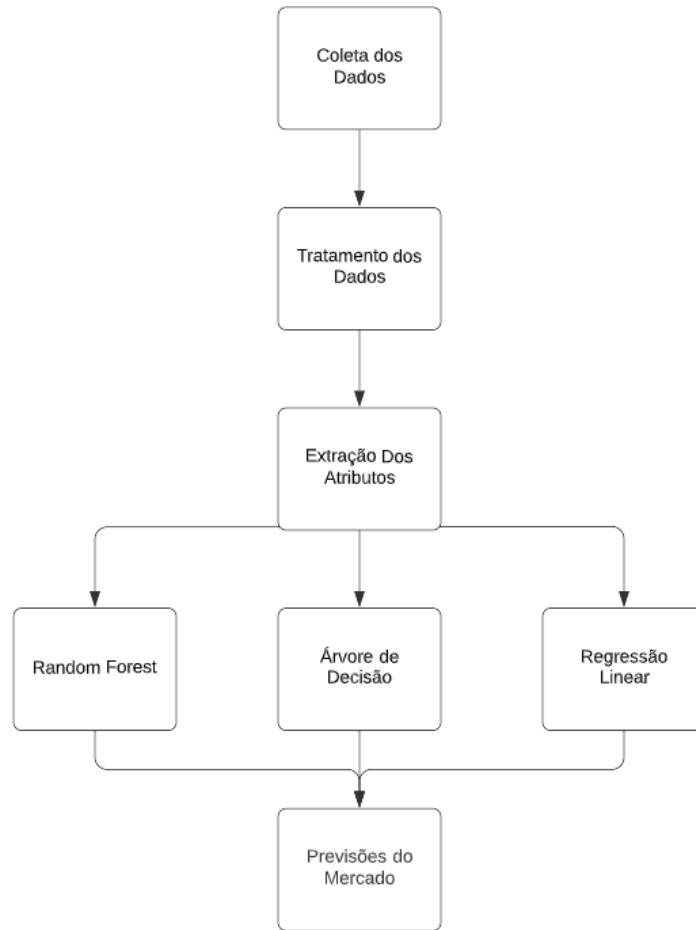
Regressão Linear é um algoritmo de aprendizado de máquina baseado no aprendizado supervisionado. Ele realiza uma tarefa de regressão. A regressão modela um valor de previsão de destino com base em variáveis independentes (GEEKSFORGEEKS, 2022, tradução nossa).

Esta forma de análise estima os coeficientes da equação linear, envolvendo uma ou mais variáveis independentes que melhor predizem o valor da variável dependente. A regressão linear se encaixa em uma linha reta ou superfície que minimiza as discrepâncias entre valores de saída previstos e reais (IBM, 2022, tradução nossa).

### **3. Metodologia**

#### **3.1 Etapas da Metodologia**

Os dados do ativo ITUB4 foram coletados, tratados e disponibilizados em um *Data Frame* final para ser utilizado nos modelos. Os detalhes de cada etapa serão apresentados na Figura 1.



**Figura 1** - Metodologia do Artigo  
**Fonte:** Elaboração Própria

### 3.2 Base de Dados

Nesta pesquisa foram utilizados os dados históricos diários e do DFP (Demonstrações financeiras padronizadas) do Itaú, código da ação na B3 ITUB4, para um período de 10 anos e meio (03/01/2011 a 05/07/2022).

Os dados utilizados nesta pesquisa foram coletados do site da CVM (Comissão de Valores Mobiliários) e do site Yahoo Finance.

Tabela 1 – Base de Dados Final do Ativo ITUB4

	Open	High	Low	Close	Adj Close	Volume	Qtd Ações	Lucro Líquido	PL	VP	P/VP	P/VP Adj	ROE	LPA	P/L
2011-01-03	18.031555	18.250118	17.963253	18.067982	11.323718	10014663.0	2281650.0	14610000.0	75336000.0	33.018211	0.547213	0.342954	19.393119	6.403261	2.821684
2011-01-04	18.077089	18.318420	18.031555	18.213692	11.415043	12685182.0	2281650.0	14610000.0	75336000.0	33.018211	0.551626	0.345720	19.393119	6.403261	2.844440
2011-01-05	18.122623	18.509665	18.004234	18.486897	11.586257	12143831.0	2281650.0	14610000.0	75336000.0	33.018211	0.559900	0.350905	19.393119	6.403261	2.887107
2011-01-06	18.400381	18.436810	18.031555	18.049768	11.312298	17066720.0	2281650.0	14610000.0	75336000.0	33.018211	0.546661	0.342608	19.393119	6.403261	2.818840
2011-01-07	18.013342	18.149944	17.512465	17.608086	11.035485	15259728.0	2281650.0	14610000.0	75336000.0	33.018211	0.533284	0.334224	19.393119	6.403261	2.749862
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2022-06-29	23.469999	23.680000	23.000000	23.080000	23.062031	30692300.0	4845845.0	28384000.0	164476000.0	33.941655	0.679990	0.679461	17.257229	5.857389	3.940322
2022-06-30	22.820000	22.879999	22.530001	22.670000	22.652349	36506900.0	4845845.0	28384000.0	164476000.0	33.941655	0.667911	0.667391	17.257229	5.857389	3.870325
2022-07-01	22.320000	22.969999	22.240000	22.850000	22.850000	34834600.0	4845845.0	28384000.0	164476000.0	33.941655	0.673214	0.673214	17.257229	5.857389	3.901056
2022-07-04	22.670000	22.799999	22.510000	22.629999	22.629999	8754100.0	4845845.0	28384000.0	164476000.0	33.941655	0.666732	0.666732	17.257229	5.857389	3.863496
2022-07-05	22.500000	22.790001	22.209999	22.790001	22.790001	28623500.0	4845845.0	28384000.0	164476000.0	33.941655	0.671446	0.671446	17.257229	5.857389	3.890812

2865 rows x 16 columns

Fonte: Dados da pesquisa

Primeiramente estes dados foram coletados, após isso foram realizados diversos tratamentos e cálculos para disponibilizar a base final, como mostra a Tabela 1.

### 3.3 Extração dos Atributos

#### 3.3.1 P/VP

O Preço sobre Valor Patrimonial (P/VP) indica se o preço da ação está barato ou caro. Se o valor for menor que 1, pode ser que a empresa esteja valendo menos que seu patrimônio líquido, sendo essa uma ótima chance de valorização para o futuro (STATUS INVEST, 2020).

A fórmula para calcular o P/VP é:

$$P/VP = \text{Preço de Mercado} / \text{Valor Patrimonial}$$

#### 3.3.2 ROE

O Retorno sobre o Patrimônio Líquido (ROE) é uma relação que fornece aos investidores uma visão de quão eficientemente uma empresa (ou mais especificamente, sua equipe de gestão) está lidando com o dinheiro que os acionistas contribuíram para isso. Ou seja, o retorno sobre o patrimônio líquido mede a rentabilidade de uma empresa em relação ao patrimônio líquido. Quanto maior o ROE, mais eficiente é a gestão de uma empresa na geração de renda e crescimento a partir de seu financiamento patrimonial (FURHMANN; JAMES; LI, 2022, tradução nossa).

A fórmula para calcular o ROE é:

$$\text{ROE} = \text{Lucro Líquido} / \text{Patrimônio Líquido}$$

### **3.3.3 LPA**

O Lucro por Ação (LPA) é calculado como lucro de uma empresa dividido pelas ações em circulação de suas ações ordinárias. O número resultante serve como um indicador da rentabilidade de uma empresa. É comum uma empresa relatar LPA que é ajustado para itens extraordinários e potencial diluição de ações. Quanto maior o LPA de uma empresa, mais rentável é considerado (FERNANDO; KINDNESS; KAZEL, 2022, tradução nossa)

A fórmula para calcular o LPA é:

$$\text{LPA} = \text{Lucro Líquido} / \text{Quantidade de ações}$$

### **3.3.4 P/L**

A relação preço-lucro é a razão para a valorização de uma empresa que mede seu preço atual das ações em relação ao seu lucro por ação (LPA). A relação preço-lucro também é às vezes conhecida como o múltiplo de preço ou o múltiplo de ganhos. As relações P/L são usadas por investidores e analistas para determinar o valor relativo das ações de uma empresa em uma comparação a outra, e pode ser usado para comparar uma empresa com seu próprio registro histórico ou para comparar mercados agregados uns com os outros ou ao longo do tempo (FERNANDO; BROCK; LI, 2022, tradução nossa)

A fórmula para calcular o P/L é:

$$\text{P/L} = \text{Preço da Ação} / \text{Lucro por Ação}$$

## **3.2 Método para Avaliação das Predições**

### **3.2.1 Coeficiente de Determinação - R<sup>2</sup>**

O  $R^2$  é uma medida que representa a variação dos dados explicados pelo modelo. Estes resultados variam entre 0 e 1, e são demonstrados de forma percentual. E quanto mais ele se aproxima de 1, mais consegue explicar a relação dos dados preditos (JÚNIOR, 2021).

A fórmula para calcular o  $R^2$  é:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

### 3.2.2 Erro Médio Absoluto – MAE

MAE é uma estatística de erro que faz uma média da distância (valor absoluto da diferença entre os dados históricos reais e os dados previstos pelo modelo) entre cada par de pontos de dados de previsão reais e ajustados (POLANITZER, 2022, tradução nossa).

A fórmula para calcular o MAE é:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

### 3.3.3 Erro Percentual Absoluto Médio – MAPE

O MAPE é uma medida que indica o valor do erro dos valores preditos em relação aos reais. O cálculo dele é bem parecido com o MAE, mas acrescentando uma divisão por  $|y|$ . (JÚNIOR, 2021)

A fórmula para calcular o MAPE é:

$$MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\max(\epsilon, |y_i|)}$$

### 3.3.4 Erro Quadrático Médio – MSE

O MSE é uma medida de erro absoluta que acerta os erros para evitar que os valores positivos e negativos se cancelem. Essa medida também tende a exagerar grandes erros ponderando os grandes erros mais fortemente do que erros menores ao separá-los, o que pode ajudar na comparação de diferentes modelos de séries temporais (POLANITZER, 2022, tradução nossa).



A fórmula para calcular o MSE é:

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

### 3.3.5 Raiz do Erro Quadrático Médio – RMSE

RMSE é a raiz quadrada do MSE e é a medida de erro mais popular, também conhecida como função de perda quadrática. A RMSE pode ser definida como a média dos valores absolutos dos erros de previsão e é altamente apropriada quando o custo dos erros de previsão é proporcional ao tamanho absoluto do erro de previsão (POLANITZER, 2022, tradução nossa).

A fórmula para calcular o RMSE é:

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

## 3.4 Ferramentas Utilizadas

### 3.4.1 Python

Python é uma linguagem de programação de computador frequentemente usada para construir sites e softwares, automatizar tarefas e realizar análises de dados. Python é uma linguagem de uso geral, o que significa que pode ser usado para criar uma variedade de programas diferentes e não é especializado para quaisquer problemas específicos. Essa versatilidade, juntamente com sua simpatia iniciante, tornou-a uma das linguagens de programação mais utilizadas hoje (COURSERA, 2022, tradução nossa).

### 3.4.2 Pandas

Pandas é um pacote Python de código aberto que é mais amplamente utilizado para tarefas de análise de dados e aprendizado de máquina. Ele é construído em cima de outro pacote chamado NumPy, que fornece suporte para matrizes multidimensionais. Como um dos pacotes

de disputa de dados mais populares, o Pandas funciona bem com muitos outros módulos de ciência de dados dentro do ecossistema Python (ACTIVESTATE, 2021, tradução nossa).

### **3.4.3 NumPy**

NumPy é um dos pacotes mais usados para computação científica em Python. Ele fornece um objeto de matriz multidimensional, bem como variações como máscaras e matrizes, que podem ser usadas para várias operações matemáticas. NumPy é usado em conjunto com muitos outros pacotes Python populares, incluindo pandas e Matplotlib (ACTIVESTATE, 2022, tradução nossa).

### **3.4.4 Scikit Learning**

Scikit-learn é uma biblioteca em Python que fornece muitos algoritmos de aprendizagem não supervisionados e supervisionados. É construído sobre algumas das tecnologias como NumPy, Pandas e Matplotlib (CODECADEMY, 2022, tradução nossa).

### **3.4.5 Matplotlib**

Matplotlib é uma biblioteca de plotagem disponível para a linguagem de programação Python como um componente do NumPy. Este pacote utiliza uma API orientada a objetos para visualizar os gráficos desenvolvidos pela aplicação (TECHOPEDIA, 2022, tradução nossa).

### **3.4.6 Google Colab**

O Colab é um produto do Google Research. Colab, que permite que qualquer pessoa escreva e execute código Python através do navegador, e é especialmente adequado para aprendizado de máquina, análise de dados e educação. Mais tecnicamente, Colab é um serviço de notebook Jupyter hospedado que não requer nenhuma configuração para usar, ao mesmo tempo em que fornece acesso gratuito aos recursos de computação, incluindo GPUs (GOOGLE, 2022, tradução nossa).

## **4. Resultados e discussões**

Com o objetivo de eleger o algoritmo com a menor taxa de erros nas previsões realizadas, foram calculadas as métricas de erro para cada modelo – AD, RF e RL, utilizando os resultados do ativo ITUB4, para um período de 30 dias, de 24/05/2022 a 05/07/2022. A tabela 2 exibe os valores dos erros de cada algoritmo de previsão.

Tabela 2 – Resultado de Erros dos Modelos

Algoritmo	R <sup>2</sup>	MAE	MSE	RMSE	MAPE
Árvore de Decisão (AD)	86%	0,341	0,193	0,441	0,014
<i>Random Forest</i> (RF)	89%	0,295	0,151	0,388	0,012
Regressão Linear (RL)	91%	0,262	0,125	0,354	0,011

Fonte: O autor

Analisando a Tabela 2, é possível observar que o modelo RL teve o R<sup>2</sup> mais alto, indicando que ele sofre pouca variação, conseguindo explicar 91% dos dados preditos, enquanto os outros modelos tiveram desempenho ligeiramente menor.

Outro ponto a se observar, são as métricas MAE e MAPE, observa-se que os modelos estão errando, para mais ou para menos, entre 26% e 34% para o MAE e 1% para o MAPE.

Além disso, analisando as métricas MSE e RMSE, pode-se observar que os modelos de RL e RF ficaram com valores bem próximos, indicando uma variação de 0,354 e 0,388 entre o valor predito e o real.

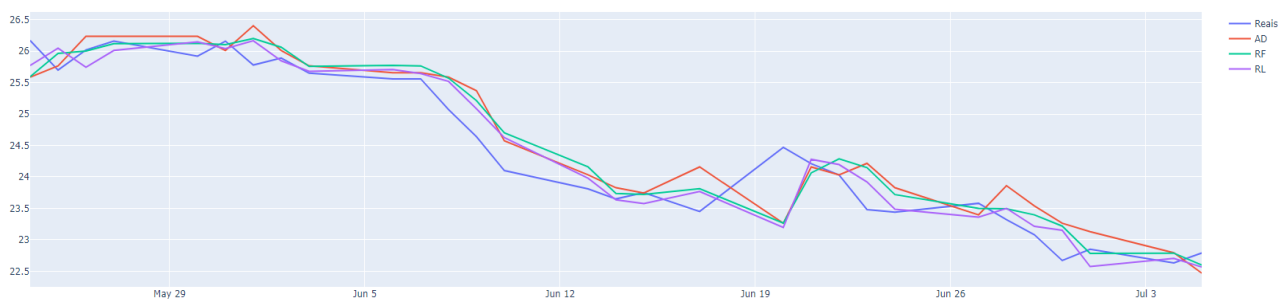


Figura 2 – Comparação entre os valores preditos e reais dos Modelos

Fonte: O autor

Através da Figura 2, pode se notar que todos os modelos conseguiram manter uma previsão bem próxima aos valores reais. Além disso, pode-se observar que em alguns dias

específicos, como na data de 20/06/2022, todos os modelos tiveram uma previsão destoante do valor real.

Outro ponto a se observar, é diferença entre os modelos de RF e AD, ambos sendo árvores de decisão, porém o RF gera uma maior quantidade de árvores de forma aleatória, e com isso melhora a predição em relação ao modelo com uma árvore de decisão simples.

## **5. Considerações finais**

O desenvolvimento deste artigo resultou na apresentação de três algoritmos de aprendizado de máquina com objetivo de prever o preço do ativo ITUB4, disponível na bolsa de valores brasileira (B3).

Foram utilizadas diversas técnicas de engenharia de dados para gerar os dados a serem utilizados nos modelos. Os resultados obtidos com este trabalho indicam baixas taxas de erros de previsão para o período em que foi proposto, indicando que o objetivo foi alcançado.

Para trabalhos futuros a ideia é testar novos algoritmos de *Machine Learning*, como *Gradient Boosting*, e *Deep Learning*, como redes neurais artificiais. E no futuro os modelos também podem realizar outros tipos de previsões, como risco de queda do valor do ativo, e indicação de compra e venda. Além disso, utilizar indicadores de análise técnica, como o MACD (*Moving Average Convergence/Divergence*) e do VWAP (*Volume Weighted Average Price*).

## 6. Referências Bibliográficas

Activestate. **What Is Pandas In Python? Everything You Need To Know.** 2021. Disponível em: <https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/>. Acesso em: 14 jul. 2022.

Activestate. **What Is Numpy Used For In Python.** 2022. Disponível em: <https://www.activestate.com/resources/quick-reads/what-is-numpy-used-for-in-python/>. Acesso em: 14 jul. 2022.

BROWN, Sara. **Machine learning, explained.** MIT Sloan. 2021 - Disponível em: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>. Acesso em: 8 jun. 2022.

Codecademy Team. **What is Scikit-Learn?.** Codecademy. Disponível em: <https://www.codecademy.com/article/scikit-learn>. Acesso em: 14 jul. 2022.

Coursera. **What Is Python Used For? A Beginner's Guide.** 2022. Disponível em: <https://www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python>. Acesso em: 14 jul. 2022.

CFI. **Fundamental Analysis.** 2022. Disponível em: <https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/fundamental-analysis/>. Acesso em: 12 jun. 2022.

FERNANDO, Jason; KINDNESS, David; KAZEL, Melody. **Earnings Per Share (EPS).** Investopedia. 2022 - Disponível em: <https://www.investopedia.com/terms/e/eps.asp>. Acesso em: 7 jul. 2022.

FERNANDO, Jason; BROCK, Thomas; LI, Timothy. **Price-to-Earnings (P/E) Ratio.** Investopedia. 2022. Disponível em: <https://www.investopedia.com/terms/p/price-earningsratio.asp>. Acesso em: 16 jun. 2022.

FINKLER, Aline Cristiane. **Aprendizagem de Máquina Aplicada à Previsão dos Movimentos do Ibovespa.** UFPR. 2017 - Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/49395/R%20-%20D%20-%20ALINE%20CRISTIANE%20FINKLER.pdf?sequence=1&isAllowed=y>. Acesso em: 17 jun. 2022.

FURHMANN, Ryan; JAMES, Margaret; LI, Timothy. **How to Calculate Return on Equity (ROE).** Investopedia. 2022. Disponível em: <https://www.investopedia.com/ask/answers/070914/how-do-you-calculate-return-equity-roe.asp>. Acesso em: 7 jul. 2022.

GeeksforGeeks. **ML | Linear Regression.** 2022. Disponível em: <https://www.geeksforgeeks.org/ml-linear-regression/>. Acesso em: 11 jun. 2022.

GIUDICCE, Thiago Lucas; ESTENDER, Antonio Carlos. **O Processo de Análise de Investimentos Financeiros em Instituições Financeiras**. Revistas PUC-SP. 2017. Disponível em: <https://revistas.pucsp.br/index.php/caadm/article/view/30867/25023>. Acesso em: 16 jun. 2022.

Google. **Colaboratory**. Disponível em: <https://research.google.com/colaboratory/faq.html>. Acesso em: 13 jul. 2022.

GUPTA. **Decision Trees in Machine Learning**. Towards Data Science. 2017. Disponível em: <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>. Acesso em: 11 jul. 2022.

IBM. **Linear regression**. Disponível em: <https://www.ibm.com/topics/linear-regression>. Acesso em: 8 jul. 2022.

IBM Cloud Education. **Machine Learning**. 2020. Disponível em: <https://www.ibm.com/cloud/learn/machine-learning>. Acesso em: 27 jul. 2022.

IG. **Fundamental analysis definition**. Disponível em: <https://www.ig.com/en/glossary-trading-terms/fundamental-analysis-definition>. Acesso em: 11 jun. 2022.

JÚNIOR, Clébio de Oliveira. **Prevendo Números: Entendendo as métricas R<sup>2</sup>, MAE, MAPE, MSE e RMSE**. 2021. Disponível em <https://medium.com/data-hackers/prevendo-n%C3%B0meros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>. Acesso em: 7 jul. 2022.

KHAIDEM, Luckyson; SAHA, Snehanshu; DEY, Sudeepa Roy. **Predicting the direction of stock market prices using random forest**. 2016. Disponível em: <https://arxiv.org/pdf/1605.00003.pdf>. Acesso em: 17 jun. 2022.

LEMOS, Eliane Prezepiorski; STEINER, Maria Teresinha Arns; NIEVOLA, Julio César. **Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining**. RAUSP. 2005 - Disponível em: <https://www.redalyc.org/pdf/2234/223417392002.pdf>. Acesso em: 21 jun. 2022.

MARQUES, Albertino. **Conceção e Análise de Projetos de Investimento**. Lisboa, 2014. 19 p. Disponível em: <https://static.fnac-static.com/multimedia/PT/pdf/9789726187813.pdf>. Acesso em: 2 jul. 2022.

MBAABU, Onesmus. **Introduction to Random Forest in Machine Learning**. 2020. Disponível em: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/>. Acesso em: 23 jun. 2022.

Nubank. **O que é B3 e como ela funciona**. 2021. Disponível em: <https://blog.nubank.com.br/o-que-e-b3-e-como-ela-funciona/>. Acesso em: 10 jun. 2022.

POLANITZER, Roi. **The Minimum Mean Absolute Error (MAE) Challenge**. 2022. Disponível em: <https://medium.com/@polanitzer/the-minimum-mean-absolute-error-mae-challenge-928dc081f031>. Acesso em: 7 jul. 2022.

STATUS INVEST. **P/VP**. 2020. Disponível em: <https://statusinvest.com.br/termos/p/p-vp>. Acesso em: 17 jun. 2022.

Techopedia. **Matplotlib**. Disponível em: <https://www.techopedia.com/definition/33861/matplotlib>. Acesso em: 14 jul. 2022.

TORO. **Mercado de capitais: desvende o que é e como funciona na prática**. 2022. Disponível em: <https://blog.toroinvestimentos.com.br/mercado-de-capitais-o-que-e>. Acesso em: 10 jun. 2022.

YIU, Tony. **Understanding Random Forest**. 2019. Disponível em: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. Acesso em: 23 jun. 2022.