

**Área Temática: Finanças**

**ANÁLISE DE INVESTIMENTOS: PREVISÃO DO PREÇO DE 5 AÇÕES DO SETOR  
BANCÁRIO BRASILEIRO UTILIZANDO *MACHINE LEARNING***

## **Resumo**

Desde o início das Bolsas de Valores ao redor do mundo, muitos investidores e cientistas tentam saber para onde vai o preço das ações visando assim diminuir seu risco e buscar os melhores retornos possíveis. Entretanto esta tarefa de tentar prever o futuro dos preços é de alta complexidade visto uma enorme quantidade de variáveis que podem afetar a previsão. Diante desses fatos e com o avanço de novas tecnologias, novos sistemas, modelos e algoritmos baseados no uso de inteligência artificial tem sido desenvolvido para auxiliar o ser humano em vários setores e um desses é justamente o mercado financeiro. Este artigo tem como objetivo utilizar dois desses métodos de inteligência artificial (*Random Forest* e *SVM*), para prever a direção do preço das ações de cinco bancos brasileiros listados na B3. Foram utilizadas as variáveis CAMELS, assim como informações contábil-financeiras e dados de mercado entre os anos de 2001 e 2021. Os resultados mostram que as variáveis fundamentalistas foram mais assertivas e, apesar da previsão média baixa (55%, variando entre 40% e 67%), uma análise qualitativa revela que fatos relevantes sinalizam erros nos modelos, o que poderia melhorar o poder preditivo quando se desconsidera períodos após divulgação dessas informações.

**Palavras-chave:** Mercado Financeiro, Ações, Bancos, Inteligência Artificial, Previsão de Tendência.

## **Abstract**

*Since the beginning of stock exchanges around the world, many investors and scientists have been trying to figure out where the stock price looking at lower their risk and seek the best returns possible. However, this task of trying to predict the future of prices is of high complexity given a huge number of variables that can affect the forecast. Given these facts and with the advance of new technologies, new systems, models, and algorithms based on the use of artificial intelligence has been developed to assist the human being in various sectors and one of these is precisely the financial market. This article aims to use two of these artificial intelligence methods (Random Forest and SVM), to predict the direction of the stock price of five Brazilian banks listed in B3. We add CAMELS indicators, as well as accounting-financial information and market data between 2001 and 2021. Our findings show that the fundamental variables were more assertive and, despite the low mean forecast (55%, ranging between 40% and 67%), a qualitative analysis reveals that relevant facts would anticipate errors in the models, which could improve predictive power when considering periods after disclosure of this information.*

**Keywords:** Financial Market, Stocks, Banks, Artificial Intelligence, Trend Forecast.

## 1 Introdução

O mercado de capitais tem como papel principal, a alocação de capital na economia (FAMA, 1970). Nele é possível negociar títulos e ativos financeiros publicamente, que possibilitam às empresas, arrecadar capital para pesquisa e desenvolvimento de produtos e serviços, crescimento e expansão, bem como aquisição de um concorrente. Ainda de acordo com a Hipótese dos Mercados Eficientes (HME), Fama (1970, 1991), observa que os preços dos títulos negociados no mercado financeiro, refletem todas as informações publicamente disponíveis. Vale ressaltar que a hipótese dos mercados eficientes não explica a existência de bolhas na economia.

Esta hipótese defende que nenhum investidor é capaz de superar os ganhos médios do mercado utilizando estratégias de investimento e sim por sorte. Com isso, tanto a análise técnica que é o estudo de preços anteriores na tentativa de prever os preços do futuro e a análise fundamentalistas que é a análise das informações financeiras, seriam capazes de ajudar a selecionar ativos que obteriam retornos maiores do que uma seleção aleatória (MALKIEL, 2003).

Entretanto, no final da década de noventa, alguns estudos mostraram evidências de que os retornos financeiros podem ser previsíveis através do uso de informações públicas sobre índices financeiros, assim como da trajetória de preços (BALVERS; COSIMANO; MCDONALD, 1990; LO; MACKINLAY, 1988). Outros autores como Lakonishok, Shleifer e Vishny (1994) e Davis (1994) mostram que os retornos futuros podem ser previstos utilizando os tamanhos relativos do preço de mercado atual das ações e pelas informações contábeis. Para Dosamantes (2013) prever o futuro é algo complexo, e uma das maneiras de prever os ganhos futuros é analisando as demonstrações financeiras do passado.

Ao analisar a criação de um portfólio de investimentos, a Teoria de Markowitz (1952) divide em duas etapas como deve ser essa seleção do portfólio. A primeira etapa é uma fase de observação e experimentação e a segunda etapa é sobre as crenças do desempenho futuro dos títulos e finaliza com a escolha do portfólio. Em seu estudo, Markowitz (1952), concluiu que métodos melhores serão encontrados e formulados e levarão em conta mais informações, visando uma melhor seleção do portfólio diminuindo assim seu risco de retorno.

Para Faria et al. (2009) o desenvolvimento de sistemas avançados tem modificado os mercados financeiros em todo mundo nos últimos anos. Com isso a tecnologia também melhorou neste sentido possibilitando a criação de sistemas, modelos e algoritmos visando a melhorar a qualidade e o conforto da vida humana. Rossi (2018) afirma que ao longo da década de noventa e início do século XXI, uma maior sofisticação da estatística, estimulou um número maior de tentativas empíricas de encontrar evidências a favor ou contra a previsibilidade dos retornos e da volatilidade dos ativos.

A previsão de ativos no mercado financeiro é realmente difícil devido à sua natureza não linear, dinâmica e complexa. Entretanto, nos últimos anos, as técnicas de *Machine Learning* (ML) têm se mostrado eficazes (KHAIDEM; SAHA; DEY, 2016).

Neste contexto, a previsão do desempenho de preços e índices dos mercados financeiros é um desafio, mesmo para experientes estudiosos da área. Este trabalho tem como objetivo central antecipar a tendência futura no preço de ações para cinco ativos do setor bancário: Banco de Brasil (BBAS3), Bradesco (BBDC3), BTG Pactual (BPAC11), Itaú (ITUB3) e Banco Inter (BIDI4), todos listados na Bolsa de Valores Brasileira (B3). Para tanto, observa-se e analisa-se desempenho a partir de técnicas de ML, mais especificamente, *Random Forest* (RF) e *Support Vector Machine* (SVM).

A partir de dados fundamentalistas e de mercado entre 2001 a 2020, também foi apreciado a influência de informações oriundas de Fatos Relevantes.

Dentre os algoritmos utilizados, o SVM apresentou melhores percentuais de precisão dos preços dos ativos em relação a RF, embora uma análise mais detalhada mostrar que, em alguns casos, suas previsões apontam apenas uma direção. Vale ressaltar que dentre as variáveis escolhidas, os dados fundamentalistas são os que apresentam resultados mais consistentes junto a utilização dos modelos.

O estudo nos mostra que o uso de algoritmos de Inteligência Artificial pode ajudar na tomada de decisão no momento de investir, visto a possibilidade da previsão da direção dos preços, possibilitando assim uma maior maximização do lucro e a provável mitigação de riscos para investidores do mercado de capitais.

Este trabalho está organizado da seguinte forma: além desta introdução, na seção 2 a fundamentação teórica – explanação sobre o setor bancário e as determinantes utilizadas para mensurar seu desempenho, como também aplicações de ML em Finanças. Na seção 3, está contemplada com os dados e métodos utilizados. A seção 4 deste artigo traz os resultados e análises e a seção 5 apresenta as conclusões.

## **2 Revisão da Literatura**

Um mercado de ações eficiente contribui para atrair mais investimento, financiando projetos que levem ao crescimento econômico, alocando eficientemente o capital, reduzindo e diversificando o risco e facilitando a troca de bens e serviços (MISHKIN, 2001). Os trabalhos realizados por King e Levine (1993) e também Demirgüç-Kunt e Maksimovic (1998) forneceram evidências empíricas robustas de que mercados financeiros mais fortes e consolidados estão fortemente associados com melhores perspectivas de crescimento econômico futuro.

Em um desses estudos, Levine e Zervos (1998), analisaram dados de 47 países entre os anos de 1976-93 e encontraram uma correlação positiva entre as taxas atuais e futuras do crescimento econômico com a liquidez do mercado de ações.

### **2.1 Setor Bancário e suas determinantes**

A América Latina na década de noventa, estava passando por transformações resultantes do processo de liberalização financeira e integração internacional, culminando em uma rápida consolidação dos sistemas financeiros desses países. (CHORTAREAS; GARZA-GARCIA; GIRARDONE, 2010).

O setor financeiro brasileiro, tanto o setor bancário e o mercado de ações foram expostos a essa série de reformas. No final dos anos 1980 marcaram o início da reforma do setor bancário no Brasil. As reformas foram de amplo alcance e incluíram programas para: criação de instituições financeiras especializadas; reestruturação do setor privado e bancos controlado pelo Estado; e a decisão de permitir a entrada de bancos estrangeiros no sistema financeiro nacional (Carneiro et al., 1993).

Yeyati e Micco (2003) observaram esse processo, na qual bancos estrangeiros, principalmente bancos europeus e bancos americanos investiram na aquisição de bancos locais. Conseqüentemente, a entrada de bancos estrangeiros, estimulou a eficiência do setor bancário como um todo, principalmente dos bancos domésticos, que para não perderem sua participação de mercado, tiveram que melhorar o serviço oferecido (CLAESSENS; DEMIRGÜC-KUNT; HUIZINGA, 2001).

Yildirim e Philippatos (2006) promoveram um estudo entre os anos de 1993 e 2000, visando analisar as condições competitivas entre os mercados bancários de onze países na América Latina. Os resultados mostraram que a inserção de bancos estrangeiros, pode estimular a concorrência forçando os bancos domésticos a

melhorarem seus serviços operacionais. Ainda para Chortareas, Garza-Garcia e Girardone (2010), o resultado desse processo é visto atualmente, em que o mercado bancário latino-americano é altamente concentrado tendo assim, efeitos na conduta e lucratividade destes bancos.

Em outro estudo, Levine e Zervos (1993) usaram diferentes métricas de desenvolvimento bancário para vários países. Eles descobriram que o desenvolvimento desse setor pode estimular o crescimento econômico no longo prazo. Boyd e Prescott (1986) e Stiglitz e Weiss (1981) argumentam que o desenvolvimento do setor bancário pode desempenhar um papel importante na promoção do crescimento econômico, pois os bancos são melhores do que o mercado de ações quando se trata de alocação de recursos.

Molyneux e Thornton (1992) foram os primeiros a explorar completamente as determinantes da lucratividade bancária em uma amostra de 18 países europeus durante o período de 1986 a 1989. Eles encontraram uma associação positiva e significativa de concentração bancária de bancos ligados ao governo com lucratividade. Nos estudos de Kunt e Huizinga (1998), eles analisaram os determinantes da lucratividade dos bancos comerciais usando dados bancários de 80 países do período de 1988 a 1995, eles incorporaram vários fatores como características do banco, condições macroeconômicas, impostos, e a estrutura financeira geral. Os resultados encontrados mostram que os bancos capitalizados têm maior lucratividade devido ao baixo custo de captação de recursos devido a menor perspectiva de falência.

Uma característica inerente aos bancos que vale ser ressaltada, são os diversos riscos que essas empresas enfrentam. Leo et al. (2019) argumentam que existem o risco da taxa de juros, de mercado, risco de crédito, risco operacional, cambial, de liquidez, risco país, tecnológico entre outros. E dentre o desempenho apresentado pelos bancos, a gestão eficiente desses riscos é fundamental.

Para Wirnkar e Tanko (2008), a utilização de índices financeiros são frequentemente usados para medir a solidez geral de um banco e sua gestão. Outra abordagem para medir o desempenho de um banco é utilizando uma combinação de índices financeiros, benchmarking, desempenho do orçamento (AVKIRAN, 1995).

Existe ainda a classificação CAMELS. Ela é usada para avaliar a saúde financeira e o desempenho do banco (ROZZANI; RAHMAN, 2013). Esse modelo foi criado em 1991, nos EUA e desde então tem sido uma ferramenta eficaz para medir o desempenho dos bancos (TODOROVIC; FURTULA; DURKALIC, 2018). Entretanto, segundo Hirtle e Lopez (1999), os ratings CAMELS, são estritamente destinados aos gestores da alta administração dos bancos assim como os órgãos reguladores competentes, tornando assim uma informação altamente confidencial. Conforme Hirtle e Lopes (1999), a sigla CAMELS refere-se às seguintes condições avaliadas: Adequação de Capital (C), Qualidade dos Ativos(A), Gestão (M), Resultados (E), Liquidez (L) e Sensibilidade ao risco de mercado (S).

A literatura relacionada trás diversos estudos que utilizaram o modelo CAMELS. Em um desses estudos, DeYoung (1998), analisou a qualidade de gestão e ineficiência nos bancos. Os resultados obtidos foram que em bancos bem administrados os custos unitários estimados eram mais baixos que em bancos mal administrados. Outro trabalho interessante utilizando o modelo CAMELS, foi realizado por Christopoulos, Mylonakis e Diktapanadis (2011), na qual utilizaram o modelo para analisar sobre a possível falência do Banco Lehman Brothers, e que os resultados encontrados mostram o declínio do banco.

Em outro estudo utilizando o modelo CAMELS, Nandi (2013), comparou os bancos da Índia, separando-os em bancos públicos e privados. Os resultados mostraram que os bancos públicos apresentaram um desempenho melhor do que os bancos privados. Esse fator se deu pela dependência de receita de juros por parte dos bancos privados. Em um estudo similar, Khatri, Arora e Kumar (2018), divergiram do estudo anterior. Os autores encontraram por meio do modelo CAMELS, que não existe uma diferença significativa do desempenho de bancos públicos e privados na Índia.

## **2.2 Análise de Investimentos com *Machine Learning***

Os investidores no mercado de ações querem maximizar seus retornos comprando ou vendendo seus ativos no momento certo. Entretanto, essa análise dos preços futuros dos ativos é bem complexa. Uma vez que os dados do mercado de ações são altamente variantes no tempo e normalmente estão em um padrão não linear, prever o preço futuro de uma ação é altamente desafiador.

A previsão fornece informações sobre o status atual do movimento de preços. Assim, pode ser utilizado na tomada de decisão para os clientes no momento de comprar ou vender ações específicas de uma determinada empresa. (PATEL; YALAMALLE, 2014). A literatura empírica mostra que o retorno médio das ações ordinárias está relacionado às características da empresa, como tamanho, lucro / preço, fluxo de caixa / preço, patrimônio líquido de contabilização e outros indicadores financeiros. Banz (1981), Koehn e Santomero (1980), Blum (1999), Cooper et al. (2003), relataram uma relação significativa entre micro variáveis e o retorno das ações.

Dentro do campo da inteligência artificial, o Aprendizado de Máquina foi concebido com o objetivo de desenvolver métodos computacionais, que implementaram várias formas de aprendizagem, para solucionar problemas que necessitem de soluções algorítmicas (KUBAT; BRATKO; MICHALSKI, 1998). Em seu livro "*The Discipline of Machine Learning*", Mitchell (2006), descreve que as aplicações do aprendizado de máquina são as mais variadas, desde sua utilização para diagnósticos médicos, como para construção de robôs e veículos autônomos. Palit e Roy (2020), em seu estudo, descrevem uma série de aplicações práticas de *machine learning*, entre elas pode-se citar as seguintes: Cuidados de saúde, Conversas Inteligentes, Processos de Contratação, Detecção de Fraudes em bancos e seguradoras, Sistemas de Recomendações (anúncios) e nos Transportes.

Existem, diversos estudos e trabalhos envolvendo *Machine Learning*, em um desses estudos relacionados a área de finanças, Barboza, Kimura e Altman (2017), testaram diversos modelos de aprendizado de máquina, entre eles o de RF, na previsão de falências. Esse estudo analisou dados de mais de 10.000 empresas norte-americanas entre os anos de 1985 a 2013. Os resultados mostraram que a utilização de modelos de *Machine Learning*, apresentaram em média, 10% a mais de precisão em relação a modelos tradicionais. O mais interessante deste estudo foi o que o uso de RF apresentou uma precisão de 87% na previsão, enquanto os modelos de regressão logística e análise discriminante linear, apresentaram 69% e 50% de precisão respectivamente.

Em outro estudo da área de finanças, agora envolvendo criptomoedas, Valencia, Gómez-Espinosa e Valdés-Aguirre (2019), utilizaram as ferramentas de *Machine Learning* para prever o movimento de quatro criptomoedas: Bitcoin, Ethereum, Ripple e Litecoin. Os autores mostraram que é possível prever o mercado de criptomoedas usando aprendizado de máquina. Segundo Ferreira, Gandomi e Cardoso (2021) as principais vantagens do uso computacional baseado em inteligência artificial, seria automatizar o processo de investimento. Os autores afirmam que essa automação

ajudaria a eliminar a “irracionalidade momentânea”, ou simplesmente evitar que decisões sejam tomadas com base em emoções.

Um estudo de Gu, Kelly e Xiu (2018) analisou o mercado acionário norte americano utilizando ferramentas e técnicas de ML. Os autores puderam concluir que os ganhos econômicos das previsões de aprendizado de máquina foram realmente grandes. Ainda de acordo com este estudo, as previsões de tendências de preço, liquidez e volatilidade foram as que tiveram maior sucesso. O aprendizado de máquina tem um grande potencial para melhorar o prêmio do risco de cálculo, que é fundamentalmente um problema de previsão. (GU; KELLY; XIU, 2018, tradução nossa)

Em outro estudo utilizando modelo de ML aplicado no mercado financeiro, Laboissiere, Fernandes e Lage (2015), utilizaram de Redes Neurais Artificiais para prever os preços máximos e mínimos de três empresas do setor de energia no Brasil. A metodologia utilizada foi eficaz em estimar a faixa de preço das ações e podem ser utilizadas para auxiliar na tomada de decisão dos investidores.

Dentre os métodos de *Machine Learning*, existe o *Random Forest* (RF) que significa floresta aleatória. O RF são de fácil entendimento e tendem a ter um bom desempenho na previsão do preço das ações (BASAK et al., 2019; KHAN et al., 2020). Para Breiman (2001), idealizador do modelo, trata-se de um dos métodos mais conhecidos e utilizados métodos que apresenta ótimos resultados e tem várias aplicações práticas. Esse modelo foi baseado nas árvores de decisão e duas características usadas na construção do modelo são o ensacamento e a seleção aleatória em cada nó. (BREIMAN, 1996). Durante a construção do modelo, o melhor recurso é selecionado dentre outros e uma divisão de nó é realizada com base nesse recurso. Com isso, aumenta-se a diversidade e a aleatoriedade do modelo, tendo assim um melhor resultado (PALIT; ROY, 2020).

Para Soni (2018) o uso de RF está em expansão na previsão de preço no mercado de ações. Desde que foi nomeado um dos mais fáceis de usar e de grande precisão, um número de cientistas cada vez maiores estão se dedicando a montar estratégias para o modelo ficar cada vez mais preciso. Para Han, Pei e Kamber (2011) a precisão do algoritmo de RF é muito eficiente visto que são mais robustas para erros e outliers. Outro ponto interessante, é a combinação de várias árvores de decisão para aumentar a acurácia de previsão do modelo. Esse método segundo Ballings et al. (2015), estão entre os métodos que apresentam os melhores resultados em aplicações relacionadas ao mercado de ações.

As Máquinas de Vetores de Suporte conhecidas por sua sigla SVM, é um tipo de máquina de aprendizagem. Este algoritmo foi descrito inicialmente por Vapnik (1999). O autor afirma que o SVM possui algumas propriedades úteis como o fato da otimização para encontrar uma solução única e pela rapidez de aprendizagem no processo de construção do algoritmo. Reddy (2018) afirma que o SVM é considerado como um dos algoritmos existentes mais adequados para a previsão de séries e pode ser utilizado tanto para regressão como para classificação, a qual este trabalho utiliza. Em um trabalho de Patel et al. (2015), foi analisado 10 anos de dados históricos de 2003 a 2012 das empresas Reliance Industries e Infosys Ltd. e dois índices de preços de ações CNX Nifty e S&P Bombay Stock Exchange (BSE) Sensex. Foram utilizados os métodos de RNA, SVM, RF e naive-Bayes para prever a direção do movimento das ações e índices. Inicialmente, aplicaram dez indicadores técnicos. O resultado mostrou que o melhor modelo de previsão foi o de RF com 83,56% de precisão. Já na segunda abordagem foram utilizados dados determinísticos de tendência e, assim, todos os modelos apresentaram uma precisão maior, com destaque para o modelo de

RF com 89,98% de precisão, enquanto o melhor modelo foi a *naive-Bayes* com 90,19% de precisão do modelo.

No estudo de Khaidem, Saha e Dey (2016), produziu resultados expressivos ao atingir uma faixa entre 85% e 95% de precisão para o longo prazo com RF, o que mostra a robustez do modelo. Em outro estudo parecido, Sadorsky (2021) obteve 80% de precisão para um horizonte de 10 dias. Já para um horizonte de 20 dias, as taxas de precisão do modelo de *Random Forest* ficaram entre 85% e 90% de exatidão.

### 3 Metodologia

Este estudo tem característica exploratória, com a finalidade de analisar o uso de duas ferramentas de *Machine Learning* para a previsão da direção dos preços de 5 ativos do setor bancário brasileiro que segundo dados coletados juntos ao site da B3 (2021) representam juntas cerca de 10% do índice Bovespa, e figuram entre os maiores valores de mercado dentre os bancos listados no índice. Para a realização do trabalho, a Metodologia foi definida seguindo a ordem de passos na Figura 1 (a seguir).

Entre os bancos selecionados, o Banco do Brasil (BBAS3) é o único banco de controle estatal brasileiro analisado enquanto os demais bancos são pertencentes à iniciativa privada. Entretanto existe ainda uma diferença entre tais bancos. Os bancos Bradesco (BBDC3) e Itaú (ITUB3) são considerados parte do “grupo dos grandes *players*” dentro do cenário bancário brasileiro, justamente por conta do enorme número de correntistas, tempo de existência, receitas e lucros figurando entre os maiores do Brasil. Já o BTG Pactual (BPAC11) tem como característica principal o fato de ser um banco especializado em investimentos, sendo considerado o maior banco da América Latina neste segmento bancário. Por fim, o banco mais novo analisado é o Inter (BIDI4), que em anos anteriores mantinha uma estrutura de banco tradicional, mas ao longo dos últimos, mudou sua forma de atuação, sendo considerado hoje no meio por ser um banco digital, oferecendo seus serviços principalmente na internet.

**Figura 1** – Diagrama de blocos para Metodologia proposta



Fonte: Os autores.

Os dados fundamentalistas, como os Balancetes e Balanços Patrimoniais, foram coletados juntos ao site do Banco Central do Brasil (2021). Os dados de mercado, que incluem preço de abertura e fechamento, como também o volume, foram coletados no site Yahoo Finance. Os indicadores técnicos foram calculados por meio da biblioteca TA-Lib: *Technical Analysis Library* (2021), que possui cerca de 200 indicadores utilizados no mercado financeiro. Foram utilizados também dados do PIB brasileiro, assim como dados do desemprego coletados junto ao site do Instituto Brasileiro de Geografia e Estatística (2021). As fórmulas para calcular os indicadores CAMELS foram abordadas com detalhes por Barboza et. al (2021).

Vale ressaltar que o período dos dados utilizados na pesquisa foi de 2001 a 2020. A Figura 2 (a seguir) ilustra o preço das ações citadas. Devido a cada ativo ter iniciado sua negociação na Bolsa de Valores em épocas diferentes, os períodos analisados também serão diferentes. BBDC3, BBAS3 e ITUB3 são os ativos com maior quantidade de dados, enquanto BPAC11 e BIDI4 têm quantidade menor de dados coletados devido à recente entrada na Bolsa de Valores conforme a Tabela 1.

**Tabela 1** – Data Início e Fim dos dados coletados. A diferença entre os períodos analisados baseia-se na disponibilidade de dados, uma vez que nem todos eles possuem período extenso de negociação de ações em bolsa. Além disso, a necessidade de 80% da base de dados para desenvolvimento dos modelos exija o devido ajuste e melhor aproveitamento dos dados.

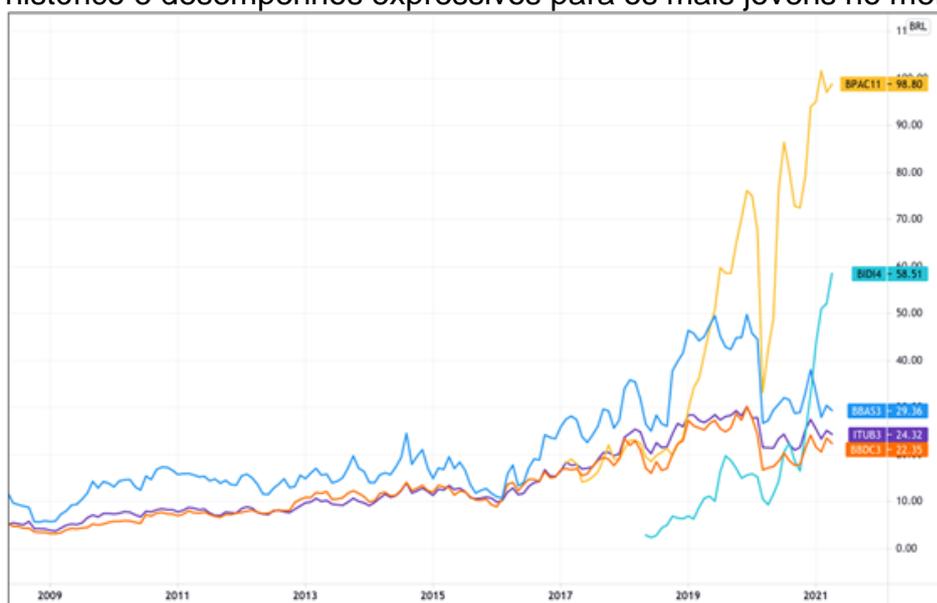
Ação	Treino		Teste	
	Data Início	Data Fim	Data Início	Data Fim
BBDC3	03/12/2001	24/05/2017	25/05/2017	23/03/2021
BBSA3	01/12/2001	08/06/2017	09/06/2017	23/03/2021
BIDI4	02/05/2018	14/09/2020	15/09/2020	23/03/2021
BPAC11	24/02/2017	22/06/2020	23/06/2020	23/03/2021
ITUB3	03/12/2001	24/05/2017	25/05/2017	23/03/2021

Fonte: Os autores.

Para os algoritmos de *Machine Learning* utilizados, tanto o de *Random Forest* como SVM, foi utilizada a linguagem de programação Python. As bibliotecas utilizadas para rodar os modelos foram a Pandas e a Scikit-Learn. Por fim, o algoritmo foi desenvolvido no ambiente computacional Jupyter Notebook.

O pré-processamento dos dados (tratamentos e remoção de observações com dados faltantes) é o primeiro passo necessário para o emprego adequado dos modelos. Após esse tratamento inicial, são adicionados os atributos fundamentalistas, de mercado e as CAMELS no modelo. Por fim, é necessário separar os dados cronologicamente em treino e teste, sendo que a maior parte dos dados coletados sejam destinados para treino (80%) e o restante para teste (20%), conforme ilustrado na Tabela 2.

**Figura 2** – Preço dos Ativos analisados (BBDC3 [laranja]; BBAS3 [azul]; BIDI4 [ciano]; BPAC11 [amarelo]; ITUB3 [roxo]). Percebe-se uma certa correlação entre os bancos de maior histórico e desempenhos expressivos para os mais jovens no mercado.



Fonte: Trading View

**Tabela 2** – Quantidade de observações coletadas para as amostras de Treino e Teste.

Ação	Treino	Teste	% Treino	%Teste
BBDC3	3792	949	80%	20%
BBSA3	3750	938	80%	20%
BIDI4	496	125	80%	20%
BPAC11	734	184	80%	20%
ITUB3	3792	949	80%	20%

Fonte: Os autores.

Finalizados todos esses processos, os algoritmos estão prontos para criar os modelos de previsão baseados em todas as informações coletadas, gerando assim as previsões para um período futuro de 90 dias. Cabe ressaltar que esses o período de previsão, são dias úteis em que ocorreu pregão na bolsa de valores brasileira, desconsiderando assim dias não uteis como finais de semana e feriados. Dentre os dados coletados durante os anos analisados, os preços reais das ações subiram mais do que desceram conforme apresentado na Tabela 3.

A fim de observar o desempenho com diferentes dados de entrada, foram gerados três modelos para cada algoritmo de aprendizado de máquina. O primeiro modelo utilizou apenas os dados fundamentalistas, o segundo modelo utilizou apenas os dados de mercado e por fim, o último modelo de previsão utilizou ambos os dados para gerar as previsões.

**Tabela 3** – Quantidade de eventos (quedas e altas nos preços dos ativos, em dias) ocorridos em cada uma das amostras.

Ação	Treinamento		Teste	
	Queda	Alta	Queda	Alta
BBDC3	1391	2401	348	601
BBSA3	1558	2192	424	514
BIDI4	154	342	0	125
BPAC11	204	530	29	155
ITUB3	1353	2439	383	566
<b>TOTAL</b>	<b>4660</b>	<b>7904</b>	<b>1184</b>	<b>1961</b>

Fonte: Os autores.

Para avaliar as previsões dos modelos analisados, assim como sua robustez, são utilizados métricas estatísticas (NABIPOUR et al., 2020), que são dadas pelas seguintes fórmulas:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Especificidade} = \frac{FN}{TN + FP} \quad (4)$$

onde TP = Quantidade de previsões de alta classificadas corretamente; TN= Quantidade de previsões de queda classificadas corretamente; FP = Quantidade de previsões de alta classificadas incorretamente; e, FN = Quantidade de previsões de queda classificadas incorretamente.

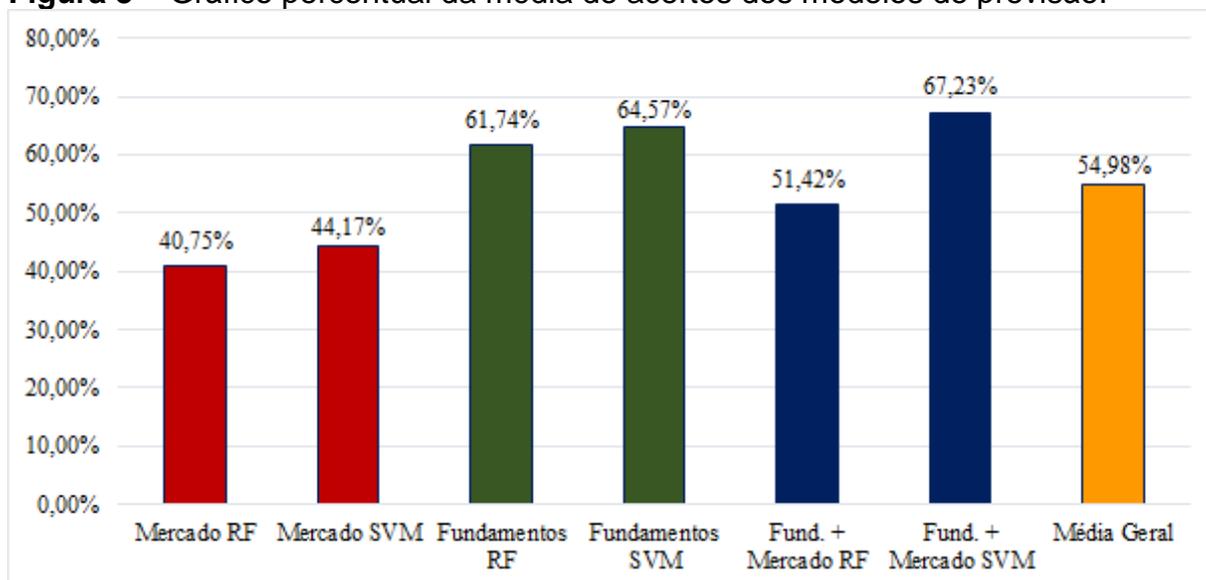
Após a análise estatística dos algoritmos de previsão, foi feita uma pesquisa junto ao site da Comissão de Valores Mobiliários (CVM) por fatos relevantes que justificassem alguns erros nas previsões realizadas na tentativa de identificar potenciais erros devidos a eventos imprevistos, que tem chances de serem apresentados aos investidores por meio de fatos relevantes.

## 4 Resultados

### 4.1 Performance das Previsões por *Random Forest* e SVM

Os algoritmos apresentaram juntos uma média geral de acertos nas previsões considerável, cerca de 54,98%, conforme a Figura 3. Em relação aos algoritmos separadamente, o SVM utilizando dados ambos os dados fundamentalistas, assim como os dados de mercado foi o modelo que mais conseguiu acertar em média nas previsões realizadas para os cinco bancos, aproximadamente dois a cada três (67,23%) previsões deste modelo acertou para qual direção o preço do ativo se movimentou. Vale ressaltar que os resultados analisados se referem a amostra teste. Ainda observando a Figura 3, percebe-se que a utilização dos dados fundamentalistas nos modelos, apresentaram um percentual de acertos acima de 60%, ficando assim acima da média geral e também ficaram bem próximos tanto apresentam uma diferença percentual de apenas 2,83%.

**Figura 3** – Gráfico percentual da média de acertos dos modelos de previsão.



Fonte: Os autores.

Por fim, percebe-se que o uso de dados técnicos de mercado, alinhado com os dois modelos propostos, apresentaram resultados bem ruins nas previsões. Ficando assim tanto abaixo da média geral com o percentual do modelo de SVM em 44,17% e o de *Random Forest* com um resultado pior ainda com 40,75% de acertos.

### 4.2 Análise por Ativo

Ao analisar os bancos individualmente, tem-se de observar as características e especificidades pontuais de cada um. A seleção dos bancos deu-se justamente observando essa questão. Esta análise compreende os dois modelos que tiveram maior percentual de acertos na previsão por banco. O objetivo é buscar os principais fatos relevantes que, ao serem divulgados ao mercado, alterem o rumo real dos ativos, ocasionando o erro dos modelos. A Figura 4 ilustra tais resultados.

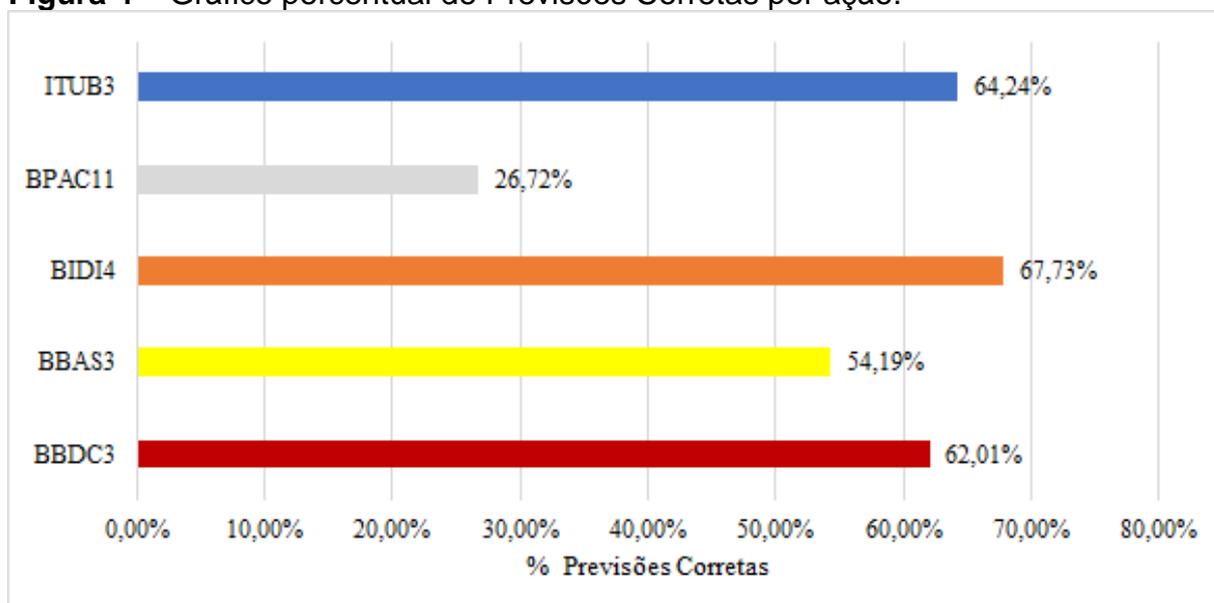
O BBAS3 conforme visualizado na Figura 2, estava apresentando uma tendência de alta no seu preço. Essa tendência foi percebida pelos modelos utilizados que traçaram previsões da alta do ativo de 2017 até o início do ano de 2019. Entretanto, houve um período de cerca de um trimestre, mais precisamente no início do ano de 2018, no qual o ativo sofreu uma queda. Uma das evidências encontradas para essa queda do ativo e no erro dos modelos, poderia ter sido o Fato Relevante publicado no dia 22/02/2018, na qual divulga informações da estimativa da Margem Financeira Bruta, a qual teria um decréscimo percentual estimado de -5 a 0.

Outro período interessante em que a previsão dos modelos errou foi entre os meses de agosto a outubro de 2019. Em fato relevante publicado em 29/07/2019, o Banco do Brasil comunicou, o redimensionamento da estrutura organizacional nos níveis estratégicos (direção-geral), táticos (superintendências) e demais órgãos regionais e agências. Tal informação agradou o mercado e com isso as ações tiveram uma valorização nos meses seguintes, afetando assim a previsão do modelo que estava prevendo uma queda.

Por fim, outro fato relevante publicado em 07/04/2020, consta que o BBAS3 reduziria seu percentual de dividendos, para o mínimo obrigatório. Com isso, observou-se uma tendência de queda dos preços, enquanto os modelos estavam prevendo uma tendência de recuperação alta.

O BBDC3 durante os anos analisados, estava apresentando uma tendência de alta, com alguns curtos períodos de queda. Em uma dessas quedas – que foi no início de 2018 – aconteceu exatamente a falha dos modelos. A partir de fatos relevantes sucessivos neste período, destaca-se a renúncia e a troca de vários diretores, bem como mudanças na Presidência e no Conselho de Administração do banco. Esses fatos relevantes podem ser uma explicação plausível para a queda ocorrida e justificar o erro das previsões.

**Figura 4** – Gráfico percentual de Previsões Corretas por ação.



Fonte: Os autores.

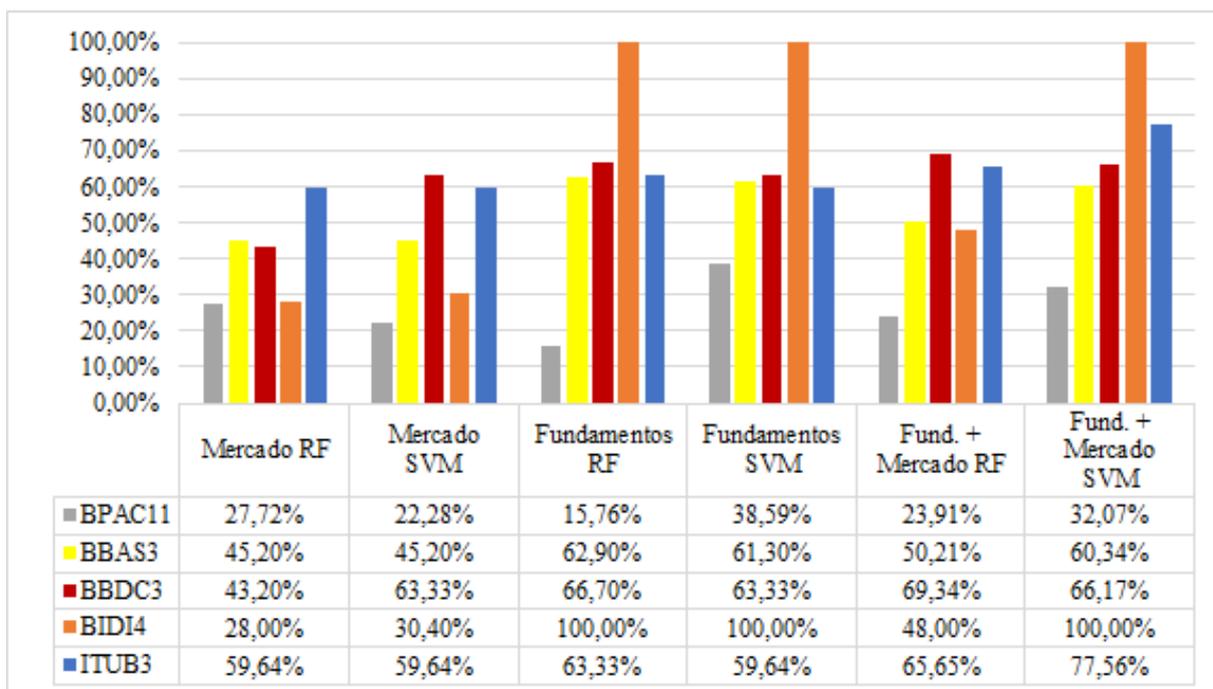
Outro período de erro dos modelos foi relacionado ao Fato Relevante Publicado em 06/08/2019, na qual a empresa divulga seus resultados institucionais do segundo trimestre de 2019. O resultado foi positivo e de crescimento em diversos indicadores. O mercado no geral reagiu positivamente com os números apresentados, posto que ocorreu uma alta no preço após tal divulgação, ou seja, precificando assim os bons resultados apresentados e ocasionando o erro do modelo que previa uma queda.

BIDI4 por apresentar um curto período analisado, os modelos que utilizaram variáveis gráficas apresentaram resultados na casa dos 30% de precisão de acordo com a Figura 5. Já na previsão utilizando as informações fundamentalistas apresentaram resultados de 100% de precisão tanto para o modelo de *Random Forest*, como para o SVM. Vale ressaltar que o Banco Inter apresenta uma tendência muito forte de alta de seu preço em um curtíssimo espaço de tempo conforme a Figura 2. Na média geral foi o que apresentou os melhores resultados, com o percentual de precisão chegando perto de 70%.

BPAC11 apresentou o pior percentual médio de acerto das previsões conforme Figuras 4 e 5. As ações do BPAC11 subiram em cerca de 85% dos dias analisados, enquanto os modelos previram a queda do ativo em cerca de 91% dos dias para o modelo de *Random Forest* e cerca de 83% de queda no modelo de SVM, observando que os dois modelos utilizaram tanto as informações de mercado com as fundamentalistas. Outro fator é o curto período do ativo analisado. Uma das hipóteses levantadas a respeito do erro geral dos modelos em relação a este ativo poderia ser os dados de governança que poderiam ser incorporados ao modelo para aumentar a precisão.

ITUB3 foi o banco que apresentou o melhor percentual de previsões corretas entre os “grandes *players* do mercado”, cerca de 64% em média. Alguns períodos interessantes na quais os modelos erraram foi no início de 2018, no qual no dia 05/02/2018, foi apresentada por meio de Fato relevante, as informações financeiras e contábeis consolidadas do ano de 2017. Tais informações podem ter colaborado para o erro do modelo visto que as ações do banco apresentaram tendência de alta e após resultados, viraram para queda.

**Figura 5** – Comparativo percentual dos modelos de previsão por bancos



Fonte: Os autores.

Outro Fato relevante que impactou a direção do ativo ITUB3 foi o relatório de demonstrações contábeis publicado em 29/10/2018 que inclui dados de janeiro a setembro daquele ano. As ações estavam apresentando um período de alta, o qual foi interrompido por uma queda do preço do ativo e poderia justificar assim o erro dos modelos neste período analisado.

Contudo, é possível observar que existem evidências de que fatos relevantes influenciam as decisões dos investidores de forma que tais informativos sejam sinalizadores de alerta para as previsões baseadas em inteligência artificial e, dessa forma, mitigar erros que a máquina se mostrou incapaz de corrigir.

## 5 Conclusão

O objetivo deste artigo é tentar prever a direção dos preços dos cinco ativos escolhidos utilizando dois modelos de Aprendizado de Máquina: *Random Forest* e SVM, a partir de dados fundamentalistas e técnicos, assim como as variáveis CAMELS.

Em suma, a previsibilidade média de 54% na direção dos preços no mercado acionário brasileiro foi observada. Para este estudo, o melhor algoritmo preditivo foi o SVM no qual utilizou-se os dados fundamentalistas como os dados de mercado, apresentando 67% de acertos em média. A melhor acurácia dos modelos de RF foi o que utilizou apenas informações fundamentalistas, chegando a 61%.

Outro fator observado, foi o impacto que informações divulgadas no mercado, como os fatos relevantes, podem impactar a direção de um ativo, tanto para alta, como para queda do preço. Os modelos de previsão não conseguem prever a data de publicação dessas informações e muito menos os conteúdos abordados. Com isso, analisou-se o papel dos fatos relevantes nos erros nas previsões, que mostraram relativa importância.

Entretanto, estudos futuros que utilizem inteligência artificial aplicado ao mesmo problema são necessários. Vale destacar que este estudo considerou apenas o setor bancário, abrindo oportunidades para mais pesquisas que englobem os demais.

## Referências

- AVKIRAN, N. K. Developing an Instrument to Measure Customer Service Quality in Branch Banking, **International Journal of Banks Marketing**, v. 12, p. 10-18, 1995.
- B3, 2021. Ações. Disponível em: [http://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm](http://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm) Acesso em: 01/03/2021.
- BALLINGS, M. et al. Evaluating multiple classifiers for stock price direction prediction. **Expert Systems with Applications**, v. 42, n. 20, p. 7046-7056, 2015.
- BALVERS, R. J.; COSIMANO, T. F.; MCDONALD, B. Predicting stock returns in an efficient market. **Journal of Finance**, v. 55, p. 1109-1128, 1990.
- BANZ, R. The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics*, v. 9, p. 3-18. 1981.
- BARBOZA, F.; KIMURA, H.; ALTMAN, E. Machine learning models and bankruptcy prediction. **Expert Systems with Applications**, v. 83, p. 405-417, 2017.
- BARBOZA, F.; BARBOSA, J. ; KIMURA, H.; SANTOS, G.; CORTEZ, P. Early Warning System for preventing bank distress in Brazil. *IJBRS*, in press, 2021.
- BASAK, S.; KAR, S.; SAHA S.; KHAIDEM, L.; DEY, S. R. Predicting the Direction of Stock Market Prices Using Tree-Based Classifiers. **The North American Journal of Economics and Finance**, v. 47, p. 552-67, 2019.
- BANCO CENTRAL DO BRASIL, 2021. Balancetes e Balanços Patrimoniais (Transferência de arquivos). Disponível em: <https://www.bcb.gov.br/estabilidade/financeira/balancetes/balancospatrimoniais>. Acesso em: 01/03/2021.
- BLUM, J. Do Capital Adequacy Requirements Reduce Risk in Banking?. **Journal of Banking and Finance**, v. 23, p. 755-71, 1999.
- BOYD, J. H.; PRESCOTT. E. Financial intermediary-coalitions. **Journal of Economic Theory**, v. 38, p. 211-232. 1986.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.
- BREIMAN, L. Bagging predictors. **Machine Learning**, v. 24, n. 2, p. 123-140, 1996.
- CARNEIRO, D. D.; WERNECK, R. L. F.; GARCIA, M. G. P.; BONOMO, M. A. Strengthening Brazil's financial economy, **Inter-American Development Bank - Working Paper Series**, v. 142, 1993.
- CHRISTOPOULOS, A.G.; MYLONAKIS, J.; DIKTAPANADIS, P. Could Lehman Brothers' collapse be anticipated? An examination using CAMELS rating system. **International Business Research**, v. 4, n. 2, p. 11-19, 2011.
- CHORTAREAS, G. E.; Garza-Garcia, J, G.; GIRARDONE, C. Banking sector performance in some Latin American countries: Market power versus efficiency. **Banco de México - Working Papers**, n. 2010-20, 2010.
- CLAESSENS, S.; DEMIRGÜC-KUNT, A.; HUIZINGA, H. How does foreign entry affect domestic banking markets?. **J. of Banking and Finance**, v. 25, p. 891-911, 2001.
- COOPER, M.; JACKSON, W.; PATTERSON, G. Evidence of Predictability in the Cross Section of Bank Stock Returns. **J. of Banking and Finance**, v. 27, p. 817-850, 2003.
- COMISSÃO DE VALORES MOBILIÁRIOS, 2021. Consulta de Documentos de Companhias Abertas – RAD. Disponível em: <https://www.rad.cvm.gov.br/ENET/frmConsultaExternaCVM.aspx>. Acesso em: 01/03/2021.
- DAVIS, J. L. The cross-section of realized stock returns: The pre-Compustat evidence. **Journal of Finance**, v. 49, p. 1579-1593, 1994.
- DEMIRGUÇ-KUNT, A.; MAKSIMOVIC, V. Law, Finance and Firm Growth. **Journal of Finance**, v.53, p. 2107-2137, 1998.
- DeYOUNG, R. Management quality and X-inefficiency in national banks. **Journal of Financial Services Research**, v.13, n. 1, p. 5-22, 1998.

DOSAMANTES, C. A. D. The relevance of using accounting fundamentals in the Mexican stock market. **J. of Econ. Fin. & Administrative Sci.**, v. 18, p. 2-10, 2013.

FAMA E. F. Efficient capital markets: A review of theory and empirical work. **Journal of Finance**, v. 25, p. 383-417, 1970.

FAMA, E. F. Efficient capital markets: II. **J. of Finance**, v. 46, n. 5, p. 1575-1617, 1991.

FARIA, E. L. et al. Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. **Expert Systems with Applications**, v. 36, n. 10, p. 12506-12509, 2009.

FERREIRA, F. G. D. C.; GANDOMI, A. H.; CARDOSO, R. T. N. Artificial Intelligence Applied to Stock Market Trading: A Review. **IEEE Access**. v. 9, p. 30989-30917, 2021.

GU, S.; KELLY, B.; XIU, D. **Empirical asset pricing via machine learning - Technical report**, National Bureau of Economic Research, 2018.

HAN, J.; PEI, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. Amsterdam: Elsevier, 2011.

HIRTLE, B. J.; LOPEZ, J. A. Supervisory Information and the Frequency of Bank Examinations, **FRBNY Economic Policy Review**, p. 1-19, 1999.

KHATRI, S.; ARORA, D.; KUMAR A. Implicating machine learning towards development of intelligent system for wart treatment therapy identification. **International Journal of Engineering & Technology**, v. 7, p. 754-758, 2018

KHAIDEM, L.; SAHA, S.; DEY, S. R. Predicting the direction of stock market prices using random forest. **arXiv preprint arXiv:1605.00003**, 2016.

KHAN, W.; GHAZANFAR, M. A.; AZAM, M. A.; KARAMI, A.; ALYOUBI, K. H.; ALFAKEEH, A. Stock Market Prediction Using Machine Learning Classifiers and Social Media, News. **Journal of Ambient Intelligence and Humanized Computing**, 2020.

KING, R. G.; LEVINE, R. Finance and Growth: Schumpeter Might be Right. **Quarterly Journal of Economics**, v.108, p.717-738, 1993.

KOEHN, M.; SANTOMERO, A. Regulation of Bank Capital and Portfolio Risk. **Journal of Finance**, v. 35, p. 1235-1244, 1980.

KUBAT, M.; BRATKO, I.; MICHALSKI, R.S. A Review of Machine Learning Methods. In: MICHALSKI, R.S.; BRATKO, I.; KUBAT, M. (Org.) **Machine Learning and Data Mining: Methods and Applications**. London: John Wiley & Sons, 1998. p. 3-69.

KUNT, D.; HUIZININGA. Determinants of Commercial Bank Interest Margins and Profitability, The World Bank Development Research Group. **Policy Research Working Paper**, v. 1900, p. 1-35, 1998.

IBGE, 2021. Painel de Indicadores. Disponível em: <https://www.ibge.gov.br/indicadores#desempregoes>. Acesso em: 01/03/2021.

LABOISSIERE, L. A.; FERNANDES, R. A.; LAGE, G. G. Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. **Applied Soft Computing Journal**, v. 35, p. 66-74, 2015.

LAKONISHOK, J.; SHLEIFER, A.; VISHNY, R. 1994, Contrarian investment, extrapolation, and risk. **Journal of Finance**, v. 49, p. 1541-1578, 1994.

LEO, M.; SHARMA, S.; MADDULETY K. Machine Learning in Banking Risk Management: A Literature Review. **Risks**, v. 7, n. 1, 2019.

LEVINE, R.; ZERVOS, S. J. What We have Learned about Policy and Growth from Cross-country Regressions?. **American Economic Review**, v. 83, p. 426-430, 1993.

LEVINE, R.; ZERVOS, S.J. Stock Markets, Banks, and Economic Growth. **American Economic Review**, v. 88, p.537-58, 1998.

LO, A. W.; MACKINLAY, A. C. Stock market prices do not follow random walk: Evidence from a simple specification test. **Rev. of Fin. Studies**, v. 1, p. 41-66, 1988.

MALKIEL, B. G. The efficient market hypothesis and its critics. **Journal of Economic Perspectives**, v. 17, p. 59-82, 2003

MARKOWITZ, H. Portfolio selection. **Journal of Finance**, v. 7, n. 1, p. 77-91, 1952.

MISHKIN F. S. **The Economics of Money, Banking, and Financial Markets**, ed. 6. New York: Addison Wesley Longman, 2001.

MITCHEL, T. M. **The Discipline of Machine Learning**. Pittsburgh: Carnegie Mellon University, 2006.

MOLYNEUX, P.; THORNTON, J. Determinants of European Bank Profitability: A Note. **Journal of Banking and Finance**, v.16, p. 1173-1178, 1992.

NABIPOUR, M. et, al. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data: A Comparative Analysis, **IEEE Access**, v. 8, p. 150199-150212, 2020.

NANDI, J. K. Comparative Performance Analysis of Select Public and Private Sector Banks in India: An Application of CAMEL Model. **Journal of Institute of Public Enterprise**, v. 36, p. 1-29, 2013.

PALIT, S.; ROY, C. S. Stock Market Prediction Using Machine Learning. **Journal of Computer Engineering**, v. 22, n. 4, p. 8-17, 2020.

PATEL, M. B.; S. R. YALAMALLE. Stock Price Prediction Using Artificial Neural Network. **Int. J. of Innovative Research in Science, Engineering and Technology**, v. 3, n. 6, p. 13755-13762, 2014.

PATEL, J; SHAH, S.; THAKKAR, P.; KOTECHA, K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. **Expert Systems with Applications**, v. 42, p. 259-268, 2015.

REDDY, K. S. Stock Market Prediction Using Machine Learning. International Research **Journal of Engineering and Technology**, v. 5, n. 10, p. 1033-1035, 2018.

ROSSI, A. G. Predicting Stock Returns with Machine Learning. **Working Paper**, 2018.

ROZZANI, Nabilah; RAHMAN, Rashidah Abdul. Camels and performance evaluation of banks in Malaysia: conventional versus Islamic. **Journal of Islamic Finance and Business Research**, v. 2, n. 1, p. 36-45, 2013.

SADORSKY, P. A Random Forests Approach to Predicting Clean Energy Stock Prices. **Journal of Risk and Financial Management**, v. 14, n. 2, 2021.

SONI, V. D. Prediction of Stock Market Values using Artificial Intelligence. International **J. of Adv. Res in Elec., Electronics & Instr. Engineering**, v. 7, p. 1844-1848, 2018.

STIGLITZ, J. E.; WEISS, A. Credit Rationing in Markets with Imperfect Information. **American Economic Review**, v. 71, p. 393-410, 1981.

TA-Lib: Technical analysis library (2021). Disponível em: <https://ta-lib.org/>. Acesso em: 01/03/2021.

TODOROVIC, V.; FURTULA, S.; DURKALIC, D. Measuring Performance of the Serbian Banking Sector Using Camels Model. **Journal for Social Sciences**, v. 42, n. 3, p. 961-977, 2018.

VALENCIA, F.; GÓMEZ-ESPINOSA, A.; VALDÉS-AGUIRRE, B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. **Entropy**, v. 21, n. 6, 2019.

VAPNIK, V. An overview of statistical learning theory. **IEEE Transactions on Neural Networks**, v. 10, p. 988-999, 1999.

YEYATI, E. L.; MICCO, A. Concentration and Foreign Penetration in Latin American Banking Sectors: Impact on Competition and Risk. **Inter-American Development Bank - Working Paper**, n. 499, 2003.