

Aplicação de mineração de dados para avaliação de indicadores financeiros

Diogo Tavares Cavalcanti de Moraes ^{1,2}  orcid.org/0000-0001-7524-6206

Gabriela de Vasconcelos ^{1,2}  orcid.org/0000-0002-2632-7242

Guilherme Teixeira Ribeiro de Albuquerque ^{1,3}  orcid.org/0000-0003-0610-5834

Kelly Bezerra da Mota ^{1,3}  orcid.org/0000-0001-6404-6401

Leonardo de Leon Dias ^{1,2}  orcid.org/0000-0003-2336-9858

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

² Pós-graduação em Engenharia da Computação, Escola Politécnica de Pernambuco, Pernambuco, Brasil,

³ Graduação em Engenharia da Computação, Pernambuco, Brasil.

E-mail do autor principal: Leonardo de Leon Dias lld@ecomp.poli.br

Resumo

Entender a movimentação do mercado financeiro e poder identificar tendências de variações com o intuito de prever possíveis alterações de valores de ativos significa elevar a probabilidade de obter lucro, investidores tomam suas decisões avaliando características das empresas e do mercado como um todo e buscam com isso fazer uma prospecção da movimentação do mercado, esta é uma tarefa complexa e muitos pesquisadores, tanto da indústria quanto da academia, constantemente destinam esforços para prever estas tendências no mercado. Para enfrentar o desafio de identificar bons investimentos, normalmente busca-se amparo em dois tipos distintos de abordagem: Abordagem Técnica e Abordagem Fundamentalista. Na Abordagem Fundamentalista, o valor intrínseco de um ativo é o fator determinante e busca avaliar indicadores como a saúde financeira da empresa, suas áreas de investimento, setor de atuação, dados macroeconômicos, assim como diversos outros indicativos que demonstram a possibilidade ou não de determinada empresa elevar seu valor de mercado. A Abordagem Técnica, por sua vez, foca apenas em movimentos anteriores em tabelas de preços, e faz análises gráficas para direcionar suas crenças sobre possíveis comportamentos futuros. Este trabalho buscou avaliar o desempenho de diferentes indicadores técnicos, aplicados individualmente ou em conjunto, a diferentes séries históricas de abertura e fechamento de mercado, e identificar os indicadores que melhor se aplicam a cada uma delas de forma direta ou categorizada, sendo esta categorização feita por período, setor da atuação da empresa ou observando outras características compartilhadas e advindas de análise fundamentalista. Identificou-se grupos de indicadores mais adequados para análise por grupo de empresas e foi possível obter bons insights ao comparar diferentes setores de ativos, assim como comparando outros agrupamentos gerados a partir de dados fundamentalistas. Assim, ao contribuir para a escolha adequada dos indicadores, a pesquisa identifica os melhores indicadores técnicos para cada grupo analisado.

Palavras-Chave: Predição de Série Temporal; Mercado Financeiro; Indicadores Técnicos.

Abstract

Understanding the movement of the financial market and being able to identify trends in variations in order to predict possible changes in asset values means increasing the likelihood of making a profit, investors make their decisions by evaluating characteristics of the companies and the market as a whole and seek to do so. a prospect of market movement, this is a complex task and many researchers, both from industry and academia, constantly devote efforts to predict these trends in the market. To face the challenge of identifying good investments, support is usually sought in two distinct types of approach: Technical Approach and Fundamentalist Approach. In the Fundamentalist Approach, the intrinsic value of an asset is the determining factor and seeks to evaluate indicators such as the company's financial health, its investment areas, sector of activity, macroeconomic data, as well as several other indicators that demonstrate the possibility or not of a certain company to raise its market value. The Technical Approach, in turn, focuses only on previous movements in price lists, and makes graphic analyzes to direct your beliefs about possible future behaviors. This work sought to evaluate the performance of different technical indicators, applied individually or together, to different historical series of market opening and closing prices and to identify the indicators that best apply to each of them, in a direct or categorized way, being this categorization done by period, sector of the company's performance or observing other characteristics shared and arising from fundamentalist analysis. It identified groups of indicators more suitable for analysis by a group of companies and it was possible to obtain good insights when comparing different sectors of assets, as well as comparing other groupings generated from fundamentalist data. Thus, by contributing to the appropriate choice of indicators, the paper identifies the best technical indicators for each group analyzed.

Key-words: Time Series Forecasting; Financial market; Technical Indicator.

1 Introdução

O mercado de ações é um sistema que envolve os processos de compra e venda de ativos financeiros por investidores, cuja dinâmica é bastante complexa [1]. Dentro deste mercado, os investidores têm como objetivo tomar decisões de negócio cuja probabilidade de obter lucro seja a maior possível.

A principal estratégia utilizada para atingir esse objetivo se dá através da Análise Fundamentalista combinada à Análise Técnica dos ativos [1]. Cada uma destas análises podem levar a decisões mais assertivas com vistas ao aumento do lucro para o investidor.

Nesse contexto, a fim de auxiliar os investidores na tomada de decisão no mercado de capitais, vários métodos têm sido propostos pela literatura, entre os quais a utilização de inteligência artificial, o que tem atraído cada vez mais a atenção dos investidores e pesquisadores [4]. Somado a isso, o avanço

tecnológico de processamento de dados, além dos algoritmos de aprendizado de máquina, resultam em esforços para desenvolver modelos de predição quantitativos [5].

Investidores profissionais e empresas de consultoria dentro do mercado financeiro apontam sobre a importância de compreender bem os indicadores dos ativos a serem analisados, visto que, dependendo do setor ou da situação financeira da empresa/organização, por exemplo, os indicadores com maior grau de importância nas análises técnicas e fundamentalistas podem variar [1]. No entanto, observa-se que a maioria dos estudos relacionados a essa problemática, estabelecem a previsão de modelos que usam indicadores financeiros, que podem sofrer com a diminuição do desempenho de predição ao longo do tempo [5] devido à volatilidade do mercado de ações.

De tal modo, surge a questão norteadora desta pesquisa: Como identificar indicadores técnicos que

se adequem melhor para determinado segmento de ativos a fim de otimizar as decisões dos investidores? Diante disso, o objetivo geral deste estudo é identificar um grupo de indicadores técnicos que se apliquem a determinados grupos de ativos capazes de contribuir para decisões de investimentos.

Para tanto, buscar-se-á otimizar o processo de análise de ativos, com base em indicadores financeiros negociados na Bolsa de Valores brasileira, a fim de amparar o processo de decisão dos investidores. Busca-se, por meio da aplicação de técnicas de predição e de mineração de dados e aprendizado de máquina, encontrar quais indicadores técnicos podem ser melhor aplicados a um determinado ativo a partir da observação das suas características.

Desse modo, estabeleceu-se os seguintes objetivos específicos: Identificar os indicadores, utilizados individualmente ou em conjunto, que apresentam melhor desempenho na tarefa; Avaliar a contribuição de diferentes indicadores técnicos, aplicados individualmente ou em conjunto para elevar a performance de um modelo de predição; Comparar a aplicação dos indicadores financeiros em diferentes ativos ou grupos de ativos.

Destarte, analisar e entender bem a dinâmica do mercado de capitais são fatores de extrema importância para o investimento em ações, embora seja desafiador escolher e estudar os diversos indicadores que devem ser considerados de diferentes formas dependendo do setor ou situação dos ativos em questão. Nessa perspectiva, a negociação de curto prazo é uma tarefa difícil devido à flutuação entre demanda e oferta das ações, de modo que observa-se a necessidade de aprimoramento de modelos de previsão a partir da análise fundamentalista e da análise técnica [4].

Na literatura são encontrados uma série de exemplos capazes de demonstrar a importância de se avaliar indicadores e as diretrizes que podem ser obtidas através da correta análise destes. Observou-se através de um estudo nos anos de 2008 e 2009 que a *performance* empresarial da Petrobras foi acima da média mesmo em um contexto pós-crise, onde grandes empresas do segmento foram duramente atingidas. O resultado foi traduzido nas análises de indicadores, em que a liquidez

corrente de 2009 apresentou um aumento de 30% em relação ao ano anterior. Este fato pode ser diretamente atribuído a fatores facilmente identificáveis através de técnicas de mineração de dados [5].

Uma pesquisa realizada pelo Valor Data em 2020 demonstrou que o Ibovespa, principal índice da bolsa de valores brasileira, teve a maior volatilidade anual desde a crise de 2008. A flutuação de preços das ações, no período analisado, oscilou entre 25% e 117%, sendo a Taesa e as ações preferenciais da Gol as representantes dos valores máximo e mínimo do período, respectivamente [2]. Outra pesquisa realizada no mesmo ano demonstra uma alta do volume de negociações da bolsa de 84,6% em relação ao ano anterior, chegando a R\$ 26 bilhões por dia no primeiro semestre de 2020, um novo recorde da bolsa brasileira [3]. Essas pesquisas demonstram a dinamicidade do mercado de ações e refletem a necessidade de possuir ferramentas para embasar as decisões de compra e venda de ativos.

Diante disso, buscou-se comprovar a hipótese de que diferentes indicadores apresentam resultados distintos quando aplicados a ativos com características diversas, este trabalho se justifica por propor contribuições que podem auxiliar na seleção e análise de indicadores financeiros das ações da bolsa de valores brasileira, com o intuito de otimizar o retorno das decisões dos investidores.

Por fim, destaca-se que não está dentro do escopo deste trabalho analisar dados macroeconômicos, políticos, ou realizar análise de sentimento e opiniões para predição de tendências do mercado acionário. Ademais, não pretende-se explorar operações de *day trade* e nem de longo prazo.

2 Fundamentação Teórica

2.1 Métodos de análise do Mercado de Ações

O mercado de ações e de títulos compõem o mercado financeiro, que por sua vez, pode ser definido como o local em que agentes superavitários e deficitários promovem a negociação direta. No caso do mercado de capitais, são negociados títulos de longo prazo, tais como ações e obrigações, ao passo que no mercado monetário, as transações giram em torno de obrigações de curto prazo, ou títulos

negociáveis [6]. Conforme evidenciado no esquema de fluxo de fundos na Figura 1, que apresenta o ambiente transacional do mercado financeiro.

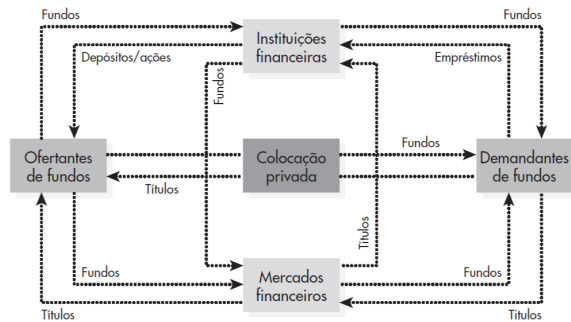


Figura 1: Fluxo de fundos para instituições e mercados financeiros
Fonte: Gitman (2010).

No âmbito do mercado de capitais, a negociação de curto prazo é uma tarefa difícil devido à flutuação entre demanda e oferta das ações, o que reflete nos preços desses ativos. Diante disso, os indicadores técnicos são de grande valia para operações de swing trade [4]. Assim, em termos simplificados, o valor de qualquer ativo é o valor presente de todos os fluxos de caixa futuros que se espera receber dele durante o período em questão [6].

Diante do exposto, observa-se que, embora prever o preço das ações em curto prazo seja uma tarefa desafiadora devido à natureza volátil do mercado de ações, essa questão tem atraído pesquisadores e acadêmicos para aprimorar modelos de previsão. Investidores baseiam suas negociações em diferentes formas de análise do mercado, de tal modo que existem várias formas de análise de ações que podem ser realizadas, como análise técnica e fundamentalistas que respectivamente destinam-se à avaliação de gráficos de histórico de preços e à avaliação da saúde financeira das empresas [31].

A análise fundamentalista consiste em avaliar o preço das ações de uma empresa, com base em seus dados históricos, debruçando-se sobre qualquer fator que possa afetar os preços de suas ações, incluindo os lucros, condições macroeconômicas e da economia de modo geral, condições da indústria e da gestão da empresa [4]. Na análise fundamentalista são avaliados indicadores como: saúde financeira da empresa, áreas de investimento, setor de atuação, e diversos outros que demonstram a possibilidade ou não de determinada empresa elevar seu valor de mercado.

Por outro lado, a análise técnica observa a movimentação anterior dos valores de um ativo em questão, principalmente através da observação de gráficos, com o intuito de obter indícios sobre possíveis comportamentos futuros daquele ativo, e direcionar a tomada de decisão baseando-se nisso [1]. A análise técnica é um método de mensuração de ações, que envolve análises estatísticas de dados do mercado, tais como o histórico de preços de um determinado ativo contendo preços de abertura, fechamento, volume, preço mais alto e preço mais baixo do dia [32].

2.1.1 Performance de diferentes indicadores aplicados ao Mercado de Ações

O princípio básico subjacente à análise técnica é que o preço de mercado de um ativo reflete todas as informações disponíveis que podem impactar o em sua precificação, analistas técnicos geralmente acreditam que os preços mudam de acordo com as tendências e a observação das séries históricas pode permitir identificar essas tendências.

Os indicadores técnicos são uma forma estatística de análise técnica onde se aplicam várias fórmulas matemáticas a preços e volumes. Existe uma diversidade de indicadores técnicos, e com diferentes níveis de complexidade, um dos mais comuns são as médias móveis, que suavizam os dados de preços para facilitar a localização de tendências, já dentre os mais complexos encontra-se a divergência e convergência da média móvel (MACD) , que analisa a interação entre várias médias móveis [7].

Frequentemente analistas técnicos utilizam vários indicadores para montar suas estratégias visto que cada indicador possui a capacidade de expressar um determinado tipo de informação sobre as séries históricas, com isso é possível estabelecer algumas classes de indicadores de acordo com a informação que o mesmo permite observar, a partir de quatro classes de indicadores, conforme disposto no Quadro 1 [7].

Quadro 1: Distribuição de indicadores por classes.

Momentum	Trend	Volatility	Chart studies
Stochastic	ADX	Bollinger Bands®	Horizontal lines
RSI	Moving averages	Standard deviation	Fibonacci
CCI	ATR		Supply / Demand

<i>Williams %</i>	<i>MACD</i>	<i>Keltner Channel</i>	<i>Trend lines</i>
MACD	Parabolic SAR	Envelopes	
	Bollinger Bands®		
Ichimoku Cloud			

Fonte: SCHOLTMANN e CZUBATINSKI (2019).

Os indicadores de momento são frequentemente conhecidos como osciladores, pois oscilam entre os limites superior e inferior definidos, ajudam a analisar a proporção de compradores e vendedores para entender qual grupo de participantes do mercado tem a maioria e com que força os estão empurrando em uma direção ou se o preço está perdendo força; já os indicadores de tendência analisam uma tendência prevalecente, geralmente não são eficazes quando o preço flutua lateralmente; indicadores de volatilidade verificam o grau de volatilidade e até que ponto o preço oscila, apresentam informações necessárias para definir paradas, definir metas e determinar o tamanho da posição, já os indicadores baseados em estudo de gráficos oferecem uma forma visual de interpretar tendências [7].

Alguns trabalhos relacionados apresentam indicadores técnicos e fundamentalistas aplicados a diferentes ativos e permitem verificar a variação de suas performances em diferentes ambientes. Um estudo realizado com 33 combinações entre diferentes indicadores técnicos aplicados a 2 ativos identificou um comportamento distinto em ambos os ambientes, o que comprova que é possível melhorar um modelo de predição para um determinado ativo escolhendo um conjunto de indicadores que melhor se adeque ao papel analisado [4].

Por sua vez, em uma comparação de indicadores técnicos com indicadores fundamentalistas, outro estudo observou que variáveis fundamentalistas geram previsões mais precisas quando a economia está em recessão enquanto que as variáveis técnicas têm um desempenho melhor quando a economia está em expansão, o que mostra que os dois tipos de variáveis fornecem informações complementares ao longo do ciclo de negócios e que combinando informações econômicas e técnicas é possível fazer previsões mais confiáveis do que combinando

qualquer um dos tipos de informações isoladamente [8].

2.2 Técnicas de Mineração de Dados

Nos últimos anos uma série de trabalhos relevantes foram propostos buscando prever os movimentos de mercado a partir de abordagens em diferentes ramos de pesquisa, como estatística, uso de técnicas de reconhecimento de padrões, análise de sentimentos, aprendizado supervisionado e não supervisionado [33]. Dentre as inúmeras técnicas de mineração utilizadas no estudo do mercado financeiro, nesta pesquisa destaca-se a clusterização e a regressão, as quais foram empregadas neste trabalho.

Os problemas de clusterização já são bastante estudados na literatura, na computação, este tema foi impulsionado pela popularização do conceito de mineração de dados [15]. A clusterização é utilizada para separar os registros de uma base de dados em subconjuntos ou clusters, de tal forma que os seus elementos compartilhem propriedades comuns, representando uma configuração em que cada elemento possua uma maior similaridade com qualquer elemento do mesmo cluster em relação a elementos de outros clusters, isto é, servindo para distinguir os elementos em outros clusters, tendo como objetivo maximizar similaridade intra-cluster e minimizar similaridade inter-cluster [13].

Um dos maiores benefícios da utilização dos modelos de clusterização para realização de agrupamentos (divisão em clusters) e sua capacidade de aplicação não supervisionada, fazendo com que eles funcionem bem quando não se sabe ao certo as etiquetas das classes desejadas, ou seja, quando se é desconhecida a característica comum responsável pela atribuição e verificação do grau de similaridade entre os objetos. Isso permite a realização de análises aprofundadas mesmo sem o desenvolvimento de conhecimento prévio sobre o tema estudado [16].

Um método bastante utilizado quando não se sabe as classes a qual se deseja encontrar é o DBSCAN, muito presente em inúmeras pesquisas por ser bastante eficiente e por possuir uma simples aplicabilidade. Este método baseia-se em um conceito de densidade focado no centro do agrupamento, que leva em consideração o raio de alcance dos objetos. Ele deve ser calibrado buscando o equilíbrio de um dos seus principais parâmetros, o raio em relação aos objetos, que permanecendo nessa zona pré-determinada, podem ser classificados em: Interiores ou Centrais; Contidos em Grupos ou Baseados na Densidade; Limítrofes, que não são

centrais mas podem ser alcançados por pelo menos um objeto central e; Ruídos, que não são centrais e também não podem ser alcançados [19].

Já a técnica de regressão tem como principal objetivo aproximar uma função que representa os dados inseridos no modelo, permitindo que dado um conjunto de dados iniciais seja possível prever valores futuros. Um dos modelos de regressão bastante utilizados são as redes neurais, dentre elas o MLP (*Multilayer Perceptron*) *destaca-se tanto por ser uma das primeiras redes neurais concebidas, como pela sua fácil implementação.*

Por ser uma rede neural, naturalmente a informação se encontra nos pesos sinápticos e, diante disso, os sinais de entrada são transmitidos de camada em camada até chegar na saída, chamada “sinal de saída” [21]. Uma forte limitação, entretanto, desses modelos é que há grande dificuldade no tocante às questões de interpretação do que está ocorrendo internamente na rede, onde as etapas do que acontece no código são de difícil interpretação. Na literatura existem alguns exemplos de utilização de MLP em diversos contextos, incluindo séries temporais não lineares.

2.3. Trabalhos Relacionados

O presente estudo levou a cabo uma revisão de literatura para identificar o estado da arte da temática relativa à aplicação de mineração de dados no mercado financeiro de ações, de modo que, considerando os últimos 5 anos observou-se um total de 158 papers em diversas bases de dados como ACM, Scopus, Capes, IEEE, Springer e Village, destes, após a análise do abstract, chegou-se a alguns trabalhos relevantes relacionados com a análise do mercado de ações por meio de técnicas de mineração de dados.

Muitos destes estudos tentaram prever o desempenho corporativo e os preços das ações para aumentar a lucratividade do investimento usando abordagens qualitativas. No entanto, o desenvolvimento da tecnologia de processamento de dados e algoritmos de aprendizagem de máquina resultou em esforços para desenvolver modelos de previsão quantitativa em várias áreas, inclusive no mercado de ações, neste sentido, vários estudos propõem um modelo quantitativo de predição aplicando indicadores financeiros e indicadores técnicos para previsão lastreado em mineração de dados [9].

Na literatura encontra-se uma série de outros exemplos capazes de demonstrar a pertinência dos estudos de inteligência artificial voltados para o mercado de ações. Sendo assim, é imprescindível, antes de tudo, possuir ciência da importância da utilização dos dados gerados pelas empresas para a

identificação de comportamentos comuns e também para a predição de movimentos de mercado. Uma pesquisa investigou a associação entre os indicadores de desempenho ROA (*return of assets*) e ROE (*return of equity*) com a distribuição de riquezas de acordo com o exposto nas Demonstrações de Valor Adicionado (DVA) das empresas listadas na B3 [10]. Há também um trabalho focado na utilização da mineração de opinião para predição de aspectos relacionados à bolsa de valores [11]. Neste último, algo muito interessante é que o autor utilizou comentários publicados na rede social *Twitter* para compor uma base de dados que serviu de insumo para utilização de técnicas como processamento de linguagem natural (PLN) e Máquinas de Vetor de Suporte (SVM) para predizer comportamentos de ativos, o que figura uma abordagem um pouco mais incomum de ser encontrada, mas também bastante interessante.

Outro estudo debruçou-se na identificação de uma análise comparativa focada na observação do desempenho de indicadores econômicos e financeiros de demonstrações contábeis da Vale S/A entre os anos de 2014 e 2018 [12]. O trabalho levou em consideração os dados de liquidez, rentabilidade e endividamento e contribuiu com a geração de importantes evidências diretamente relacionadas com a adaptação das empresas aos últimos pronunciamentos contábeis no período de 2020 e também conseguiu estabelecer uma relação lógica do desempenho organizacional com a adesão às normas internacionais de contabilidade.

Mais recentemente, pesquisas foram desenvolvidas no contexto da mineração de dados do mercado financeiro de ações, utilizando redes neurais para pré-processamento de preços originais, sugerindo que os modelos profundos testados têm um desempenho melhor do que os modelos mais simples e podem aprender com precisão uma variedade de estratégias de negociação [26]. Outro estudo investigou as regras técnicas que são mais eficazes em mercados em baixa do que em mercados em alta por meio de auto regressão vetorial e utilização de regras técnicas como média móvel simples, média móvel dupla e quebra de faixa de negociação, bem como método bootstrap para inferências estatísticas [27].

Realizou-se também a previsão de movimento direcional de dados Forex usando LSTM com base de dados que serviu de insumo para utilização de técnicas como processamento de linguagem natural (PLN) e Máquinas de Vetor de Suporte (SVM) para predizer comportamentos de ativos, o que figura uma abordagem um pouco mais incomum de ser encontrada, mas também bastante interessante [28]. Essa pesquisa, utilizou-se da ferramenta popular de aprendizado profundo, que se mostrou

muito eficaz em muitos problemas de previsão de séries temporais, para fazer previsões de direção no Forex, utilizando dois conjuntos de dados diferentes (macroeconômicos e indicadores técnicos), propondo assim, um modelo híbrido de previsão de direção no Forex.

Por sua vez, em outra pesquisa buscou-se prever o comportamento flutuante do mercado de ações a partir da análise dos sentimentos dos investidores considerando as eleições presidenciais dos EUA e os efeitos da Covid 19 como um fator de flutuação explícito que afeta o desempenho do mercado de ações [29]. Além disso, buscou identificar regras de e pontos de negociação corretos e melhores indicadores técnicos a serem considerados por meio de formulações matemáticas para determinar quando comprar ou vender ações. Para isso, adotou-se um modelo de previsão de ações usando o modelo log bilinear (LBL) para aprendizagem de padrão de sentimento do mercado de ações de curto prazo e rede neural recorrente (RNN).

Por fim, dentre os vários estudos realizados no âmbito do mercado de ações, chamou atenção, o estudo realizado sobre a previsão da volatilidade do mercado de ações usando uma ampla gama de indicadores técnicos construídos com base no comportamento anterior do preço das ações, volatilidade e volume de negócios, por meio da técnica de auto regressão, o qual identificou que a incorporação de variáveis técnicas no benchmark de auto regressão pode produzir previsões de volatilidade significativamente mais precisas [30]. O desempenho de previsão da combinação de indicadores técnicos é ainda comparada com o dos indicadores econômicos populares. As variáveis técnicas têm um desempenho melhor do que as variáveis econômicas quando a economia está em expansão, enquanto as variáveis econômicas geram previsões mais precisas quando a economia está em recessão. Esses dois tipos de variáveis fornecem informações complementares ao longo do ciclo de negócios.

Dado o exposto, fica evidente que existem uma série de técnicas de mineração de dados capazes de contribuir de maneira significativa para o mercado financeiro como um todo, não excetuando o mercado de ações. Essa constatação prévia forneceu importantes insumos para o presente trabalho e permitiu uma ampliação do horizonte de ideias capazes de agregar valor à execução dos processos descritos a seguir e consequentemente ao resultado final da pesquisa.

Assim, pode-se observar que dentre os estudos mais recentes acerca da análise do mercado de ações, é recorrente a utilização de indicadores técnicos e fundamentalistas aliados a aplicação de

técnicas de mineração de dados e inteligência artificial, o que tem gerado contribuições para uma maior acurácia na predição do mercado acionário, no entanto, ainda existem recortes temáticos que podem ser melhor explorados, tal qual o adotado na presente pesquisa.

3 Materiais e Métodos

O enquadramento metodológico adotado neste estudo caracteriza-se pelo seu objetivo exploratório-descritivo, de natureza aplicada, mediante a aplicação de procedimentos metodológicos de uma pesquisa de levantamento de dados primários e secundários, e por meio de uma abordagem quantitativa.

Por ser um estudo exploratório-descritivo, tipologia de pesquisa usualmente utilizada em pesquisas aplicadas [30] com o objetivo de aprofundar-se a temática estudada, a fim de levantar as variáveis do mercado financeiro pertinentes para avaliação de ativos financeiros, e descrever a relação entre os indicadores analisados e seu impacto na predição do valor de ações.

Nessa perspectiva, classifica-se como um estudo pragmático, produzindo conhecimentos úteis para solucionar problemas reais [22], que nesse caso, diz respeito a otimização das análises dos papéis por meio de indicadores que possam predizer o valor das ações por segmento de atuação da empresa, a fim de associar a um modelo de predição com vistas a melhoria de performance.

Ademais, o procedimento adotado para consecução da pesquisa se deu por meio da extração de dados quantitativos primários e secundários do *Yahoo Finance* no tocante às séries históricas de indicadores técnicos; e do site *Fundamentus*, quanto aos dados de indicadores fundamentalistas de um grupo de empresas que compõem a Bolsa de Valores Brasileira - B3. Sendo assim, a abordagem adotada foi a quantitativa, pois o estudo realizado se utiliza amplamente de técnicas estatísticas para pré-processamento de dados, bem como para sua análise por meio de clusterização e regressão [23] a fim de proceder a mineração dos mesmos, por meio do emprego de simulações, provas e testes do modelo proposto.

3.1 Descrição da Base de Dados

A base de dados utilizada foi construída a partir da importação de dados históricos de abertura e fechamento de mercado de ações negociados no mercado de capitais da B3 (Bolsa de Valores, Mercadorias & Futuros brasileiro), do *Yahoo Finance*

em 03 de março de 2021 e da importação de dados de análise fundamentalista do site Fundamentus. Para o armazenamento dos dados foi criado um banco de dados relacional, MySQL, na Amazon RDS (Relational Database Service) serviço de nuvem da Amazon para armazenamento de bancos de dados relacionais.

Os dados históricos de abertura e fechamento de mercado são compostos por 378 tabelas que representam cada uma das empresas listadas no mercado de ações brasileiro, com histórico de negociações há pelo menos 5 anos, cada uma com 8 colunas contendo informações da variação no valor do seu ativo em um intervalo de 1 dia onde cada linha representa 1 dia de negociação. Já os dados históricos de análise fundamentalista compõem uma tabela de 378 linhas e 21 colunas, sendo as linhas referentes às empresas escolhidas para importação de seu histórico de preço e as colunas compostas por indicadores fundamentalistas.

Diante do exposto, vale salientar que a base de dados utilizada no presente trabalho trata-se de um data warehouse devido à organização e sequência lógica utilizada no armazenamento.

Os dados apresentados estão limpos, tratados e agrupados de modo a facilitar a utilização na pesquisa. Devido a isso, portanto, pode-se afirmar que não se trata de uma base de dados do tipo data lake, uma vez que, para caracterizar tal estrutura, seria necessário que os dados fossem apresentados de maneira bruta, sem todos os cuidados tomados.

Nos quadros 2 e 3, ambos elaborados através de dados da pesquisa, observa-se uma organização capaz de caracterizá-los com data warehouse, com as informações bem segmentadas seguindo critérios preestabelecidos.

Quadro 2: Dicionário de dados da Tabela de Abertura e Fechamento do Mercado.

Atributo	Classe	Domínio	Descrição
Date	Simples	Data	Dia de obtenção dos dados
Open	Simples	número decimal	Preço de abertura
High	Simples	número decimal	Preço mais alto registrado no dia
Low	Simples	número decimal	Preço mais baixo registrado no dia
Close	Simples	número decimal	Preço de fechamento
Adj Close	Simples	número decimal	Preço de fechamento ajustado
Volume	Simples	inteiro grande	Volume de operações realizadas no dia

Fonte: Dados da pesquisa.

Quadro 3: Dicionário de dados da Tabela de análise fundamentalista.

Atributo	Classe	Domínio	Descrição
Papel	Determinante	texto	Papel
Setor	Composto	texto	Setor
Subsetor	Composto	texto	Subsetor
Cotação	Simples	texto	Cotação
Valor_de_Mercado	Simples	inteiro grande	Valor de mercado
PL	Simples	texto	Preço sobre lucro
PVP	Simples	texto	Preço sobre valor patrimonial
PEBIT	Simples	texto	Ganhos proprietários antes de interesses e taxas
Div_Yield	Simples	texto	Rendimento de Dividendo
EV_EBITDA	Simples	texto	Relação entre o valor da companhia e o EBITDA de uma empresa
EV_EBIT	Simples	texto	Comparação do valor da empresa com o lucro antes da incidência de taxas e impostos
Cres_Rec_5a	Simples	texto	Crescimento da Receita Líquida nos últimos 5 anos
LPA	Simples	texto	Lucro por ação
VPA	Simples	texto	Valor Patrimonial da ação
Marg_Bruta	Simples	texto	Margem do lucro bruto
Marg_Líquida	Simples	texto	Margem líquida
ROIC	Simples	texto	Retorno sobre o capital investido
ROE	Simples	texto	Retorno sobre o patrimônio líquido
Div_Br_Patrim	Simples	texto	Dívida Bruta total(dívida+debêntures) dividido pelo patrimônio Líquido.
Div_Bruta	Simples	texto	Dívida Bruta
Div_Líquida	Simples	texto	Dívida Líquida

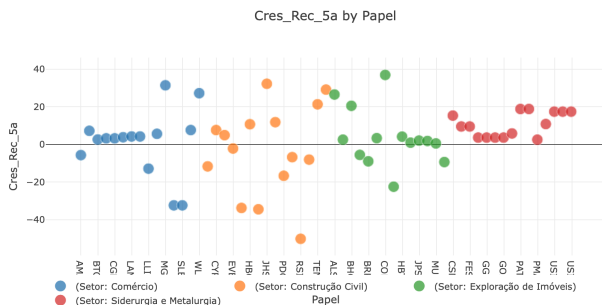
Fonte: Dados da pesquisa (2021).

3.2 Análise Descritiva dos Dados

Com relação aos dados de análise fundamentalista utilizados no projeto verifica-se uma grande variação entre cada um dos indicadores quando observado diferentes empresas. Empresas de diferentes setores podem apresentar características bastante distintas, no entanto se uma empresa for observada isoladamente pode divergir bastante do setor ao qual está inserida. Em uma análise descritiva dos dados, observando a figura 2, que apresenta um gráfico de dispersão relacionando

o crescimento da receita entre 4 setores de atuação distintos, pode-se perceber por exemplo a grande variação no indicativo de crescimento da receita em um período de 5 anos entre os setores de Siderurgia e Metalurgia e o setor de Construção Civil, que sob uma óptica inicial deveriam apresentar uma maior relação.

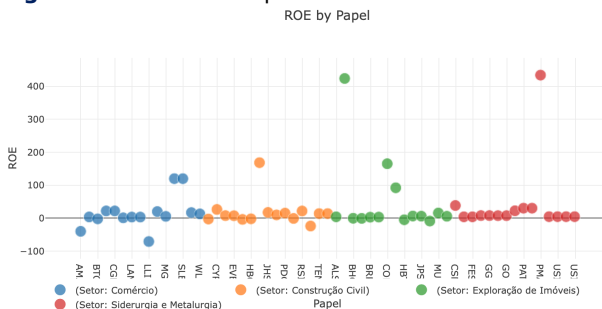
Figura 2: Gráfico de dispersão Crescimento da receita em 5 anos.



Fonte: Dados da pesquisa (2021).

Observando a figura 3 que apresenta a relação do valor do ROE entre diferentes setores, é possível observar uma boa relação entre os setores mencionados anteriormente.

Figura 2: Gráfico de dispersão do ROE

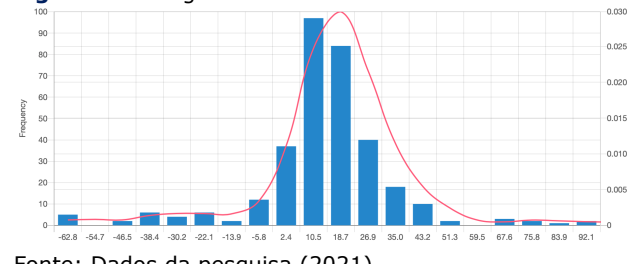


Fonte: Dados da pesquisa (2021).

Observando o histograma de frequência do ROE na figura 3 para os ativos presentes na base de dados utilizada, após a aplicação de um filtro de outliers que removeu valores menores que -100 e maiores de 100 percebe-se que existe uma concentração entre 10 e 20 evidenciando uma distribuição dos dados mais à direita do ponto zero,

ou seja, a maioria dos valores do ROE dos ativos presentes na base apresenta valores positivos.

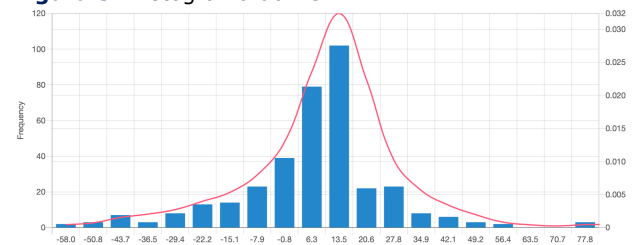
Figura 3: Histograma do ROE.



Fonte: Dados da pesquisa (2021).

Já se observarmos o histograma de frequência da do Crescimento da Receita em 5 anos, figura 4, podemos concluir que a maioria dos ativos da base possuem valores positivos e sua maior concentração encontra-se no intervalo entre 6 e 13.

Figura 3: Histograma do ROE.



Fonte: Dados da pesquisa (2021).

Inúmeras constatações interessantes podem ser feitas a partir do de uma análise descritiva de cada uma das colunas presentes na base, é possível perceber ainda a existência de outliers em cada uma delas, no entanto pela característica do mercado financeiro e o que se espera das observações aqui realizadas, optou-se pela não remoção de nenhum dos outliers observados, esta decisão foi baseada na orientação de alguns profissionais da área de investimento que colaboraram para o desenvolvimento deste projeto. O principal suporte a esta decisão recai sobre expectativa de que a busca por encontrar clusters de ativos com alguma semelhança possa oferecer insights interessantes até mesmo com relação aos outliers, outro embasamento para esta decisão recai sobre o fato de que a remoção dos outliers de todos as colunas de indicadores fundamentalistas resultaria em uma limitação muito grande da base de ativos analisados.

3.3 Pré-processamento dos Dados

Para acesso e manipulação da base foi utilizado um notebook na plataforma Google Collaboratory.

Inicialmente foram tratados os valores faltantes representados com hífen “-” em formato tipo TEXT, no qual todos os valores com esta representação foram convertidos para NULL.

Em seguida, a grande maioria dos dados numéricos que encontravam-se também em formato de texto foram convertidos para o tipo DOUBLE.

Por fim os dados foram normalizados e os dados faltantes, cuja grande maioria se referem a indicadores de dívidas, foram preenchidos com o valor zero.

Para realização da clusterização dos dados utilizando a tabela de indicadores fundamentalistas, convergindo com a pesquisa realizada com os stakeholders, utilizamos as colunas de indicadores mais relevantes para realizar a clusterização dos papéis, realizando uma redução vertical dos dados da tabela em questão.

Todo o procedimento para tratamento dos dados foi realizado utilizando Python.

3.4 Metodologia Experimental

3.4.1 Etapa de Clusterização

O desenvolvimento do projeto ocorreu com base no método *CRoss Industry Standard para Data Mining* (CRISP-DM), o qual consiste em um modelo de processo com seis fases que permite planejar, organizar e implementar um projeto de ciência de dados que possui uma característica cíclica passando pelas etapas de compreensão do negócio, compreensão de dados, identificação dos dados disponíveis e necessários, pré processamento e preparação de dados, organização dos dados para modelagem, definição das técnicas de modelagem, definição do modelo que atende melhor aos objetivos de negócios e por fim a apresentação e avaliação dos resultados.

A primeira etapa do projeto teve como objetivo criar grupos de ativos negociados no mercado de ações brasileiro que possuam características em comum, que não sejam já conhecidas, como por exemplo o setor de atuação. Em razão disto, foi decidido utilizar a técnica de clusterização denominada DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*).

A justificativa para escolha deste método recai sobre as características distintas de cada uma das

colunas dos dados, onde há uma significativa diferença entre cada um dos indicadores. Além disso, também optou-se por não definir um número de clusters previamente, visto que não há nenhuma referência ao quantitativo de grupos que melhor representarão os dados.

Por fim, a técnica DBSCAN foi escolhida também porque, além do exposto, outro benefício esperado recai sobre sua capacidade de encontrar ruídos nos dados, que neste caso podem surgir por se ter optado pela não remoção de outliers, devido a sua característica bastante distinta, a remoção de outliers acabaria por reduzir grandemente as opções de investimento.

Durante os testes iniciais foi observado que a utilização completa do conjunto de dados resultava em uma baixa qualidade da clusterização, por isso a quantidade de indicadores fundamentalista foi reduzida em torno de 60%, onde optou-se por manter na base os principais indicadores orientados pelos stakeholders consultados durante o projeto. Para o processo de clusterização, além do algoritmo DBSCAN, foram utilizados também os algoritmos PCA (*Principal Component Analysis*) e t-SNE (*t-Distributed Stochastic Neighbor Embedding*) da biblioteca skit-learn.

Após o pré-processamento dos dados, foram avaliados alguns métodos na busca pelo alcance do resultado esperado na pesquisa, dentre eles, aplicamos o algoritmo PCA para compactar o conjunto de dados em um subespaço de recurso de dimensão inferior mantendo a maioria das informações relevantes. Foram realizados testes com e sem a utilização do PCA e devido a melhoria da qualidade dos resultados optou-se por manter o seu uso. O resultado adquirido no PCA foi utilizado para alimentar o algoritmo t-SNE, que serviu ao propósito de fazer um mapeamento dimensional dos dados.

Cabe a ressalva que mesmo o conjunto de dados não sendo de uma dimensão elevada, a utilização dos dois algoritmos de redução de dimensionalidade conjuntamente resultou em uma clusterização mais adequada às expectativas.

Após as etapas citadas o DBSCAN foi executado diversas vezes, variando o número mínimo de amostras e distância máxima, para realizar a observação dos resultados com diferentes parâmetros. Foram executados ainda testes com a variação dos parâmetros de métrica, taxa de aprendizado e número de iterações para o algoritmo t-SNE.

Na tabela 5 estão explicitados os parâmetros utilizados nos algoritmos para realizar a clusterização.

Tabela 5: Parâmetros utilizados para o primeiro resultado do PCA/t-SNE/DBSCAN

Parâmetros da clusterização	
Número mínimo de amostras	10
Distância Máxima	3
Métrica	Euclidiana
Taxa de aprendizado	368
N de iterações	2500

Fonte: Dados da pesquisa (2021).

3.4.2 Etapa de Regressão

Em etapas iniciais deste trabalho foram aplicados métodos de regressão, MLP e Regressão Logística como classificadores, para identificar uma tendência de valorização ou desvalorização dos ativos analisados, para isso foram utilizados as séries históricas de preços com o incremento dos indicadores técnicos, foram executados testes variando o número de classe entre duas a seis, onde as classes representaram desde um movimento binário de valorização ou desvalorização, até um range de variações formadas por percentuais de valorização ou desvalorização diários, no entanto devido à característica bastante limitada resultante da modelagem da tarefa de predição em encontrar apenas classes, optou-se por utilizar uma rede neural MLP, como regressor para fazer predição das séries de preço dos ativos analisados, e avaliar os melhores indicadores com esta segunda técnica testada.

Foram realizadas previsões de curto prazo, com o intuito de prever o dia posterior, dado uma janela temporal de dez dias, utilizando como dados de entrada o histórico de preços dos ativos analisados. As séries históricas constituídas inicialmente de seis colunas, foram expandidas a partir da inserção de diversos indicadores técnicos, para complementar a base de dados e principalmente verificar a possibilidade de obter algum incremento de performance no modelo de predição.

O intuito desta etapa recai exatamente sobre o objetivo principal deste trabalho que se destina a identificar indicadores técnicos que melhor se

aplicam a grupos de ativos analisados. Para verificação da influência de cada indicador em uma dada execução utilizou-se o algoritmo de seleção de features, Random Forest Regressor, disponível em com uma biblioteca do Keras, para verificar qual ou quais dos indicadores inseridos na base realmente oferecem uma melhor contribuição para o modelo.

Durante as etapas iniciais os parâmetros do MLP foram ajustados com valores orientados pela expertise da equipe e posteriormente treinados por meio de uma ferramenta de *tuning* denominada *Grid Search*. Verificou-se que com os parâmetros selecionados manualmente os resultados foram superiores aos resultados obtidos pelo tuning realizado pela ferramenta *Grid Search*, por isso os resultados finais apresentados foram baseados na configuração manual da MLP utilizada.

O código foi desenvolvido na plataforma COLAB, utilizando python e as bibliotecas pandas, scikit-learn, technical-analyst-lib, keras, orange, numpy, tensor flow e matplotlib.

Em relação aos dados para treinamento e teste foram utilizadas 21 colunas contendo valores de preço de abertura, fechamento, máximo, mínimo, ajuste de fechamento e volume negociado. A partir deste histórico, calculamos os seguintes indicadores técnicos: SMA e EMA para 10, 20, 50, 100 e 200 períodos, ATR, CCI, ROC, RSI e WCL. Para o cálculo dos indicadores escolhidos foi utilizada a biblioteca denominada Technical Indicators Lib.

Após o cálculo dos indicadores técnicos, foi realizada a normalização dos dados, assim como o tratamento dos valores faltantes oriundos do próprio cálculo dos indicadores, que basicamente representaram a exclusão dos períodos iniciais onde não foi possível calcular tais indicadores.

Em seguida, realizou-se a divisão dos dados em treinamento e teste, com valores de 70% e 30% respectivamente. Posteriormente foram feitos os ajustes necessários para implementação de um modelo para predição da série temporal com múltiplos parâmetros de entrada, tendo como saída a previsão do valor de fechamento do mercado de um período de 1 dia.

Para o ajuste manual do MLP foi utilizado uma camada de entrada que variou durante os testes entre vinte a 200 neurônios, uma camada escondida com 300 neurônios com função de ativação RELU totalmente conectada, por fim foi compilado utilizando o otimizador ADAM, e MSE (*Mean squared*

error) como métrica de avaliação. O tamanho do Batch foi mantido com o valor *default* 32, e a taxa de aprendizagem também foi mantida a *default* 0,001. O modelo foi testado para diferentes épocas que variaram entre 300 e 2000.

Tabela 3: Parâmetros MLP - Ajuste manual.

Parâmetros do MLP	
Camadas de entrada	n x 10
Camada oculta	300
Função de ativação	RELU
Camada de saída	1
Otimizador	Adam
Épocas	500
Taxa de aprendizado	0,001
Métrica de avaliação	MSE
Batch	32

Legenda: "n", número de features (indicadores técnicos + dados das séries históricas). Fonte: Dados da pesquisa (2021).

Para atingir o objetivo desejado, primeiramente foram executadas duas rodadas de treinamento e teste para cada um dos 378 papéis analisados, utilizando como input todos os dados existentes, neste momento para cada rodada de treinamento foi armazenado o RMSE como métrica de avaliação do modelo, e ainda foram verificados quais as features de maior importância. Em um segundo momento, a base de dados inicial foi alterada e manteve-se apenas os indicadores de maior relevância observados na primeira etapa.

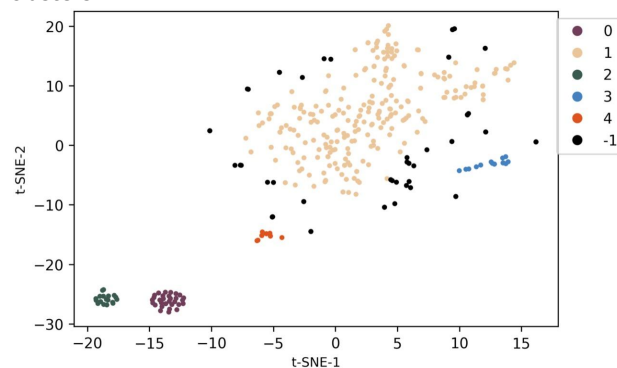
Neste segundo momento manteve-se 25% dos indicadores e novamente foram executadas 2 rodadas de treinamento e teste, no entanto armazenando apenas o RMSE (*Root Mean squared error*) como métrica de avaliação. A base de dados gerada nesta etapa contendo os valores das métricas de avaliação e os indicadores de maior relevância para cada empresa analisada foi associada à tabela dos ativos analisados e a seus respectivos setores de atuação. Foi ainda inserido nesta tabela a classe de cada ativo, obtida na etapa de clusterização. Com estes dados gerados foi possível obter uma série de informações relevantes relacionadas ao objetivo proposto.

4 Análise e Discussão dos Resultados

Primeiramente, foram obtidos os resultados da etapa de clusterização dos ativos financeiros a partir dos seus indicadores fundamentalistas.

Nesta etapa, foi encontrado o resultado final de seis *clusters*, distribuídos conforme mostra a figura 5.

Figura 5: Gráfico do resultado do DBSCAN com 6 clusters.



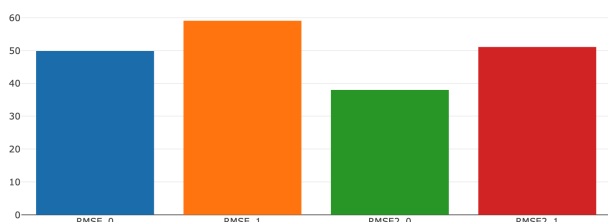
Legenda: Neste gráfico, os clusters 0, 1, 2, 3, 4 e -1 foram chamados de A, B, C, D, E e R respectivamente. Fonte: Dados da pesquisa (2021).

Diante deste resultado, foi realizada uma análise comparativa com os setores dos ativos presentes em cada cluster. Nesta análise, foi observado, por exemplo, que praticamente todos os papéis contidos no grupo A são do setor de Bancos, com exceção de um único papel do setor de Transporte, o VSPT3. Neste caso, o referido ativo pode ser identificado como outlier em seu setor, por conter características em seus indicadores fundamentalistas que se assemelham aos demais papéis contidos neste mesmo cluster. Fazer uma análise detalhada dos ativos presentes na base de dados utilizada foge do escopo deste trabalho, mas trazemos algumas referências ao ativo em [24] e [25], onde podem ser observadas algumas informações sobre ele. Foi observado também que todos os papéis do setor Agropecuário foram contidos no grupo B, assim como todos os demais papéis do setor de Transporte. Em relação aos diversos outros setores, não foi feita uma análise mais aprofundada para encontrar outras possíveis peculiaridades.

Para identificar as relações existentes entre os diferentes indicadores técnicos utilizados e os papéis analisados, observou-se tanto os setores de atuação das empresas quanto as classes obtidas na etapa de clusterização. A unificação destas informações, com

a relação dos indicadores mais influentes na avaliação de cada papel e o RMSE obtido, permitiu validar a hipótese de que existem indicadores técnicos que se aplicam melhor a um determinado grupo de ativos. Permitiu-se observar ainda, que a utilização de um grupo seleto de indicadores mais influentes por papel, oferece benefício à tarefa de predição quando comparado à utilização de todos os indicadores no treinamento do modelo.

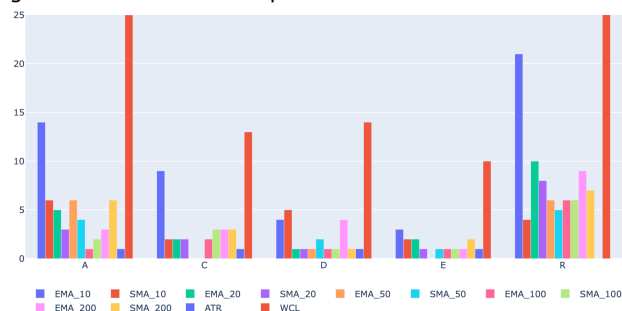
Figura 10: Gráfico da média do RMSE por ciclo de treinamento.



Fonte: Dados da pesquisa (2021).

A figura 10 demonstra em barras a média do RMSE por rodada de treinamento e teste e evidencia que selecionando um grupo de indicadores mais adequados à avaliação de uma determinada empresa, o resultado do modelo de predição tem um incremento de performance. Pode-se observar que a média do RMSE_0, referente a avaliação do erro no primeiro ciclo da rodada de treinamento com todos os indicadores, é superior ao RMSE2_0, que representa a média dos erros da primeira rodada de treinamento com apenas 25% dos indicadores, selecionados com relação a sua maior influência na análise do papel em questão. Com relação às métricas da segunda rodada de treinamento, o mesmo pode ser observado, relacionando o RMSE_1 e o RMSE2_1.

Figura 11: Gráfico do resultado dos indicadores com maior grau de importância nos clusters.

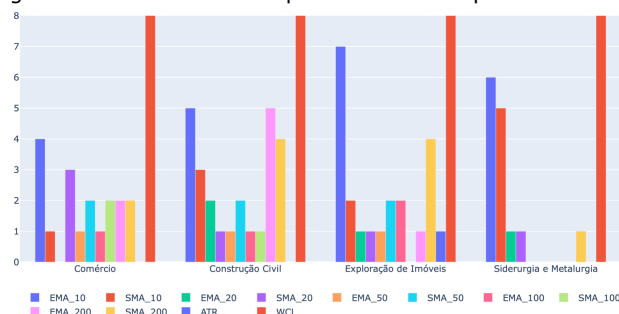


Fonte: Dados da pesquisa (2021).

A figura 11 onde o eixo x representa cada um dos clusters encontrados na etapa de clusterização, sendo A, C, D, E e R, os clusters 0, 2, 3, 4 e -1, respectivamente, permite observar a existência de diferentes indicadores técnicos com maior incidência em cada um dos clusters encontrados, comprovando que definitivamente é possível obter indicadores técnicos mais indicados para observação de um grupo de ativos.

Como exemplo observando o cluster A, que foi composto quase totalmente por ativos do setor financeiro, e por um ativo do setor de transporte, identifica-se que os indicadores técnicos mais influentes são o WCL, EMA_10, SMA_10, EMA_50, SMA_200, já o setor C, possui os indicadores WCL, EMA_10, SMA_100, EMA_200, SMA_200 como mais influentes.

Figura 12: Gráfico do resultado dos indicadores com maior grau de importância por setor.



Fonte: Dados da pesquisa (2021).

A figura 12 onde o eixo x representa os setores de Comércio, Construção Civil, Exploração de Imóveis, e Siderurgia e Metalurgia, permite observar a existência de diferentes indicadores técnicos com maior incidência em cada um dos setores observados, comprovando também que é possível identificar indicadores técnicos mais indicados em grupos bastante convencionais, que é o caso do setor de atuação da empresa analisada. Nesta figura é possível observar, por exemplo, a diferença entre indicadores técnicos mais adequados ao setor de comércio e construção civil.

5 Conclusões

O mercado de ações por dinâmica volátil e permeada de incertezas, tem sido alvo de muitos estudos científicos, a fim de contribuir para melhorar a sua predição e subsidiar as decisões de investimentos. Nessa análise, o presente estudo se debruçou na análise de indicadores técnicos e fundamentalistas com o intuito de tecer considerações para suportar e otimizar a análise dos ativos por parte dos investidores.

Diante disso, identificou-se grupos de indicadores técnicos mais adequados para análise por grupo de empresas e gerou-se bons insights ao comparar os clusters com diferentes setores de ativos. Assim, ao contribuir para a escolha adequada dos indicadores, a pesquisa permitiu a otimização do modelo de predição por parte dos investidores.

Por fim, propõe-se como sugestão para trabalhos futuros a identificação de clusters baseado na relação entre as séries temporais de preços a partir dos preços de abertura, de fechamento, mais alto do dia, mais baixo dia, adjunto de fechamento e volume.

Referências

- [1] SHAH, D.; ISAH, H.; ZULKERNINE, F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *Int. J. Financial Stud.* 2019, 7, 26. Disponível em: <<https://doi.org/10.3390/jifs7020026>>. Acesso em: 9 mar. 2021.
- [2] GREGÓRIO, R. Ibovespa tem maior volatilidade desde 2008, e ações oscilam até 117%; veja ranking. *Valor Investe*, São Paulo, nov. 2020. Disponível em: <<https://valorinveste.globo.com/mercados/renda-variavel/bolsas-e-indices/noticia/2020/11/09/ibovespa-tem-maior-volatilidade-desde-2008-e-acoes-oscilam-ate-117percent-veja-ranking.ghtml>>. Acesso em: 18 mar. 2021.
- [3] ANDRADE, J. B3 bate recorde e movimenta R\$26 bilhões por dia em 2020, jul 2020. Disponível em: <<https://einvestidor.estadao.com.br/investimentos/b3-recorde-26-bilhoes-dia/>>. Acesso em: 18 mar. 2021.
- [4] NAIK, N.; MOHAN, B. R. (2019) Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique. A. K. Somani et al. (Eds.): ICETCE 2019, CCIS 985, páginas 261–268, 2019.
- [5] REZENDE, I. C. C.; Duarte, J. M. S.; Silva, P. Z. P. da; Silva, G. R. da. Análise da performance empresarial da PETROBRAS: um estudo sob o enfoque dos indicadores econômico-financeiros. *REVISTA AMBIENTE CONTÁBIL*, v. 2, n. 1, p. 54-69, 2010.
- [6] GITMAN, L. J. *Princípios de Administração Financeira*. 12º ed. São Paulo: Pearson, 2010.
- [7] SCHOLTMANN, R.; CZUBATINSKI, M. *Trading: Technical Analysis Masterclass: Master the financial markets*. February 18, 2019.
- [8] LIU, L.; PAN; Z. Forecasting stock market volatility: The role of technical variables. *Economic Modelling*, ISSN 0264-9993, volume 84, páginas 55-65, 2020.
- [9] LEE, L.; KIM, G. PARK, S.; JANG, D. Hybrid Corporate Performance Prediction Model Considering Technical Capability. *MDPI, Open Access Journal*, vol. 8(7), pages 1-13, July 2016.
- [10] OLIVEIRA, A. M. de; SILVA, A. T. da. Associação entre o valor adicionado e indicadores de desempenho de companhias listadas na bolsa de valores (B3). *Revista Organização Sistêmica*, v. 6, n. 10, 2020.
- [11] LIMA, M. L. Um modelo para predição de bolsa de valores baseado em mineração de opinião. 2016.
- [12] LANDIM, I. M.; OLIVEIRA, R. da S.; LIMA JUNIOR, C. G. de.; MENEZES, J. C. de.; SILVA NETO, O. L. da.; GOMES, D. W. R. G. Indicadores econômicos e financeiros: um estudo de caso da empresa brasileira de mineração Vale SA. *Brazilian Journal of Development*, v. 6, n. 7, p. 50244-50255, 2020.
- [13] GALVÃO, N. D.; MARIN, H. de F.. Data mining: a literature review. *Acta Paulista de Enfermagem*, v. 22, n. 5, p. 686-690, 2009.
- [15] OCHI, L. S.; DIAS, C. R.; SOARES, S. S. F. Clusterização em mineração de dados. Instituto de

Computação-Universidade Federal Fluminense-Niterói, v. 1, p. 46, 2004.

[16] AMO, S. de. Técnicas de mineração de dados. Jornada de Atualização em Informática, 2004.

[17] MEDEIROS, C. J. F.; COSTA, J. A. F. Uma Comparação de Métodos de Redução de Dimensionalidade Utilizando Índices de Preservação da Topologia. In: Proc. Anais do IX Congresso Brasileiro de Redes Neurais/Inteligência Computacional (IX CBRN). 2009.

[18] PRADO, P. de O. Aplicação de técnicas multivariadas para visualização de dígitos manuscritos.

[19] SEMAAN, G. S.; Cruz, M. D.; Brito, J. A. M.; Ochi, L. S. Proposta de um método de classificação baseado em densidade para a determinação do número ideal de grupos em problemas de clusterização. Journal of the Brazilian Computational Intelligence Society, v. 10, n. 4, p. 242-262, 2012.

[20] WU, X.; KUMAR, V.; QUINLAN, J. R.; GHOSH, J.; YANG, Q.; MOTODA, H.; McLachlan, G. J.; Ng, S.K.A.; Liu, B.; Yu, P. S.; Zhou, Z.; Steinbach, M.; Hand, D.; Steinberg, D. Top 10 algorithms in data mining. Knowledge and Information Systems.

[21] FLORES, J. H. F. Comparação de modelos MLP/RNA e modelos Box-Jenkins em séries temporais não lineares. 2009.

[22] PRODANOV, C. C.; FREITAS, E. C. de. Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico. 2 ed. Novo Hamburgo: Feevale, 2013.

[23] RICHARDSON, R. J Pesquisa social: métodos e técnicas. 3 ed. São Paulo: Atlas, 2012.

[24] VALE. Ferrovia Centro-Atlântica FCA, jan. 2000. Disponível em: <<http://www.vale.com/brasil/EN/investors/information-market/press-releases/Pages/ferrovia-centro-atlantica-fca.aspx>>. Acesso em: 12 mai. 2021.

[25] BMFBOVESPA. Ferrovia Centro-Atlântica S.A., mai. 2020. Disponível em: <<http://bvmf.bmfbovespa.com.br/cias-listadas/empresas-listadas/>

ResumoEmpresaPrincipal.aspx?codigoCvm=15369&idioma=pt-br>. Acesso em: 12 mai. 2021.

[26] TSANTEKIDIS, Avraam; TEFAS, Anastasios. Transferring trading strategy knowledge to deep learning models. Knowledge and Information Systems, v. 63, n. 1, p. 87-104, 2021.

[27] KUNG, James J.; LIN, Wen-Ying; CHEN, Hsiu-Li. How effective are technical rules in predicting the 2008 global financial crisis? The case of the four Asian tigers. Journal of the Asia Pacific Economy, v. 26, n. 1, p. 13-33, 2021.

[28] YILDIRIM, Deniz Can; TOROSLU, Ismail Hakkı; FIORE, Ugo. Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. Financial Innovation, v. 7, n. 1, p. 1-36, 2021.

[29] GURAV, Uma; KOTRAPPA, Dr S. Predict stock market's fluctuating behaviour: Role of investor's sentiments on stock market performance. SSRG International Journal of Engineering Trends and Technology, v. 68, p. 72-80, 2020.

[30] LIU, Li; PAN, Zhiyuan. Forecasting stock market volatility: The role of technical variables. Economic Modelling, v. 84, p. 55-65, 2020.

[31] NTI, Isaac Kofi; ADEKOYA, Adebayo Felix; WEYORI, Benjamin Asubam. A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, p. 1-51, 2019.

[32] ZWART, Gerben De; MARKWAT, Thijs; SWINKELS, Laurens; DIJK, Dick Van. The economic value of fundamental and technical information in emerging currency markets. Journal of International Money and Finance, 28(4), p. 581- 604, 2009.

[33] SHAH Dev, ISAH Haruna, and ZULKERNINE Farhana. Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 2019.