

# Aplicação de Mineração de Dados para Avaliação de Indicadores Financeiros

*Implementing data mining to for financial indicators evaluation*

**Leonardo de Leon Dias<sup>1</sup>**

 [orcid.org/0000-0003-2336-9858](https://orcid.org/0000-0003-2336-9858)

**Diogo Tavares Cavalcanti de Moraes<sup>1</sup>**

 [orcid.org/0000-0001-7524-6206](https://orcid.org/0000-0001-7524-6206)

**Kelly Bezerra da Mota<sup>1</sup>**

 [orcid.org/0000-0001-6404-6401](https://orcid.org/0000-0001-6404-6401)

**Guilherme Teixeira Ribeiro de Albuquerque<sup>1</sup>**

 [orcid.org/0000-0003-0610-5834](https://orcid.org/0000-0003-0610-5834)

**Gabriela de Vasconcelos<sup>1</sup>**

 [orcid.org/0000-0002-2632-7242](https://orcid.org/0000-0002-2632-7242)

<sup>1</sup>Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. E-mail: [lld@ecomp.poli.br](mailto:lld@ecomp.poli.br)

**DOI: 10.25286/rep.v6i5.1753**

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018: DIAS, L. de L.; MOTA, K. B. de; VASCONCELOS, G. de; MORAES, D. T. C. de; ALBUQUERQUE, G. T. R. de. Aplicação de mineração de dados para avaliação de indicadores financeiros. Revista de Engenharia e Pesquisa Aplicada, Recife, v.6, n. 5, p. 37-46, Novembro, 2021.

## RESUMO

A identificação de tendências de modificação nos valores de ativos do mercado financeiro é uma forma eficaz de auxiliar a tomada de decisão e de melhorar o desempenho de investidores. Essa tarefa é bastante complexa até mesmo para profissionais e pesquisadores experientes. Para enfrentar esse desafio, normalmente busca-se amparo em informações obtidas através da análise de indicadores técnicos e fundamentalistas. Dito isso, 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 avaliando-os de modo a identificar aqueles que forneceram os melhores resultados de forma direta ou categorizada, sendo esta categorização feita por período, setor da atuação da empresa ou observando outras características compartilhadas advindas de análise fundamentalista. Por fim, foi possível obter bons insights capazes de contribuir para a identificação de indicadores técnicos que melhor se aplicam a cada grupo de ativos analisados.

**PALAVRAS-CHAVE:** Predição de Série Temporal; Mercado Financeiro; Indicadores Técnicos.

## ABSTACT

Identifying changing trends in financial market asset values is an effective way to aid decision-making and improve investor performance. Although this task is quite complex, even for experienced professionals and researchers, to face this challenge, one usually seeks support in information obtained through the analysis of technical and fundamental indicators. That said, this work aimed to evaluate the performance of different technical indicators applied, individually or together, to various historical series of market opening and closing, considering them to identify those that provided the best results directly or categorized. This categorization is done by period, sector of activity of the company, or observing other shared characteristics arising from fundamental analysis. Finally, it was possible to obtain good insights capable of contributing to identifying technical indicators that best apply to each group of analyzed assets.

**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, 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 técnica e fundamentalista [1]. Nesse contexto, a fim de auxiliar os investidores na tomada de decisão, vários métodos têm sido propostos pela literatura, entre os quais a utilização de mineração de dados e inteligência artificial [2; 3]. Investidores e empresas de consultoria apontam a importância de decidir adequadamente quais indicadores aplicar aos diferentes ativos, visto que indicadores com maior grau de importância nas análises técnicas e fundamentalistas podem variar de acordo com a empresa analisada [1].

A questão norteadora desta pesquisa visa identificar grupos de indicadores técnicos que melhor se apliquem a determinados segmentos de ativos. Para tanto, buscou-se otimizar o processo de escolha de indicadores que melhor se apliquem a ativos negociados na Bolsa de Valores, Mercadorias & Futuros (B3) a fim de amparar o processo de decisão de investidores. Para atingir o objetivo, foram utilizadas diversas técnicas de mineração de dados e de aprendizagem de máquina.

Destarte, analisar e entender a dinâmica do mercado são fatores importantes para o investimento em ações, portanto, estudar os diversos indicadores a serem considerados na análise de uma empresa é uma tarefa fundamental para se obter lucro em investimentos [2].

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 no período analisado, chegou a oscilar entre 25% e 117% [4]. Somado a isso, relatórios da B3 de 2021 apontam um crescimento de 17% no volume médio de negociações diárias no primeiro trimestre, atingindo a marca de R\$36,8 bilhões, em comparação ao trimestre anterior. Este comportamento reflete a necessidade de possuir ferramentas para embasar as decisões de investidores.

Diante disso, convergindo com a hipótese de que

é possível melhorar um modelo de predição para um determinado ativo escolhendo um conjunto de indicadores que melhor se adequem ao papel analisado [2], buscou-se identificar indicadores técnicos que se mostram mais eficazes quando aplicados a determinados ativos. Este estudo propõe contribuições que podem auxiliar na seleção e análise de indicadores financeiros, com o intuito de otimizar o retorno de investimentos. 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 se pretende 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 compõe o mercado financeiro, que 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 [5].

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 [2]. Assim, em termos simplificados, o valor de qualquer ativo é o valor presente de todos os fluxos de caixa futuros que se espera durante o período em questão [5].

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 fundamentalista que respectivamente destinam-se à avaliação de

gráficos de histórico de preços e à avaliação da saúde financeira das empresas [6].

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 [2].

Por outro lado, 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 ativo [7], dessa forma é possível observar a movimentação anterior dos valores de um ativo através da observação de gráficos, com o intuito de obter indícios sobre possíveis comportamentos futuros, e direcionar a tomada de decisão [1].

### 2.1.1 Performance de diferentes indicadores aplicado ao mercado de ações

O princípio básico subjacente à análise técnica tem como base a hipótese do mercado eficiente. Este viés acredita que todas as informações a respeito de um ativo em um dado momento estão incorporadas em seu preço [8].

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 [9].

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 se permite observar a partir de quatro classes de indicadores, conforme disposto no Quadro 1 [9].

Os indicadores de momento são frequentemente conhecidos como osciladores, pois oscilam entre os limites superiores e inferiores definidos e 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 prevaiente; 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 [9].

**Quadro 1** - Distribuição de indicadores por classes.

CLASSE	INDICADOR
Momentum	Stochastic; RSI; CCI; Williams %; MACD; Ichimoku Cloud
Trend	ADX; Moving averages; ATR; MACD; Parabolic SAR; Bollinger Bands®; Ichimoku Cloud
Volatility	Bollinger Bands®; Standard deviation; ATR; Keltner Channel; Envelopes
Chart studies	Horizontal lines; Fibonacci; Supply / Demand; Trend lines

**Fonte:** Adaptado de [9].

## 2.2 MINERAÇÃO DE DADOS

Nos últimos anos uma série de trabalhos relevantes foi proposta 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 [10]. Dentre as inúmeras técnicas de mineração utilizadas no estudo do mercado financeiro, destacam-se a clusterização e a regressão, as quais foram empregadas neste trabalho.

A clusterização é utilizada para separar os registros de uma base de dados em subconjuntos, de tal forma que os seus elementos compartilhem propriedades, apresentando uma configuração em que cada elemento possua maior similaridade com elementos do mesmo cluster em relação ao de outros grupamentos, isto é, servindo para distinguir os elementos entre clusters distintos, uma vez que maximiza a similaridade intra-cluster e minimiza a similaridade inter-cluster [11].

Um dos maiores benefícios da utilização dos modelos de clusterização para realização de agrupamentos é a sua capacidade de aplicação não supervisionada, fazendo com que eles se apliquem quando não se sabe as etiquetas das

classes desejadas. Isso permite a realização de análises aprofundadas mesmo sem o desenvolvimento de conhecimento prévio sobre o tema estudado [12].

Diante disso, o DBSCAN figura como um método bastante utilizado quando não se sabe a classe desejada, tanto por sua eficiência quanto por possuir uma simplicidade. Este método baseia-se em um conceito de densidade focado no centro do agrupamento, que leva em consideração o raio de alcance dos objetos e deve ser calibrado buscando o equilíbrio de um dos seus principais parâmetros, que permanecendo nessa zona pré-determinada, podem ser classificados em: interiores ou centrais; contidos em grupos ou baseados na densidade; limítrofes; e ruídos [13].

Já a regressão tem como principal objetivo aproximar uma função que representa os dados inseridos no modelo, permitindo que a partir de um conjunto de dados iniciais seja possível prever valores futuros. Um dos modelos bastante utilizados são as redes neurais, conjunto de técnicas aplicadas a uma grande gama de cenários.

Dentre as redes neurais, o MLP (*Multilayer Perceptron*) destaca-se tanto por ser um dos primeiros modelos concebidos, como pela sua fácil implementação. Por ser uma rede neural, naturalmente a informação se encontra nos pesos sinápticos em que os sinais de entrada são transmitidos por camadas totalmente conectadas. Essa macroestrutura, com apenas uma camada escondida, é capaz de aproximar qualquer função contínua oferecida em sua camada inicial. Entretanto, uma forte limitação desses modelos é a grande dificuldade de interpretação do que está ocorrendo internamente na rede [14].

## 2.3 TRABALHOS RELACIONADOS

Muitos 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 resultaram 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 técnicos para previsão baseado em mineração de dados [15]. O Quadro 2 apresenta alguns trabalhos relacionados a esta pesquisa.

**Quadro 2:** Estado da arte acerca da mineração de dados no mercado de ações

REF.	INDICADOR
[16]	Apresentou um método de pré-processamento de preços por meio da aplicação de redes neurais. Os resultados apontaram que os modelos de aprendizado profundo testados apresentaram desempenho superior aos modelos mais simples e aprenderam com precisão estratégias de negociação.
[17]	Investigou as regras técnicas mais eficazes em mercados em baixa do que em alta utilizando autorregressão vetorial; regras técnicas como médias móveis; e método bootstrap para inferências estatísticas.
[18]	Utilizou uma ferramenta de aprendizado profundo, Long short-term memory (LSTM), que se mostrou eficaz para fazer previsões a partir de um modelo híbrido de previsão de direção Forex, utilizando dados macroeconômicos e indicadores técnicos.
[19]	Analisou sentimentos de investidores nas eleições presidenciais dos EUA e os efeitos da Covid-19 como um fator de flutuação que afeta o desempenho do mercado de ações, para determinar quando comprar ou vender ações observando melhores regras, pontos de negociação e indicadores técnicos, utilizando o modelo de previsão Log Bilinear (LBL) e Rede Neural Recorrente (RNN) para aprendizagem de padrão de sentimento do mercado.
[20]	Buscou prever a volatilidade do retorno das ações usando indicadores técnicos construídos com base no histórico de preços, volatilidade e volume de negócios por meio de autorregressão. Observou que as variáveis fundamentalistas geram previsões mais precisas quando a economia está em recessão enquanto, as técnicas têm um desempenho melhor quando a economia está em expansão, mostrando que a combinação dessas informações permite fazer análises mais confiáveis do que utilizando-as isoladamente.

**Fonte:** Os autores.

Assim, observa-se 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, gerando contribuições para uma maior acurácia na predição do mercado acionário.

### 3 MATERIAS 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 [21] com o objetivo de aprofundar-se na 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, este estudo classifica-se como pragmático, produzindo conhecimentos úteis para solucionar problemas reais [21], que nesse caso, diz respeito a otimização das análises dos papéis por meio de indicadores que possam prever o valor das ações por segmento de atuação da empresa, a fim de associar a um modelo de predição com vistas à melhoria de performance.

Assim, para consecução desta pesquisa, buscou-se, a princípio, encontrar grupos de ativos a partir da observação de características semelhantes advindas de uma análise fundamentalista; 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.

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 B3. Sendo assim, a abordagem adotada foi a quantitativa, pois o estudo realizado utiliza amplamente de técnicas quantitativas de mineração de dados, por meio do emprego de simulações, provas e testes do modelo proposto [22].

#### 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, 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*).

Os dados históricos de abertura e fechamento de mercado utilizados nesta pesquisa 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. Com relação aos dados históricos de análise fundamentalista, após a redução de uma base preliminar contendo 58 colunas, obteve-se uma tabela de 378 linhas (empresas) e 21 colunas (indicadores fundamentalistas).

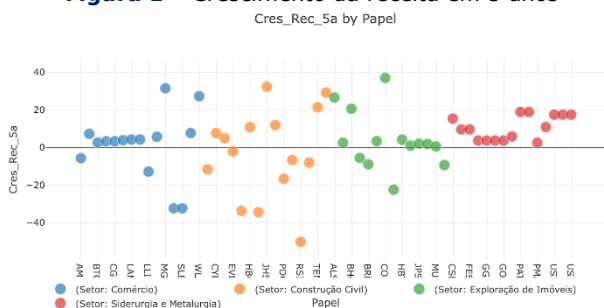
Os dados de histórico de preço utilizados para consecução da pesquisa consideraram os seguintes atributos: data, preço de abertura, preço mais alto do dia, preço mais baixo do dia, preço de fechamento, ajuste do preço de fechamento e volume de operações realizadas. No que concerne aos dados da análise fundamentalista, foram utilizadas as colunas de indicadores mais relevantes.

Após a realização da redução dos dados preliminares, obtiveram-se os seguintes atributos: Papel; Setores; Subsetor; Cotação; Valor de mercado; PL (Preço sobre lucro); PVP (Preço sobre valor patrimonial); PEBIT (Ganhos proprietários antes de interesse e taxas); Div\_Yield (Rendimento de dividendo); EV\_EBITDA (Relação entre o valor da companhia e o EBITDA); EV\_EBIT (Comparação do valor da empresa com o lucro antes da incidência de taxas e impostos); Cres\_Rec\_5ª (Crescimento da Receita Líq. nos últimos 5 anos); LPA (Lucro por ação); VPA (Valor Patrimonial da ação); Marg\_Bruta (Margem do lucro bruto); Marg\_Liquida (Margem líquida); ROIC (Retorno sobre o capital investido); ROE (Retorno sobre o patrimônio líquido); Div\_Br\_Patrim (Dívida Bruta total ÷ Patrimônio Líquido); Div\_Bruta (Dívida Bruta); Div\_Liquida (Dívida Líquida).

### 3.2 ANÁLISE DESCRITIVA E PRÉ-PROCESSAMENTO DOS DADOS

Com relação aos dados de análise fundamentalista utilizados no estudo, verifica-se uma grande variação entre cada um dos indicadores. Empresas de diferentes setores podem apresentar características bastante distintas. 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 1 – Crescimento da receita em 5 anos



Inúmeras constatações interessantes podem ser feitas a partir 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, optou-se pela não remoção de nenhum dos outliers observados. Esta decisão foi tomada conjuntamente com profissionais e especialistas da área de investimentos que contribuíram com o desenvolvimento desta pesquisa, totalizando 1 gerente de investimentos; 1 assessor de investimento; 4 especialistas no mercado de ações e 2 pesquisadores doutores em finanças.

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 todas as colunas de indicadores fundamentalistas resultaria em uma limitação muito grande da base de ativos analisados.

### 3.3 METODOLOGIA EXPERIMENTAL

#### 3.3.1 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 e possui uma característica cíclica, passando pelas etapas de compreensão do negócio, compreensão dos dados, identificação, pré-processamento, preparação e organização dos dados disponíveis e necessários, definição das técnicas de modelagem e do modelo que atende melhor aos objetivos de negócio e por fim, apresentação e avaliação dos resultados [23].

Inicialmente objetivou-se criar grupos de ativos negociados no mercado de ações brasileiro, a partir da extração de características provenientes da interpretação dos dados. Em razão disto, foi decidido utilizar a técnica de clusterização DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [24].

A justificativa para escolha deste método recai sobre a necessidade de tratar a existência de características distintas de cada uma das colunas dos dados, onde há uma significativa diferença entre cada um dos indicadores fundamentalistas; e pela capacidade de encontrar ruídos nos dados, que neste caso podem surgir por se ter optado pela não remoção de outliers uma vez que acabaria por reduzir significativamente a base de dados inicial. Além disso, optou-se por não definir um número de clusters previamente, visto que não há nenhuma referência ao quantitativo de grupos que podem melhor representar os dados.

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%, optando-se por manter na base os principais indicadores apontados por especialistas no mercado de ações consultados ao longo do estudo.

Após o pré-processamento dos dados, aplicou-se o algoritmo PCA para compactar o conjunto de dados em um subespaço de recurso de dimensão inferior mantendo a maioria das informações relevantes. O resultado foi utilizado para alimentar o algoritmo t-SNE [25], que serviu ao propósito de fazer um mapeamento dimensional dos dados, fornecendo os dados de entrada do algoritmo DBSCAN.

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.

### 3.3.2 Etapa de regressão

Com o intuito de atingir o objetivo principal deste trabalho, que se destina a identificar quais indicadores técnicos contribuem melhor para a análise de diferentes grupos de ativos analisados, optou-se por utilizar uma rede neural MLP como regressor para fazer previsão das séries de preço dos ativos analisados, e a técnica Random Forest Regressor para avaliar os melhores indicadores, com isso foram realizadas previsões de curto prazo, com o intuito de prever o valor de um ativo em 1(um) dia posterior, dado uma janela temporal de dez dias, utilizando como entrada o histórico de preços dos ativos analisados acrescidos de uma série de indicadores técnicos.

Os dados das séries históricas, constituídas inicialmente de seis colunas, contendo valores de preço de abertura, fechamento, máximo, mínimo, ajuste de fechamento e volume negociado, foram expandidos a partir da inserção de diversos indicadores técnicos calculados com base nas colunas inicialmente citadas. Essa expansão resultou em um total de 21 colunas de dados contendo, além das 6 colunas iniciais, os seguintes indicadores técnicos: SMA e EMA para 10, 20, 50, 100 e 200 períodos, ATR, CCI, ROC, RSI e WCL.

Para atingir o objetivo desejado, primeiramente foram executadas duas rodadas de treinamento e teste, com respectivamente 70% e 30% dos dados, para cada um dos 378 papéis analisados utilizando como input o conjunto total de indicadores existentes. Para cada rodada de treinamento foi armazenado o RMSE (Root Mean squared error) como métrica de avaliação do modelo. Em seguida, foram verificados quais as features de maior importância (indicadores técnicos de maior importância) utilizando o Random Forest Regressor. Por fim, a base de dados inicial foi alterada mantendo-se apenas 25% dos indicadores, sendo estes os de maior relevância, e então realizou-se novamente duas rodadas de treinamento e teste.

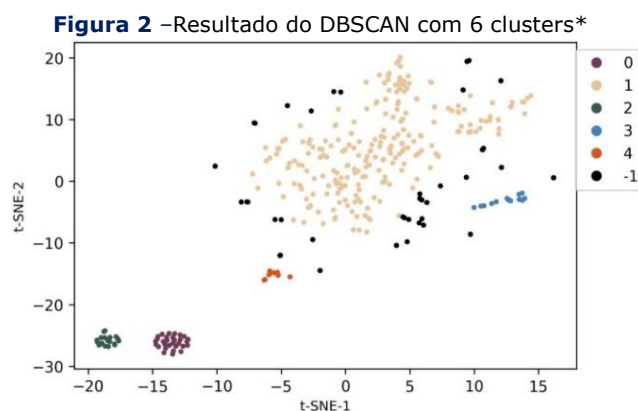
Em seguida, foi implementado um modelo de previsão utilizando um MLP com múltiplos parâmetros de entrada, tendo como saída a previsão do valor de fechamento do mercado de 1 dia posterior. Para construção do MLP foi utilizado

uma camada de entrada de 200 neurônios, 300 neurônios na camada escondida e função de ativação RELU. A rede foi compilada utilizando o otimizador ADAM, MSE (Mean squared error) como métrica de avaliação, Batch de 32 e taxa de aprendizagem de 0,001.

Após as etapas anteriores foi gerado uma nova base de dados contendo, para cada papel analisado: o RMSE obtido na primeira rodada da MLP, os indicadores mais relevantes de acordo com o Random Forest Regressor, e os valores de RMSE da segunda rodada da MLP. Por fim, os dados gerados, assim como os clusters obtidos na etapa de clusterização e os setores correspondentes de cada ativo, foram associados gerando uma nova base de dados. Com estes dados gerados, foi possível obter uma série de informações relevantes relacionadas ao objetivo proposto.

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

O presente estudo buscou por meio da aplicação de mineração de dados, encontrar quais indicadores técnicos podem ser melhor aplicados a um determinado grupo de ativos a partir da observação das suas características semelhantes. Para tanto, 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 2.



\*Legenda: Clusters 0, 1, 2, 3, 4 e -1 foram denominados A, B, C, D, E e R respectivamente.

**Fonte:** Os Autores.

Diante deste resultado, foi realizada uma análise comparativa com os setores dos ativos presentes em cada cluster, onde foi observado, por

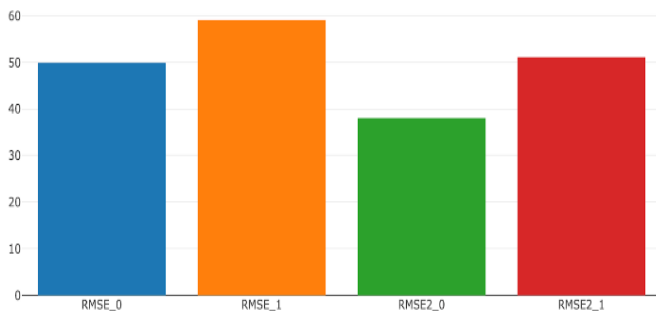
exemplo, que praticamente todos os papéis contidos no grupo A são do setor de Bancos, com exceção de um único papel do setor de Transporte, o VSPT3, logo, o referido ativo pode ser identificado como outlier em seu setor. No entanto, fazer uma análise detalhada dos ativos presentes na base de dados utilizada foge do escopo deste trabalho<sup>1</sup>. Foi constatado ainda que todos os papéis do setor Agropecuário estavam contidos no grupo B, assim como todos os demais papéis do setor de Transporte.

Assim, foi possível identificar os indicadores, utilizados individualmente ou em conjunto, que apresentaram melhor desempenho para realizar a análise dos ativos, respeitando as relações existentes entre os diferentes indicadores técnicos e os papéis analisados, através da observação tanto dos setores de atuação das empresas quanto das 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. Foi possível observar ainda que a utilização de um grupo seletivo de indicadores mais influentes por papel beneficia a tarefa de predição quando comparado à utilização de todos os indicadores no treinamento do modelo.

A figura 3 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.

**Figura 3** – Gráfico da média do RMSE por ciclo de treinamento



Fonte: Os Autores.

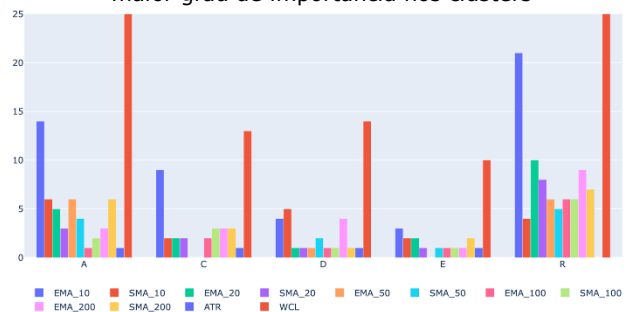
Pode-se observar que a média do RMSE\_0, referente a avaliação do erro no primeiro ciclo da

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.

A fim de avaliar e comparar a contribuição de diferentes indicadores técnicos por cluster e por setor, foram obtidas as seguintes representações: (1) por cluster; (2) por setor, conforme demonstrado nas figuras 4 e 5.

Na avaliação por cluster disposta na figura 4 - em que 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 observados na figura 5 - é possível 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.

**Figura 4** – Gráfico do resultado dos indicadores com maior grau de importância nos clusters



Fonte: Os Autores.

Na comparação por cluster observa-se que o cluster A, composto quase totalmente por ativos do setor financeiro e por um ativo do setor de transporte, apresentou como indicadores técnicos mais influentes o WCL, EMA\_10, SMA\_10, EMA\_50, SMA\_200, ao passo que no setor C, os mais influentes foram WCL, EMA\_10, SMA\_100, EMA\_200, SMA\_200.

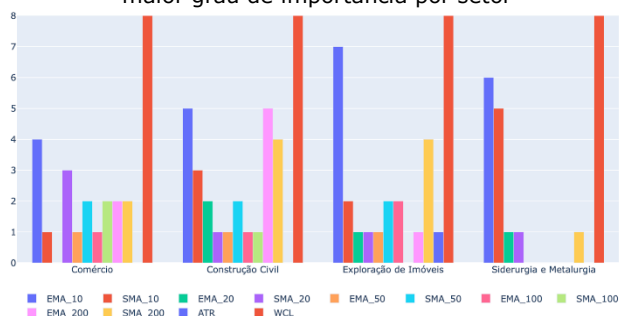
Já na avaliação por setor, observada na figura 5 - em que o eixo x representa os setores de Comércio, Construção Civil, Exploração de Imóveis, e Siderurgia e Metalurgia - pode-se observar a existência de diferentes indicadores técnicos com maior incidência em cada um dos setores observados, comprovando também que é

<sup>1</sup> Mais informações sobre o ativo podem ser encontradas em [26] e [27].



possível identificar indicadores técnicos mais influentes em grupos bastante convencionais, como por exemplo o setor de atuação da empresa.

**Figura 5** – Gráfico do resultado dos indicadores com maior grau de importância por setor



Fonte: Os Autores.

Na comparação por setor, representada na figura 5, é possível observar que os três primeiros indicadores técnicos mais adequados ao setor de comércio foram WCL, EMA\_10 e SMA\_20, enquanto no setor de construção civil destacaram-se WCL, EMA\_10 e EMA\_200, ao passo que, no setor de Siderurgia e Metalurgia, sobressaíram-se os indicadores WCL, EMA\_10 e SMA\_10.

Diante do exposto, observou-se que a partir das técnicas de mineração de dados utilizadas foi possível encontrar quais indicadores técnicos podem ser melhor aplicados a um determinado grupo de ativos, a partir da observação das características inerentes aos papéis analisados.

## 5 CONCLUSÕES

O mercado de ações por sua 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. Nesse sentido, 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 conjuntos de indicadores técnicos mais adequados para análise de diferentes grupos de empresas, gerou-se bons insights ao comparar clusters gerados a partir de ferramentas de mineração de dados, com diferentes setores de ativos. Assim, a pesquisa permitiu a otimização de um modelo de predição, ao contribuir para a escolha adequada dos indicadores técnicos a serem utilizados para análise de diferentes empresas.

Destarte, este estudo mostra-se relevante uma vez que teve contribuições importantes para auxiliar nas decisões de investimentos, encurtando o caminho do painel de análise de indicadores, reduzindo o escopo de parâmetros analisados. Ademais, contribui para o avanço da literatura que tangencia a aplicação de mineração de dados no cenário do mercado de ações.

Como limitação, destaca-se a baixa capacidade de processamento do ambiente onde foi desenvolvida a pesquisa, o que impossibilitou a utilização de algumas ferramentas de machine learning que poderiam ter se mostrado mais eficazes.

Por fim, propõe-se como sugestão para trabalhos futuros a identificação de clusters baseado na relação entre as séries temporais de preços das empresas analisadas, propõe-se ainda uma abordagem semelhante à desenvolvida neste estudo no que tange a clusterização e demais métodos, mas com a utilização de diferentes indicadores técnicos.

## REFERÊNCIAS

- [1] SHAH, D.; ISAH, H.; ZULKERNINE, F. **Stock Market Analysis: A Review and Taxonomy of Prediction Techniques**. Int. J. Financial Stud. 2019, v. 7, n. 26. Disponível em: <https://doi.org/10.3390/ijfs7020026>. Acessado em: 9 mar. 2021.
- [2] NAIK, N.; MOHAN, B. R. **Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique**. A. K. Somani et al. (Eds.). ICETCE, 2019.
- [3] REZENDE, I. C. C. *et al.* **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.
- [4] 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>. Acessado em: 18 mar. 2021.

- [5] GITMAN, L. J. **Princípios de Administração Financeira**. 12. ed. São Paulo: Pearson, 2010.
- [6] NTI, I. K. *et al.* A systematic review of fundamental and technical analysis of stock market predictions. **Artificial Intelligence Review**, p. 1-51, 2019.
- [7] ZWART, G. *et al.* The economic value of fundamental and technical information in emerging currency markets. **Journal of International Money and Finance**, v. 28, n. 4, p. 581- 604, 2009.
- [8] EUGENE, F. F. **The behavior of stock-market prices**. The journal of Business, v. 38, n. 1, p.34-105, 1965.
- [9] SCHOLTMANN, R.; CZUBATINSKI, M. **Trading: Technical Analysis Masterclass**. Master the financial markets. February 18, 2019.
- [10] SHAH, D. *et al.* Stock market analysis: A review and taxonomy of prediction techniques. **International Journal of Financial Studies**, v. 7, n. 2, 2019.
- [11] GALVÃO, N. D.; MARIN, H. de F. Data mining: a literature review. **Acta Paulista de Enfermagem**, v. 22, n. 5, p. 686-690, 2009.
- [12] AMO, S. de. **Técnicas de mineração de dados**. Jornada de Atualização em Informática, 2004.
- [13] SEMAAN, G. S. *et al.* 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.
- [14] FLORES, J. H. F. Comparação de modelos MLP/RNA e modelos Box-Jenkins em séries temporais não lineares. 2009.
- [15] LEE, L. *et al.* **Hybrid Corporate Performance Prediction Model Considering Technical Capability**. MDPI, Open Access Journal, v. 8, n. 7, p. 1-13, 2016.
- [16] TSANTEKIDIS, A.; TEFAS, A. Transferring trading strategy knowledge to deep learning models. **Knowledge and Information Systems**, v. 63, n. 1, p. 87-104, 2021.
- [17] KUNG, J. J. *et al.* 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.
- [18] YILDIRIM, D. C. *et al.* Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. **Financial Innovation**, v. 7, n. 1, p. 1-36, 2021.
- [19] GURAV, U.; KOTRAPPA, D. 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.
- [20] LIU, L.; PAN; Z. **Forecasting stock market volatility: The role of technical variables**. Economic Modelling, v. 84, p. 55-65, 2020.
- [21] 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.
- [22] RICHARDSON, R. J. **Pesquisa social: métodos e técnicas**. 3 ed. São Paulo: Atlas, 2012.
- [23] GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações**. Rio de Janeiro: Elsevier, 2015.
- [24] CHEN, Y.; RUYS, W.; BIROS, G. **KNN-DBSCAN: a DBSCAN in high dimensions**. Journal CoRR, 2020. Disponível em: <https://arxiv.org/abs/2009.04552>. Acessado em: 25 jun. 2021.
- [25] LINDERMAN, G. C.; STEINERBERGER, S. **Clustering with t-SNE, Provably**. SIAM Journal on Mathematics of Data Science, v. 1, n. 2, p. 313-332, 2019.
- [26] VALE. **Ferrovias Centro-Atlântica FCA**, jan. 2000. Disponível em: <http://www.vale.com/brasil/EN/investors/information-market/press-releases/Pages/ferrovia-centro-atlantica-fca.aspx>. Acessado em: 12 mai. 2021.
- [27] BMFBOVESPA. **Ferrovias Centro-Atlântica S.A.**, mai. 2020. Disponível em: <http://bvmf.bmfbovespa.com.br/cias-listadas/empresas-listadas/ResumoEmpresaPrincipal.aspx?codigoCvm=15369&idioma=pt-br>. Acessado em: 12 mai. 2021.