UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MATHEUS MARRONE CASTANHO

Modelo de análise multidisciplinar para previsão de tendência mensal na bolsa de valores brasileira

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dra. Renata Galante

Porto Alegre
2022

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Castanho, Matheus Marrone

Modelo de análise multidisciplinar para previsão de tendência mensal na bolsa de valores brasileira / Matheus Marrone Castanho. – Porto Alegre: PPGC da UFRGS. 2022.

106 f.: il.

Orientadora: Renata Galante.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação. Porto Alegre, BR – RS, 2022.

1. Séries temporais. 2. Classificação 3. Previsão 4. Bolsa de valores 5. Algoritmos de Machine Learning. I. Galante, Renata. II. Título.

RESUMO

O avanço tecnológico tem permitido uma ampla utilização de soluções computacionais no mercado financeiro, especialmente aquelas relacionadas às áreas de ciência de dados e inteligência artificial. Investidores e empresas de investimentos vêm buscando soluções que auxiliem sua tomada de decisão para a compra de ativos. Parte dessas soluções visa a previsão do preço dos ativos num curto espaço de tempo. Através da análise de séries temporais e da utilização de algoritmos de aprendizado de máquina diversas pesquisas têm se mostrado promissoras em sua capacidade de previsão. Contudo, a utilização das cotações históricas dos ativos tem se mostrado insuficiente para aumentar a precisão das previsões. O preço dos ativos possui um comportamento bastante diversificado, sofrendo a influência do cenário macroeconômico global ou local, do comportamento dos investidores em diferentes períodos do ano e do movimento de outros índices acionários. O trabalho busca construir um modelo de análise para previsão de tendências mensais na bolsa de valores que utilize um conjunto de indicadores macroeconômicos, efeitos comportamentais e comportamento de outros índices de ativos. Esse modelo é testado considerando o cenário brasileiro, sua bolsa de valores e seus indicadores macroeconômicos. Esse modelo apresentou resultados interessantes para a classificação mensal em meses de baixa ou de alta, obtendo, em sua média, resultados acima de 60% para os diferentes algoritmos aplicados.

Palavras-chave: Séries temporais. Classificação. Previsão. Bolsa de valores. Algoritmos de Machine Learning.

A multidisciplinary analysis model for Monthly trend prediction in the Brazilian Stock Market

ABSTRACT

Technological advances have allowed for a wider application of computational solutions within the financial market, especially those related to the areas of data science and artificial intelligence. Investors and investment companies have been searching for solutions that assist in their decision making for assets allocation, most of which aim at forecasting the assets price within a short-term. Via the analysis of time series and the use of Machine Learning algorithms, several researches show promise in their predictive capacity. However, the use of historical asset prices has proved insufficient in increasing the accuracy of forecasts. Asset prices reveal highly diversified behavior, influenced by the global or local macroeconomic scenario, the behavior of investors in different periods of the year and the movement of other stock indices. This work aims at building a monthly trend forecasting stock market analysis model, considering macroeconomic indicators, behavioral effects and movement of other asset indices. The model is tested bearing in mind the Brazilian scenario, its stock exchange and its macroeconomic indicators. The algorithm presented interesting results for sorting months into negative and positive returns, achieving, on average, results above 60% for the different algorithms applied.

Keywords: Time series. Classification. Prediction. Stock market. Machine learning algorithm.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ANN	Artificial Neural Networks
ARIMA	Modelo autorregressivo integrado de médias móveis
BCB	Banco Central do Brasil
BGRU	Bidirectional Gated Recurrent Unit
BOVA11	ETF do Ibovespa
CDI	Certificado de Depósito Interbancário
CNN	Convolutional Neural Networks
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
EEM	iShares MSCI Emerging Markets ETF
ETF	Exchanged-Trading Fund
FFNN	Feed-Forward Neural Network
GCN	Graph Convolutional Network
GRU	Gated Recurrent Unit
IAU	iShares Gold Trust
IBGE	Instituto Brasileiro de Geografia e Estatística
Ibovespa	Índice Bovespa
IFIX	Índice de Fundos Imobiliários
IMA-B	Índice de Mercado Ambima série B
IPCA	Índice de Preços ao Consumidor Amplo
IVVB11	ETF brasileiro do S&P 500
IXIC	Nasdaq Composite
KNN	K-Nearest Neighbors
LSTM	Long term short memory
ML	Machine Learning
MLP	Multilayer Perceptron
MTFA	Algoritmo de previsão de tendências mensais
NASDAQ	Bolsa de valores americana
NB	Naïve Bayes
NBB	Naive Bayes Bernoulli
NN	Redes neurais

PCA	Principal Component Analysis
PIB	Produto Interno Bruto
PNADC	Pesquisa Nacional por Amostra de Domicílios Contínua
RF	Random Forest
RLOG	Regressão Logística
RNN	Rede neural recorrente
SMAL11	ETF do SMLL
SMGA	Sell in May and Go Away
SMLL	Índice das Small Caps da bolsa brasileira
SVC	Support Vector Classification
SVM	Support Vector Machine
SVR	Support Vector Regression
S&P 500	Índice Standard & Poor's 500
TCN	Temporal Convolutional Network
VIX	Índice de Volatilidade da Chicago Board Options Exchange

SUMÁRIO

1 INTRODUÇÃO

Ao longo dos últimos anos, uma grande transformação está em curso no mercado brasileiro de capitais. Com um cenário inédito de baixas taxas de juros, um grande número de investidores pessoa física vem migrando seus investimentos para ativos de renda variável em busca de uma maior rentabilidade. Esse crescimento é importante para a consolidação do mercado financeiro do país, uma vez que contribui para seu fortalecimento ao demandar uma melhora nos processos e uma redução dos custos aos clientes, ao propiciar o desenvolvimento de tecnologias inovadoras.

Contudo, o conhecimento de base é fundamental para ajudar o investidor a tomar melhores decisões. Compreender o funcionamento do mercado financeiro e suas características é essencial, assim como as relações existentes entre a economia e os principais índices acionários. Também devem ser levados em conta os diferentes comportamentos dos investidores, visando identificar oportunidades das quais estes se beneficiem. A grande diversidade existente no mercado financeiro, cada vez mais globalizado, permite ao investidor local buscar diversas classes de ativos para compor seu portfólio, mas, com a ausência de conhecimento, este poderá concentrar seus investimentos ao invés de diversificar, se expondo a eventuais riscos.

Uma das melhores formas de se obter conhecimento sobre todos esses fatores é através de estudo de dados históricos (JACKSON et al., 2021), que avaliam o comportamento do mercado ao longo de grandes períodos, fornecendo importantes *insights* para as tomadas de decisões futuras. Diversas pesquisas são conduzidas com a temática de previsão de tendências futuras (LIU H. et al., 2021) (GIACOMEL F. et al., 2015). Cientistas de dados vêm aplicando técnicas de aprendizado de máquina em séries temporais (LIU H. et al., 2021) (GIACOMEL F. et al., 2015) visando identificar padrões para investimentos de curto-prazo. Importantes estudos analisam o cenário macroeconômico de um país (SALISU A. et al., 2021) (TORAMAN C et al., 2014) (STONA F. et al., 2018) para compreender como este pode influenciar o mercado financeiro. Estudiosos em finanças comportamentais também buscam estudar e identificar anomalias no comportamento dos investidores (AHMED B, 2020) (PLASTUN A. et al., 2020) (CHEN Z. et al., 2018).

Com o avanço tecnológico, em especial o aumento da capacidade de processamento e de armazenamento de dados, estudos que antes demandavam uma enorme quantidade de tempo, se tornaram muito mais rápidos e mais precisos. E por isso, cada vez mais o mercado de trabalho

financeiro vem agregando engenheiros, analistas e cientistas de dados que possuem a capacidade de manipular e extrair informações de grandes volumes de dados, criando modelos e analisando tendências.

No mercado financeiro, investidores buscam constantemente novas estratégias com o objetivo de superar o desempenho médio do mercado de ações. Estratégias mais tradicionais acabam sendo deixadas de lado para análises de curto prazo que visam obter lucros em um curto período, muitas vezes fazendo o uso de previsão de tendências através da análise gráfica ou de indicadores.

No sistema financeiro atual existem os chamados fundos *Quants*[1], fundos que utilizam uma abordagem quantitativa com foco em modelos computacionais para conseguir se antecipar aos movimentos do mercado e obter lucro. Com o potencial de processamento cada vez maior, a utilização de algoritmos de inteligência artificial no mercado financeiro (DE MELLO ASSIS J. et al. 2018) é uma área bastante explorada, seja por investidores, economistas e cientistas da computação.

Entre as aplicações de inteligência artificial no mercado financeiro, destaca-se a utilização de técnicas de aprendizado de máquina para a previsão do comportamento futuro de um ou mais ativos. Com base nessa previsão, investidores podem posicionar seu portfólio para aproveitarem tanto os movimentos de alta quanto de baixa. A previsão desses movimentos é vista por muitos com bastante ceticismo, gerando debates e visões divergentes entre economistas e investidores sobre seus resultados.

No lado tecnológico, houve uma evolução das técnicas de previsão. Por um lado, há a previsão gráfica, que utiliza a chamada análise técnica para prever tendências futuras, sendo baseada apenas no preço do ativo alvo. Uma das vantagens das técnicas de aprendizado de máquina é a possibilidade de combinar diferentes fontes de dados para a tomada de decisão, gerando modelos mais complexos e dinâmicos.

As variáveis utilizadas nestes modelos acabam sendo um ponto chave para a performance dos mesmos. Alguns estudos (CHRISTY J. et al., 2021) avaliam aspectos financeiros das empresas, como alguns indicadores da própria empresa ou setoriais como dados úteis para a criação de um modelo de predição. Outros trabalhos (RYOTA K. et al, 2012) também consideram alguns dados macroeconômicos, como taxa de juros e cotação do barril de petróleo como indicadores que podem afetar o comportamento do preço de um ativo no futuro.

[1] www.investopedia.com/terms/q/quantfund.asp

Alguns estudos (XIA B. et al., 2013) (ZHANG L. et al., 2017) utilizam as cotações e outras informações de negociações como preço de fechamento e abertura, além do volume de negociação como dados de entrada. Outros trabalhos (HUYNH H. et al., 2017) (WANG Z. et al., 2015) buscam encontrar uma relação entre notícias e o comportamento da bolsa de valores. Alguns focam na otimização de algoritmos (TANG L. et al., 2018) para melhorar o desempenho, enquanto outros (LIU H et al., 2021) buscam uma combinação destes para obter maior precisão. As técnicas vistas na literatura são bastante variadas, indo de algoritmos de classificação mais simples como os de regressão linear, *Random Forest* (RF) e *Support Vector Machine* (SVM) (PATIL P. et al., 2020) (WANG H. et al., 2020) (LUO R., 2020), até algoritmos mais complexos que envolvem redes neurais artificiais (GIACOMEL F. et al., 2015) (DE MELLO ASSIS J. et al. 2018) ou LSTM (*Long short-term memory*) (HUANG B. et al. 2018) (OJO S. et al., 2015). A seleção da janela temporal (GONZALEZ R. et al., 2015) de análise pode impactar na obtenção de análises mais precisas, assim como a utilização de outras séries temporais, de índices ou ativos (RYOTA K. et al., 2012).

Observa-se a importância da seleção de variáveis chaves para a construção de modelos de previsão mais precisos. Esse trabalho se propõe a analisar algumas variáveis de diferentes áreas e seu relacionamento com a bolsa de valores e, com base nesse estudo, propor um modelo de previsão que será chamado de multidisciplinar. Esse modelo será avaliado juntamente com um conjunto de algoritmos de aprendizado de máquina, objetivando encontrar aquele que produz a maior precisão nas previsões.

Esse trabalho foi dividido em duas etapas. Na primeira, é proposta uma análise multidisciplinar de estudo de séries temporais. Essa análise compreende três fases. Na primeira, são analisados os indicadores macroeconômicos do país e de que forma eles impactam na bolsa de valores. Busca-se definir intervalos temporais que apresentem grandes variações nos dados macroeconômicos para se estudar os impactos no índice de ações. Na segunda, são conduzidas algumas investigações sobre o comportamento dos investidores da bolsa de valores e alguns estudos estatísticos. Por fim, são realizadas comparações entre diferentes índices locais e internacionais com a bolsa brasileira, analisando-se de que forma eles refletem no resultado desta última.

A segunda etapa do trabalho propõe um algoritmo de previsão de tendências mensais que visa avaliar o melhor conjunto de dados de entrada, composto pelas variáveis estudadas na primeira etapa do trabalho. A análise mensal visa ser um diferencial em relação a outros trabalhos, ao mesmo tempo que permite a utilização dos efeitos sazonais estudados na etapa de

análise. O algoritmo consiste de uma seleção e adaptação das variáveis de entrada e na análise de qual configuração terá maior impacto em prever o comportamento mensal do índice Bovespa. Utiliza-se um conjunto de técnicas de aprendizado de máquina (ML) comuns na literatura, visando se obter a melhor precisão a partir da média do resultado destas técnicas na avaliação do resultado mensal do Ibovespa como de alta ou de baixa.

Experimentos exaustivos mostraram que a precisão das previsões está ligada à seleção dos indicadores, apresentando resultados diversos de acordo com o conjunto escolhido. As técnicas de aprendizado de máquina em sua média obtiveram precisões parecidas, mesmo utilizando configurações mais genéricas. O trabalho também propõe algumas direções a serem seguidas que podem trazer melhores resultados, desde a diminuição da janela temporal, quanto uma seleção de outros índices financeiros ou setoriais. Além disso, é provável que uma customização mais expressiva de alguma das técnicas de aprendizado de máquina permitirá aumentar a precisão da previsão.

Este trabalho apresenta as seguintes contribuições:

- Modelo de análise de séries temporais financeiras multidisciplinar, composto por indicadores macroeconômicos, índices do mercado financeiros e efeitos comportamentais.
- Construção de um algoritmo para avaliação da predição mensal da bolsa de valores brasileira de acordo com um conjunto variado de entradas, obtidos a partir da análise multidisciplinar.
- Modelo que utiliza variadas técnicas de aprendizado de máquina para validar os diferentes conjuntos de dados de entrada.

O trabalho está estruturado em seis capítulos. No Capítulo dois é fornecida uma explicação sobre os principais conceitos utilizados no trabalho. O Capítulo três apresenta uma breve descrição dos trabalhos relacionados usados como inspiração e referência. No Capítulo quatro é especificado em detalhes o modelo proposto neste trabalho. Por fim, o Capítulo cinco apresenta os resultados das análises realizadas e o Capítulo seis oferece uma conclusão para o trabalho, bem como propostas para estudos seguintes.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados os principais conceitos apresentados no trabalho. Os conceitos financeiros que fazem parte da primeira fase do trabalho serão discutidos inicialmente, seguido pelos conceitos computacionais presentes na segunda fase.

2.1 Séries temporais

A evolução tecnológica trouxe um grande aumento no poder computacional. Esse aumento foi responsável pelo grande desenvolvimento recente da área de análise de dados. Cada vez mais presente em diversos campos e especialidades, a análise de dados aliada à ciência da computação e a estatística vem apresentando novas descobertas em uma escala cada vez maior. De posse de um grande volume de dados passados, que pouco puderam ser analisados devido às limitações tecnológicas, hoje é possível estudar e extrair conhecimento de grandes conjuntos de dados de forma muito mais rápida.

No setor financeiro, a evolução foi marcante. Se antes a contabilidade das empresas era registrada em livros caixa, os pregões da bolsa de valores feitos ao vivo e por voz, e as ordens de compra e venda de ações eram registradas em cartões, hoje tudo é digital. Felizmente, os órgãos regulatórios responsáveis e as bolsas de valores mantiveram esses dados guardados e criaram bases de dados que podem ser facilmente acessadas atualmente.

Diversos sites e empresas fornecem ferramentas para obter as cotações históricas das ações e dos principais índices de ações. Dentro do mundo de investimento, essas informações gráficas são a matéria-prima dos investidores que utilizam as estratégias de análise técnica. A análise técnica, também conhecida como análise gráfica, utiliza os gráficos das cotações e dos volumes de negociação dos ativos para identificar tendências e prever os movimentos do mercado. Muito utilizadas por investidores de curto prazo, que realizam operações de compra e venda em curtos períodos de tempo, essas análises visam identificar o comportamento atual e futuro do mercado para que o investidor obtenha vantagem através de um posicionamento que gerará resultados.

O estudo de séries temporais permite a avaliação do comportamento de diversos tipos de dados ao longo de um período de tempo. São famosas as séries temporais meteorológicas como as de temperatura máxima, mínima e regime de chuva anual, além de também serem bem

conhecidas nas ciências biológicas as séries temporais de monitoração cardíaca (REYHANI R. et al., 2011). Composto por um conjunto de valores e um índice, cada item de uma série temporal representa esses valores em um determinado período de tempo. Os itens dentro dessa série são ordenados de forma cronológica. O índice pode possuir qualquer periodicidade no formato de data/hora, como anual, mensal, diário, por hora, minuto ou segundo. A observação do comportamento temporal de dados é utilizada em diferentes áreas, dentre as quais se destaca o uso em análise estatística.

Na economia e no setor financeiro, séries históricas são utilizadas na condução de estudos sobre a evolução de indicadores, preços e outras importantes variáveis (CHRISTY J. et al., 2021). A análise histórica permite a observação de ciclos e tendências, além de auxiliar na identificação de eventos não recorrentes como crashes e bolhas (MCMILLAN D. G, 2021).

2.2 Mercado Financeiro

O mercado financeiro tem a bolsa de valores como um instrumento presente na sociedade para a negociação de títulos mobiliários entre diferentes atores. Um dos títulos mais comuns e negociados são as ações. Uma ação representa um percentual do capital social de uma empresa e pode ou não estar sendo negociada em uma bolsa de valores. São chamadas de empresas de capital aberto aquelas que negociam suas ações na bolsa de valores. Quando alguém compra uma dessas ações, acaba se tornando sócio e acionista desta empresa, tendo direito a participar do lucro e de algumas decisões de empresa, de acordo com o seu percentual de participação na empresa.

A atual bolsa de valores brasileira é a B3[2], bolsa de valores sediada em São Paulo, que possui mais de 380 empresas listadas no começo de 2021. No mundo dos investimentos, possuir um *benchmark* é fundamental para se acompanhar o desempenho das diferentes estratégias e composições de carteiras. Uma forma de representar o mercado de capital de um país e medir o seu desenvolvimento é através da utilização de um índice que o represente. Dessa forma, no Brasil temos o índice Bovespa ou Ibovespa[3], índice formado por uma carteira teórica com as maiores e mais negociadas empresas da bolsa brasileira. O índice é constituído por um

[2] www.b3.com.br

[3] www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm

percentual de cada empresa, proporcional ao valor de mercado dessa, sendo atualizado a cada quatro meses.

Na bolsa de valores brasileira, também ocorre a negociação de outras classes de ativos, como fundos de investimento imobiliários e fundos passivos. Os fundos imobiliários são fundos de investimentos que investem em imóveis, recebíveis imobiliários ou em outros fundos de investimentos imobiliários. Assim como existe o índice Bovespa para as ações brasileiras, existe o IFIX[4] para os fundos imobiliários. Os fundos passivos negociados em bolsa são conhecidos como ETF, *Exchanged-Traded Fund*. Estes fundos buscam replicar uma carteira teórica com os ativos presentes em um índice, geralmente comprando todas as ações desse índice na mesma proporção indicada. De uma forma simplificada, um investidor que deseja obter o mesmo desempenho do Ibovespa, poderá procurar investir num ETF que replique o índice Bovespa, como o BOVA11[5].

Outros importantes ETFs presentes no mercado brasileira são:

- SMAL11[6]: fundo que investe no índice *small caps* (SMLL[7]). Esse índice é composto por empresas de baixa capitalização e volume de negociação diário.
- IVVB11[8]: fundo que investe no índice americano Standard & Poor's 500[9] (S&P 500). Esse índice é formado pelas 500 maiores empresas americanas. No Brasil o IVVB11 é cotado em reais, mas acompanha a variação dos preços em dólar, uma vez que o fundo compra empresas na bolsa americana.

Nos mercados globais, especialmente nos Estados Unidos, a quantidade de ETFs listados é mais significativa se comparada com a bolsa brasileira. Diversos tipos de índices são

---

4 www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-de-fundos-de-investimentos-imobiliarios-ifix.htm

5 www.blackrock.com/br/products/251816/ishares-ibovespa-fundo-de-ndice-fund

6 www.blackrock.com/br/products/251752/ishares-bmfbovespa-small-cap-fundo-de-ndice-fund

7 www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/indice-small-cap-smll.htm

8 www.blackrock.com/br/products/251902/ishares-sp-500-fi-em-cotas-de-fundo-de-ndice-inv-no-exterior-fund

9 www.spglobal.com/spdji/pt/indices/equity/sp-500/#overview

utilizados para a criação de ETFs nesses mercados. Dentre eles, na bolsa de valores americana destacam-se:

- EEM[10]: o iShares MSCI Emerging Markets é um ETF que investe no índice MSCI Emerging Markets, formado pelas maiores empresas dos países emergentes. Esse fundo investe um pequeno percentual (inferior a 10%) nas maiores empresas brasileiras.
- IAU[11]: o iShares Gold é um ETF que investe em ouro, um dos investimentos mais antigos do mundo. Esse instrumento acaba sendo uma forma acessível do investidor ter uma posição em ouro, sem depender de negociar contratos de compra e venda de ouro cujos valores são elevados.

Na avaliação do desempenho de um ativo no mercado financeiro, utilizam-se algumas métricas como a rentabilidade, correlação e o CAGR (*Compound Annual Growth Rate*) (CHRISTY J. et al., 2021). A rentabilidade de um investimento é medida pela diferença percentual entre o investimento final e o inicial. O objetivo de todo e qualquer investidor é obter a maior rentabilidade possível em seus investimentos, sem se expor a grandes perdas.

A rentabilidade também pode ser calculada de forma periódica, tendo destaque o cálculo do CAGR, taxa anual composta de crescimento. Essa métrica informa o crescimento anual do investimento na forma de juros compostos, sendo mais significativa que a média anual da rentabilidade. Seu valor é calculado de acordo com a Equação 1:

$$CAGR = \left(\frac{Valor\ Final}{Valor\ Inicial}\right)^{\frac{1}{n}} - 1, sendo\ n\ o\ número\ de\ anos \quad (1)$$

Existem momentos em que o mercado enfrenta fortes crises e um determinado tipo de ativo pode acabar sendo impactado mais fortemente do que outro. Considerando como exemplo a queda da bolsa ocorrida em virtude da pandemia em março de 2020, alguns ativos tiveram uma maior queda do que outros. Ativos relacionados ao setor de serviço, turismo e aviação tiveram uma forte queda, ao passo que ativos do setor de e-commerce se recuperaram rapidamente.

Uma das formas de se proteger contra riscos nos investimentos é através da diversificação. Possuir ativos diferentes, como ações de setores diferentes, pode ser uma vantagem, uma vez que, cada setor pode ter um desempenho diferente, impedindo todo o portfólio de sofrer o mesmo prejuízo. Na construção de portfólios, os investidores utilizam uma

[10] www.ishares.com/us/products/239637/ishares-msci-emerging-markets-etf

[11] www.ishares.com/us/products/239561/ishares-gold-trust-fund

importante métrica que é a correlação entre os ativos. Essa métrica visa indicar como um ou mais ativos estão relacionados, em relação às variações de suas cotações. Variando entre -1 e 1, a correlação entre dois ativos é forte quando o preço de ambos tende a se movimentar para o mesmo lado. Correlação próxima a 0 indica nenhuma correlação, logo os movimentos dos preços não se relacionam. Uma correlação negativa próxima a -1 indica movimentos relacionados no preço só que de forma oposta.

A diversificação se faz presente no momento em que os ativos possuem uma baixa correlação entre si, ou até correlação negativa em alguns casos. Dessa forma, alguns ativos que possam passar por um período de baixa poderão ser compensados por outros ativos que estejam em alta. O contrário da diversificação é a concentração. Um exemplo seria possuir diversos ativos semelhantes ou do mesmo setor. Dessa forma, eles tendem a ter uma alta correlação positiva e, normalmente, quando um se encontra em momentos difíceis, os outros o seguem. Assim, o risco de ter uma carteira concentrada é maior, uma vez que algum evento pode acabar atingindo esses ativos e levar a carteira como um todo a ter um resultado negativo.

2.3 Macroeconomia

Uma noção básica de economia é sempre muito bem vinda para a sociedade. Ter a compreensão de algumas das variáveis que afetam de diferentes formas a vida da população é importante para o planejamento financeiro pessoal. Nesta seção serão apresentados algumas das principais variáveis macroeconômicas do Brasil.

A SELIC[12] é a taxa de juros da economia brasileira, sendo utilizada como taxa básica em diversas aplicações do mercado. Reajustada conforme decisão do Banco Central do Brasil (BCB), essa taxa ajuda no controle da inflação e na expansão do crédito. Como a SELIC é utilizada como indexador para grande parte dos empréstimos e das dívidas, uma redução dessa taxa acaba por aliviar o endividamento e facilitar a concessão de novos créditos, ao passo que sua elevação acarreta por diminuir a liberação de crédito, que por sua vez auxilia na retração da inflação.

[12] www.bcb.gov.br/controleinflacao/taxaselic

A inflação oficial no país é medida pelo índice de preços ao consumidor amplo (IPCA[13]), que corresponde a uma cesta de produtos e serviços cujos preços são analisados mensalmente em diversas cidades do país. A partir da média desses preços, calcula-se a variação mensal. A inflação também acaba por refletir o poder de compra da moeda, que aumenta em caso de deflação (inflação negativa) e cai em caso de inflação. O BCB estabelece uma meta de inflação para cada ano e visa, através de ajustes da taxa SELIC, manter essa taxa dentro da meta estabelecida.

A taxa de câmbio[14] é um importante indicador para a economia brasileira. Sendo o dólar a moeda mais utilizada em transações internacionais e no comércio mundial, seu valor pode influenciar fortemente a economia brasileira. Um dólar alto favorece as empresas exportadoras e que geram receitas no mercado internacional, ao passo que encarece a importação de produtos e insumos, aumentando os preços das mercadorias pagas pela população local. Um dólar baixo acaba por favorecer as importações, reduzindo o custo das mercadorias, favorecendo o consumo de produtos importados, ao passo que diminui as receitas de exportadores. A dinâmica do câmbio é complexa, mas pode se resumir ao princípio de oferta e demanda. Conforme mais pessoas compram dólares, mais o preço tende a aumentar, enquanto que um aumento das vendas de dólares para compra de real por parte de investidores internacionais acarreta numa queda da cotação do dólar.

Na economia é comum haver índices que são indexados às variáveis macroeconômicas. No cenário brasileiro destacam-se o CDI[15] e o IMA-B[16]. Os certificados de depósito interbancário (CDI) são uma forma de investimento em renda fixa que utiliza a taxa DI[17], uma taxa cobrada entre instituições financeiras para empréstimos. A taxa DI utilizada é muito próxima da taxa SELIC e tornou-se o referencial básico do custo das operações interbancárias, servindo de referência no mercado financeiro para a determinação da rentabilidade de diversos ativos de renda fixa.

---

[13] www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=o-que-e

[14] www.bcb.gov.br/acessoinformacao/perguntasfrequentes-respostas/faq_taxacambio

[15] blog.xpeducacao.com.br/cdi-o-que-e-como-funciona-e-impacta-os-seus-investimentos/

[16] www.anbima.com.br/pt_br/informar/precos-e-indices/indices/ima.htm

[17] www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/serie-historica-do-di.htm

Outro indexador utilizado em aplicações financeiras é o IMA-B, um índice que representa um conjunto de títulos públicos atrelados à inflação oficial do país. O IMAB acompanha, portanto, os rendimentos de uma carteira de títulos do Tesouro Nacional indexados ao IPCA, com diversas durações.

Para uma melhor compreensão da situação econômica do país, dois indicadores são imprescindíveis e, constantemente, são destaques na mídia, o Produto Interno Bruto[18] (PIB) e a taxa de desemprego. O PIB é um indicador macroeconômico que mede a quantidade de bens e serviços produzidos na economia de um país. Através do PIB consegue-se avaliar a condição econômica de um determinado país, medindo-se sua geração de riqueza. É calculado trimestralmente pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O IBGE também realiza a medição da taxa de desemprego mensal, realizada desde 2013 através da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC[19]).

2.4 Finanças Comportamentais

A evolução das ciências sociais ocasionou a criação de um novo campo de pesquisa na área econômica. Diversos psicólogos e cientistas sociais passaram a estudar o comportamento humano e como ele afeta a economia. O surgimento da chamada economia computacional a partir da metade do século XX trouxe um novo olhar para a economia tradicional. Partindo do princípio que as pessoas não tomam somente decisões racionais, esse campo de estudo buscou observar como se dariam os efeitos econômicos a partir dos diferentes comportamentos humanos, muitas vezes influenciados por vieses.

Dentre os grandes pesquisadores responsáveis pela evolução desta área, destacam-se os psicólogos Amos Tversy e Daniel Kahneman (Nobel de economia em 2010) e o economista Richard Thaler (Nobel de economia em 2015). Com o avanço das pesquisas em diversos ramos da economia, houve o surgimento do campo de Finanças Comportamentais. Nesse campo é avaliado o impacto do comportamento humano, principalmente do investidor, em tomadas de decisão sobre investimentos.

---

[18] www.ibge.gov.br/explica/pib.php

[19] www.ibge.gov.br/estatisticas/sociais/trabalho/17270-pnad-continua.html?=&t=o-que-e

Em um de seus livros mais famosos, o pesquisador Robert Schiller (2015) discute os diferentes vieses presentes nos investidores do mercado financeiro. Através das suas pesquisas, ele discute a eficiência dos mercados, a formação de bolhas especulativas e alguns efeitos adversos sobre o psicológico do investidor. Trazendo diversas publicações na área, o professor Schiller ilustra alguns fenômenos interessantes observados no mercado americano, como o efeito de sazonalidade.

Efeitos de sazonalidade no mercado financeiro podem ser explicados como um comportamento atípico e normalmente recorrente em determinado período do ano que acaba se repetindo ao longo de vários anos. Esses efeitos caracterizam-se por movimentos na bolsa de valores ligados ao comportamento do investidor em alguns períodos do ano, em especial, o início do período de férias e as festas de final do ano. Cabe um destaque para dois movimentos em particular, a euforia de final de ano e o efeito Halloween ou SMGA, "*Sell in May and Go Away*" (BOUMAN, S. et al. 2002).

A euforia de final de ano é caracterizada pelo comportamento do investidor em relação ao mercado financeiro ao longo dos últimos meses do ano, em especial dezembro. Observa-se que há uma tendência de retornos mensais positivos nos últimos meses do ano, fato que sugere uma atitude mais otimista e compradora dos investidores.

O efeito SMGA refere-se a uma tendência dos investidores venderem suas posições antes do início do verão no hemisfério norte, voltando a comprá-las ao final do verão. Essa pressão vendedora traz, como resultado, retornos negativos para os meses de maio a setembro.

2.5 Previsão de tendências

Prever acontecimentos ou resultados futuros sempre foi um motivador para o desenvolvimento de pesquisas na área de inteligência artificial. Através do estudo e interpretação de dados passados, busca-se encontrar semelhanças que indiquem a repetição de um comportamento em uma ocorrência futura. Diversas técnicas de aprendizado de máquina são utilizadas na busca de uma previsão com maior acurácia. Esses algoritmos necessitam de grandes bases de dados para conseguirem encontrar padrões mais perceptíveis que ajudem a apontar uma tendência. Estudos de previsão de tendências do preço de ações vêm utilizando técnicas de ML na construção de classificadores (LUO R., 2020).

A maioria das técnicas pode ser consideradas como de aprendizado supervisionado, na qual se utiliza para treinamento uma parte dos dados já contendo o resultado esperado, diferentemente das técnicas não supervisionadas que não necessitam de treinamento. Técnicas de aprendizado supervisionado visam um mapeamento dos dados de entrada para saídas esperadas conforme definidas em dados de treinamento. Elas são divididas em técnicas de regressão, que visam estudar a relação entre a variável de saída dependente e as variáveis de entrada independentes, através de operações matemáticas e vetoriais, e técnicas de classificação que visam encontrar a qual categoria a variável de saída pertence de acordo com as variáveis de entrada.

A seguir são comentadas as técnicas utilizadas nesta dissertação a partir da revisão bibliográfica realizada e que será apresentada no próximo capítulo. Como o interesse do trabalho está na utilização das técnicas e não em sua implementação, a fundamentação matemática de cada uma delas não será abordada.

2.5.1 Random Forest (RF)

Os algoritmos conhecidos como *Random Forest* são uma classe especial de árvores de decisão. A função de uma árvore de decisão é de analisar as entradas e, através de regras definidas em cada nível da árvore, chegar a uma conclusão. O algoritmo Random Forest utiliza múltiplos conjuntos de árvores de decisões para obter um resultado. Cada árvore apresentará uma configuração diferente e o conjunto de diversas dessas árvores irá convergir para um resultado mais aceito por todos. O algoritmo utiliza a ideia de sabedoria da multidão, ao gerar uma resposta que represente a decisão da maioria das árvores individuais. Nesse mesmo algoritmo é importante destacar a utilização de configurações de árvores com baixa correlação entre si, aumentando, portanto, a diversificação e gerando, assim, um resultado mais fidedigno (LUO R., 2020) (OUAHILAL M. et al., 2016).

2.5.2 Naive Bayes (NB)

O teorema probabilístico de Bayes calcula a probabilidade condicional de eventos aleatórios A e B, onde P(A|B) é a probabilidade de A ocorrer se B ocorrer (LUO R., 2020). A aplicação do teorema de Bayes permitiu a criação de uma classe de classificadores conhecidos como Bayesianos, de onde Naive Bayes é um deles. São algoritmos de classificação estatísticos,

também conhecidos como probabilísticos, uma vez que consideram a independência das variáveis de acordo com o teorema.

Existem diferentes configurações do algoritmo para determinados conjuntos de dados, como Gaussiana, Multinomial e Bernoulli. A primeira lida com entradas que possuam uma distribuição gaussiana ou normal, a segunda com distribuições multinomiais, enquanto a terceiro com a distribuição multivariada de Bernoulli (NBB). Essa última distribuição utiliza entradas que são valores binários.

2.5.3 Support Vector Machine (SVM) e Support Vector Regression (SVR)

Uma máquina de vetores de suporte é um algoritmo que visa classificar um conjunto de dados em duas classes distintas. Baseando-se no modelo linear, o algoritmo busca encontrar um hiperplano que melhor sirva para separar as duas classes uma da outra (GONZALEZ R. T. et al., 2015). Essa técnica visa minimizar o erro ao selecionar o melhor plano que melhor classifique um conjunto de entrada.

Vetor de suporte de regressão (SVR) é um algoritmo da família do SVM para regressão. De forma semelhante ao SVM, busca encontrar um hiperplano que divida as duas classes, contudo, o algoritmo irá tentar minimizar a soma dos erros quadráticos, calculados a partir da distância entre os pontos e o hiperplano (OUAHILAL M. et al., 2016). Por ser um algoritmo de regressão, buscará encontrar um valor e não uma classe à qual esse valor pertence (HENRIQUE B. et al., 2018).

2.5.4 K-Nearest Neighbors (KNN)

O algoritmo KNN é um algoritmo de classificação que visa encontrar a melhor classe para um determinado dado de entrada. Ele se baseia na semelhança entre um novo dado e seus K vizinhos mais próximos. Esse novo dado irá ser classificado na mesma classe que seus K vizinhos, através do cálculo da distância entre o novo dado e as classes já presentes. A definição de K costuma ser um dos pontos de partida do algoritmo, onde através de alguns experimentos iniciais, busca-se encontrar um valor otimizado para K (TANG L. et al., 2018).

2.5.5 Regressão Logística (RLOG)

Utilizada em análises econômicas e em grandes conjuntos de dados, a técnica de regressão logística pode ser considerada uma generalização da regressão linear (WANG H., 2020). A regressão logística é um método de regressão mais voltado para classificação de variáveis categóricas. Baseia-se no método de regressão linear, porém utiliza um modelo matemático diferente, utilizando a função logística (curva em S), em vez da função linear. Da mesma forma que a regressão linear, esse algoritmo busca encontrar uma aproximação que melhor represente uma determinada entrada de acordo com as categorias disponíveis.

2.5.6 Multilayer Perceptron (MLP)

Algoritmos de redes neurais podem ser associados ao sistema nervoso humano, composto por diversos neurônios e que combinados processam informações. Sendo uma área de estudo não tão recente, vem produzindo diversas contribuições ao longo dos últimos anos, sendo aplicadas para solucionar diversos problemas, dos mais simples aos mais complexos, formados por uma grande quantidade de variáveis (REYHANI R. et al., 2011). Podem ser aplicadas tanto para problemas de classificação quanto para de regressão.

As redes neurais são formadas por nodos que servem como entrada e que vão se ligar e outros (REYHANI R. et al., 2011) nodos de diferentes camas, formando assim uma rede. O resultado de cada nodo é obtido a partir dos nodos anteriores. As camadas ocultas internas são responsáveis por aplicar operações sobre os dados que serão transmitidos para as próximas camadas até chegar aos nodos de saída.

Dentre os algoritmos de redes neurais artificiais, os perceptron de multicamadas (MLP) são redes neurais diretas que possuem várias camadas ocultas. Os neurônios dentro da rede possuem pesos que serão ajustados conforme o treinamento e essa rede pode possuir mais de uma saída. Esse tipo de rede neural utiliza a retropropagação para atualizar os pesos dos nodos nas camadas ocultas e na saída (OJO S. et al., 2019), melhorando assim o resultado da classificação.

2.6 Considerações finais

Este Capítulo discutiu os fundamentos teóricos que auxiliaram na execução deste trabalho. Inicialmente, foram apresentados os conceitos de séries temporais, que são o ponto fundamental da análise da primeira etapa do modelo proposto. Importantes conceitos de mercado financeiro, finanças comportamentais e macroeconomia foram trazidos à tona, uma vez que a análise multidisciplinar deste trabalho engloba estes três domínios.

Por fim, foram mencionadas as técnicas de ML comuns na literatura e que são utilizadas na construção do algoritmo proposto na segunda etapa deste trabalho. As configurações de cada uma dessas técnicas serão apresentadas no Capítulo 4.

No Capítulo seguinte será mostrado um resumo dos principais estudos relacionados assim como um estudo comparativo entre eles. Por fim, apresentará as principais escolhas feitas por este trabalho em relação à literatura.

3 TRABALHOS RELACIONADOS

Este Capítulo apresenta um resumo dos principais trabalhos que motivaram e contribuíram para a metodologia proposta nesta dissertação. O Capítulo está dividido em três seções: a primeira abordando os trabalhos relativos à fase de análise de séries temporais financeiras; a segunda apresentando os trabalhos relacionados às técnicas de previsão de tendências no mercado de ações e, por fim, algumas considerações acerca do que foi aproveitado para o modelo proposto nesta dissertação.

3.1 Trabalhos relacionados a séries temporais

Os estudos de séries temporais estão presentes em uma variedade de áreas. Na economia e no setor financeiro, destacam-se trabalhos sobre a evolução de preços e cotações de diversas variáveis. Ao mesmo tempo, diversos desses trabalhos fazem uso de importantes conceitos estatísticos para analisar esse conjunto de dados.

J. Christy Jackson, J. Prassanna, Md. Abdul Quadir, V. Sivakumar (2021) buscaram prever o cenário futuro do mercado através do levantamento de dados de séries temporais ao mesmo tempo que realizaram um estudo sobre o comportamento dos dados utilizados. Esses dados foram ilustrados através de análises estatísticas que representam alguns dos principais conceitos utilizados no mercado financeiro como CAGR, médias móveis, índice Sharpe, retorno acumulado, correlação entre outros. O trabalho é finalizado com a tentativa de se prever o comportamento futuro desses ativos através do método estatístico ARIMA (modelo autorregressivo integrado de médias móveis), mas os resultados obtidos não foram muito promissores, uma vez que a cotação das ações não tende a seguir um modelo linear.

As variáveis macroeconômicas costumam ser utilizadas para se analisar o comportamento da economia e do mercado financeiro. Salisu A., Vo X. (2021) apresentaram um estudo sobre o comportamento da taxa de câmbio e da bolsa de valores em função das taxas de juros. Neste trabalho, foi analisado como o ambiente econômico distinto entre países desenvolvidos e em desenvolvimento impacta tanto na cotação do câmbio quanto na performance do mercado acionário. Da mesma forma, analisou-se a relação entre câmbio e ações. Toraman C., Başarir C. (2014) avaliaram a relação de longa duração entre o

comportamento do mercado acionário turco com as taxas de juros da economia local. O trabalho mostrou como essas duas variáveis estiveram relacionadas ao longo de catorze anos.

McMillan D. G. (2021) analisou qual o impacto dos títulos do tesouro americano de diferentes durações na economia americana. Através da análise de diferentes títulos que contemplam diferentes expectativas de taxas de juros, avaliou-se o resultado do crescimento no mercado imobiliário, no consumo, nas dívidas de empresas e em *commodities* durante diversos períodos. O estudo destacou subamostras relacionadas às taxas de juros em diferentes períodos e comparou os resultados do crescimento de diversos setores assim como do mercado de ações.

Um panorama do papel da política monetária e fiscal brasileira foi apresentado por Afonso J. R., Araújo E. C., Fajardo B. G. (2016). O estudo teve por objetivo ilustrar as políticas fiscais vigentes no país desde a criação do plano Real. Ao longo do trabalho, os autores apresentaram o comportamento das principais variáveis macroeconômicas com a SELIC, o IPCA e a taxa de câmbio ao longo do período. Complementar a esse trabalho, Stona F., Morais I. A. C., Triches D. (2018) verificaram quais as principais medidas fiscais e monetárias foram adotadas em períodos de crise na economia brasileira e quais as suas consequências. O trabalho ainda buscou criar um índice para a identificação do nível de estresse do mercado brasileiro. Esse índice ajudou a identificar períodos nos quais o risco de se investir no país aumenta. O trabalho também mostrou a evolução de alguns indicadores macroeconômicos como SELIC e inflação.

Há também trabalhos que utilizam cotações internacionais para analisar o impacto do preço destas no mercado financeiro. An Y., Sun M., Gao C., Han D., Li X. (2020) estudaram o impacto do preço do barril de petróleo em dois setores do mercado chinês, dividindo a análise em períodos de forte alta, estabilidade e forte queda e leve queda. Os autores verificaram como a volatilidade desses setores é influenciada pelas cotações do petróleo de forma diferente em cada um dos diferentes períodos.

A evolução das finanças comportamentais a partir das últimas décadas do século XX trouxe vários estudos que debateram a racionalidade dos investidores e a eficiência dos mercados. Ahmed B. (2020) apresentou um estudo sobre o sentimento dos investidores no mercado financeiro americano, examinando algumas hipóteses. Em resposta aos trabalhos que afirmam que investidores negociam de forma irracional e movidos a ruídos, o autor acredita que há fatores que os levam a tomar tais decisões, estas sendo orientadas por sentimentos. O autor observou uma importante relação entre movimentos orientados por sentimento em mercados de baixa.

Efeitos sazonais estão presentes na literatura, onde pesquisadores analisam esses comportamentos em diferentes localizações geográficas. Plastun A., Sibande X., Gupta R., Wohar M. E. (2020) também realizaram uma análise acerca de diferentes anomalias mensais como efeitos sazonais em dezembro, janeiro e outubro. Os autores fizeram uma análise estatística histórica do mercado americano, assim como em algumas bolsas de valores globais. Os resultados mostraram a evolução e a intensidade de cada efeito em cada um dos países analisados, destacando-se, por exemplo, que o efeito janeiro foi bastante presente nos Estados Unidos durante a metade do século XX, enquanto os efeitos outubro e dezembro estiveram mais presentes no resto do mundo.

Efeitos sazonais podem ter relação com diversos fatores que causam uma mudança na percepção de risco dos investidores, seja um recebimento de uma verba extra ou momentos de euforia ou medo. No estudo realizado por Chen Z., Daves P. R. (2018), verificou-se como o sentimento do investidor nos meses de janeiro pode impactar no desempenho do portfólio do investidor ao longo do ano. Partindo-se da premissa que no início do ano os investidores tendem a estar tanto mais otimistas quando a situação econômica para o ano parecer favorável, quanto mais pessimista quando do oposto. Dessa forma, os autores analisaram o retorno ao longo do resto do ano baseado no índice de sentimento do consumidor durante os meses de janeiro ao longo de quarenta anos. Os resultados mostraram que existe uma relação positiva tanto em direção quanto em magnitude entre o sentimento percebido nos meses de janeiro e o retorno obtido no resto do ano.

Chen T., Chien C (2011) analisaram como as características culturais de um país podem levar ao efeito janeiro em países como China e Taiwan. A celebração do ano novo chinês nos meses de janeiro e fevereiro faz com que diversos cidadãos recebam algum tipo de bônus nesses meses e com esse dinheiro extra os autores observaram um aumento no investimento em ativos mais arriscados, normalmente empresas de baixa capitalização.

O efeito *Halloween* ou SMGA pode ser observado como um movimento de pressão vendedora entre maio e outubro, onde os retornos do mercado ao longo desse período tendem a ser menores que o retorno após o *halloween*. Zhang C. Y., Jacobsen B. (2021) estudaram o efeito SMGA em diversos mercados globais, analisando diversos índices de ações. Eles perceberam que o período de novembro a abril apresentou um retorno em média 4% superior ao retorno do período compreendido entre maio e outubro, ao longo dos anos. Também observaram que na grande maioria dos casos os retornos tendem a ser negativos ou estáveis.

Um resultado oposto foi observado por Dichtl H., Drobetz W (2015), os quais verificaram que em análises mais recentes o efeito SGMA tende a desaparecer. Nesse trabalho, os autores buscaram utilizar uma estratégia que tirasse proveito do efeito SGMA para aumentar os ganhos, mas não obtiveram sucesso nos anos mais recentes, impactado principalmente pela queda da diferença entre os períodos. Os estudos também buscaram avaliar um conjunto de mercados globais.

Com o entendimento dos efeitos sazonais, alguns pesquisadores buscam criar estratégias para aproveitar essas anomalias e obter maiores ganhos no mercado financeiro. Guo B., Luo X., Zhang Z. (2014) analisaram o comportamento SGMA e os efeitos janeiro e fevereiro no mercado chinês, ao longo de 17 anos. A estratégia montada pelos autores conseguiu se beneficiar dessas anomalias e obteve resultados superiores às estratégias tradicionais de investimento, ao mesmo tempo que enfrentou de forma mais amena as crises financeiras no período.

A diversificação é um fator muito considerado por investidores. Investir em diferentes países e em diferentes classes de ativos permite diminuir os riscos dos investimentos ao não concentrar todos os investimentos em apenas poucas classes. Spierdijk L., Umar Z. (2014) apresentaram um estudo sobre a diversificação geográfica em investir em mercados emergentes e nas moedas locais. Através das análises sobre os dados históricos, os autores verificaram o desempenho de investidores locais e investidores estrangeiros num período de 13 anos em vinte bolsas de valores diferentes. Como resultado, concluiu-se que investimentos em países emergentes no curto prazo acabam por ser uma alternativa viável para se obter mais ganhos, ao mesmo tempo que a diversificação geográfica acaba apresentando a melhor relação risco-retorno.

Kirikkaleli D. (2020) analisou os impactos causados por uma série de fatores de risco, como político, econômico e fiscal no desempenho da bolsa de valores. Baseando-se em indicadores de risco, tanto locais quanto globais, o pesquisador buscou entender de que forma os diferentes tipos de risco estão relacionados ao desempenho da bolsa de valores de Taiwan. O autor descobriu que a estabilidade da economia local é um fator positivo para a bolsa local, ao passo que é necessário encontrar formas de se proteger das crises globais.

O papel do investimento em ouro como proteção de carteira nos mercados americano e inglês foi objeto de estudo por He Z., O'Connor F., Thijssen J. (2018). Os autores analisaram cinquenta anos de dados históricos da cotação do ouro e como ele se comportou em momentos de crise financeira. Os resultados mostraram a resiliência desse tipo de investimento e como

sua utilização como forma de diversificação e proteção de carteira obteve boa resposta em momentos de maior estresse do mercado financeiro.

A Tabela 3.1 mostra um comparativo entre os trabalhos vistos nesta seção. Além de estarem classificados conforme o domínio, é apresentado o destaque de cada um dos trabalhos em relação à temática proposta por esta dissertação.

Tabela 3.1 – Resumo dos trabalhos sobre análise de séries temporais.

Trabalhos	*Domínio*	*Destaque*
Salisu A. et al. (2021)	Macroeconomia	Câmbio, Juros e Bolsa
Toraman C. et al. (2014)	Macroeconomia	Câmbio e ações
McMillan D. G. (2021)	Macroeconomia	Juros em diferentes períodos
Afonso J. R. et al. (2016)	Macroeconomia	Selic, IPCA, câmbio no Brasil desde plano Real
Stona F. et al. (2018)	Macroeconomia	Selic, inflação em períodos de crise
An Y. et al. (2020)	Macroeconomia	Barril de petróleo com divisão de períodos
Ahmed B. (2020)	Comportamental	Sentimento no EUA. Investidor irracional.
Plastun A. et al. (2020)	Comportamental	Efeitos sazonais (JAN, OUT, DEZ) nos EUA e bolsas globais
Chen Z. et al. (2018)	Comportamental	Efeito Janeiro e relação positiva em direção e magnitude
Chen T. (2011)	Comportamental	Efeito Ano Novo chinês
Zhang C. Y. et al. (2021)	Comportamental	SMGA em diversos mercados
Dichtl H. et al. (2015)	Comportamental	SMGA desaparece em períodos mais recentes
Guo B. (2014)	Comportamental	SGMA e efeito início de ano na China
Spierdijk L. et al. (2014)	Outros índices	Diversificação em outros países
Kirikkaleli D. (2020)	Outros índices	Riscos na bolsa local
He Z. et al. (2018)	Outros índices	Investimento em ouro

Fonte: Dos Autores.

Destes estudos, cabe destacar a divisão das séries temporais em períodos de análise como períodos de crise, eleições, grandes altas e grandes quedas. Também se destaca a utilização de um conjunto das principais variáveis macroeconômicas como taxa de câmbio, juros e inflação.

Verifica-se a presença de efeitos sazonais conhecidos como SMGA e euforia de final de ano, observados em diferentes bolsas do mundo. Comparações com índices estrangeiros de

países emergentes, dos principais índices de países desenvolvidos como S&P e o índice de bolsas europeias mostraram-se bastante relevantes, em especial o uso histórico do ouro em momentos de crise.

3.2 Trabalhos relacionados a previsão de tendências

A utilização de algoritmos de ML para previsão de tendências é uma área composta por diversos artigos que exploram diferentes maneiras de aumentar a precisão das previsões. Destacam-se também trabalhos que buscam empregar diferentes técnicas tanto para diferentes etapas quanto para comparar resultados isolados.

Alguns trabalhos optaram por construir um algoritmo com diversas etapas, empregando diferentes técnicas em cada uma delas. Liu H., Long Z. (2021) procuraram prever o comportamento de alguns índices do mercado como o S&P 500 através de uma combinação de técnicas avançadas que atuam sobre os dados em uma sequência de etapas. Em cada etapa uma técnica diferente foi utilizada, gerando entradas para as novas fases e em alguns casos servindo como *feedbacks* para as fases anteriores. Destaca-se a utilização de técnicas de decomposição, LSTM e *Deep Learning*. Os resultados finais foram bastantes promissores, uma vez que o gráfico da cotação futura obtida estava muito próximo do gráfico da cotação real.

As combinações de algoritmos e fatores econômicos podem apresentar resultados satisfatórios, como no trabalho de Luo R. (2020), na qual o autor criou um modelo de previsão de preços de ações para a bolsa de Shangai. O autor começou analisando as vantagens e desvantagens da utilização de diferentes técnicas e os seus resultados. As técnicas utilizadas foram RF, *Adapting Boost Algorithm*, NB, ARIMA, *Prophet model* (fpprophet), LSTM e *Temporal Convolutional Network* (TCN). Ao final, apresentou um modelo que utiliza alguns índices macroeconômicos para prever as cotações das ações e que obteve bons resultados em análises de curto prazo.

Um dos algoritmos considerados por vários trabalhos como um dos mais adequados para a previsão de tendências foi o LSTM. Huang B., Ding Q., Sun G. (2018) propuseram um algoritmo adaptado de LSTM para predição dos preços da bolsa chinesa. Nesse trabalho, realizou-se a uma otimização do algoritmo LSTM com a adição do algoritmo bayesiano. O algoritmo utilizou sete entradas da série histórica como os valores de cotação, máxima, mínima, abertura, fechamento, volume e tempo. Ao mesmo tempo, o trabalho buscou analisar um

intervalo de tempo maior, agregando os dados em períodos de uma semana. O resultado encontrado para essa nova abordagem foi 25% superior ao modelo tradicional de LSTM.

Uma outra adaptação do algoritmo LSTM foi realizada por Ojo S., Owolawi P., Mphahlele M., Adisa J. (2019), visando a predição dos preços futuros do mercado de ações empregando o índice IXIC composto por quase todas as ações da bolsa americana NASDAQ. A modificação adotada pelos pesquisadores baseia-se no empilhamento de blocos contendo o método LSTM. Dessa forma, as saídas de um estarão conectadas às entradas de outro. Comparando-se com o método tradicional, o resultado obtido foi superior.

Um método que se aproxima do LSTM para previsões de curto prazo que foi bastante destacado por pesquisadores é o estatístico ARIMA. Joosery B., Deepa G. (2019) realizaram a predição dos valores de ações utilizando esses dois métodos, ARIMA e LSTM. Para o último, também foi desenvolvido uma adaptação chamada de *Attention*. Os autores observaram que o método ARIMA teve melhor desempenho quando utilizado grandes *datasets* (maiores que 1 ano) e o LSTM e sua variação tiveram melhor desempenho para *datasets* menores.

A utilização de ANN (redes neurais artificiais) também é bastante comum para encontrar tendências nos preços dos ativos. Giacomel F., Pereira A.C.M., Galante R. (2015) utilizaram um modelo composto por um conjunto de ANN para prever a cotação de algumas ações. Nesse modelo duas redes neurais são utilizadas com as mesmas entradas, mas com configurações que podem ser distintas, e o resultado é obtido através de fusão de ambas. No final os autores simularam o resultado de uma carteira de investimentos utilizando algumas estratégias conhecidas do setor e o modelo proposto. O modelo proposto teve um desempenho satisfatório, sempre apresentando lucro para os diferentes portfólios montados.

Uma das formas dos pesquisadores verificarem seus resultados consiste em utilizar os resultados dos seus experimentos para simular operações reais de compra e venda de ativos e verificar o desempenho final versus algum *benchmark*. J. de Mello Assis, A. C. M. Pereira, R. C. e Silva (2018) buscaram avaliar um conjunto de estratégias de *trading* baseadas em conjuntos de ANN. Utilizando diferentes modelagens e configurações de ANN, os autores buscam encontrar as melhores estratégias para negociações na bolsa de valores brasileira. Os autores selecionam as melhores ANN e utilizam esse conjunto para adaptar uma estratégia de compra e venda de ativos. Como resultado, os pesquisadores obtiveram, para todas as estratégias, ganhos positivos no período analisado.

Para prever o comportamento futuro do mercado de ações chinês Wang H. (2020) utilizou duas abordagens de aprendizado de máquina, regressão linear e SVM. Ele analisou um

conjunto de 10 anos de cotações do mercado chinês e buscou utilizar esses algoritmos para montar uma estratégia de seleção de ativos. Os resultados mostraram que ambos os métodos conseguiram gerar bons portfólios que acabaram por superar o desempenho médio do mercado.

É comum na literatura encontrar trabalhos que tentam relacionar os movimentos nas bolsas de valores com base nas informações que circulam na mídia. Para prever o comportamento futuro da bolsa de valores Tang J., Chen X. (2018) desenvolveram um modelo híbrido que leva em conta tanto as cotações quanto as notícias diárias. Nesse modelo, os autores utilizaram uma adaptação do LSTM, RNN-LSTM (*Recurrent Neural Network Long Short Term Memory*) para a análise das cotações ao mesmo tempo em que utilizaram CNN (*Convolutional Neural Networks*) para a análise das notícias. As saídas dos dois algoritmos foram combinadas para a geração da predição. Utilizando o índice *Dow Jones Industrial Average*, os resultados obtidos pelo modelo híbrido superaram os resultados dos modelos individuais, na análise de um período de sete anos.

Huynh H., Dang L., Duong D. (2018) buscaram agregar o conhecimento fornecido pelas notícias sobre o mercado financeiro com a previsão de preços futuros. Para isso, eles desenvolveram um algoritmo que utiliza a polaridade de um conjunto de notícias em um determinado período para ajustar a previsão de tendências do preço. Utilizando um modelo chamado *Bidirectional Gated Recurrent Unit* (BGRU) os autores obtiveram um resultado superior aos modelos tradicionais usando LSTM e *Gated Recurrent Unit* (GRU), tanto para previsão do mercado como um todo quanto para ações individuais.

Wang Z., Ho S., Lin Z. (2018) propuseram um método de predição de preços da bolsa de valores utilizando o sentimento das notícias. Os artigos de notícias foram filtrados e processados através de algoritmos de processamento de linguagem natural de modo a indicar o sentimento do conteúdo da notícia. O *score* atribuído ao sentimento de cada notícia foi utilizado juntamente com a série temporal dos preços das ações através de algoritmos de aprendizado com redes neurais. Os autores construíram diferentes janelas temporais de notícias para identificar qual o tamanho que apresentava maior precisão, obtendo resultados que superaram os modelos de base.

Visando a otimização dos resultados, alguns trabalhos sugerem a realização de alguns tratamentos especiais nos dados, seja através da remoção de redundâncias, filtragem, aplicação de teoria dos grafos e teoria do caos. Tang L., Pan H., Yao Y. (2018) utilizaram um algoritmo de KNN com PCA (*Principal Component Analysis*) para a predição de séries temporais financeiras. Eles utilizaram o processo de PCA para reduzir a redundância de informações de

forma a gerar entradas mais enriquecidas para o algoritmo de KNN. Através de experimentações sobre os dados da taxa de conversão euro para dólar e do índice do mercado de ações chinês num período de 10 anos, os autores obtiveram uma taxa de acerto maior em comparação ao algoritmo KNN sem PCA.

Em outro trabalho que utilizou a técnica KNN, Khattak A., Ullah H., Khalid H., Habib A., Asghar M., Kundi F. (2019) criaram um algoritmo de previsão de tendências de ações que utiliza esta técnica para tratamento dos dados, visando a redução da dispersão, ao identificar pontos fora da curva no conjunto de dados. Os resultados encontrados foram promissores e obtiveram melhor desempenho em relação a algoritmos sem tratamento de dados.

Ouahilal M., Mohajir M., Chahhou M., Mohajir B. (2016) propuseram uma abordagem híbrida para a predição do preço das ações combinando SVR com filtros Hodrick-Prescott. O filtro teve a função de otimizar as entradas do algoritmo SVR, filtrando o ruído. Utilizando como base a cotação de uma empresa, os autores obtiveram êxito em sua proposição que superou o desempenho do algoritmo SVR normal e de outras adaptações do SVR com filtros diferentes.

Patil P., Wu C., Potika K., Orang M. (2020) utilizaram a teoria dos grafos para construir um modelo para previsão de tendências de preços de ações. Analisando o comportamento espaço-temporal das relações entre diferentes ações, os autores modelaram o mercado de ações como um sistema complexo. Dois modelos de grafos foram utilizados, um de correlação entre os dados e outro baseado em casualidade que utiliza notícias relacionadas. Os autores acabaram avaliando três modelos, GCN (*Graph Convolutional Network*) baseado em modelos de *Deep Learning* para grafos, modelo de regressão linear e modelo estatístico utilizando o método ARIMA. Os resultados mostraram que a modelagem com a utilização de grafos apresentou um desempenho superior aos modelos tradicionais.

Um dos algoritmos de ANN bastante utilizado na literatura para previsão de tendências é o MLP. Conhecendo as dificuldades de se obter uma boa previsão, Reyhani R., Moghadam A. (2011) sugeriram a modelagem das séries temporais com o uso da teoria do caos, de forma a melhorar o desempenho dos algoritmos preditivos. Os autores construíram um modelo de MLP alimentado por uma nova entrada obtida a partir da série temporal caótica. Os resultados mostraram que o modelo proposto teve desempenho superior ao modelo tradicional.

A extração de informações e padrões das séries temporais foi proposta por Han T., Peng Q., Zhu Z., Shen Y., Huang H., Abid N. (2020) através da criação de uma nova forma de representação de séries temporais, com o uso de DTW (*Dynamic Time Warping*). Para validar

o modelo proposto, os pesquisadores utilizaram três classificadores, 1NN, árvores de decisão e MLP. Os resultados obtidos pela solução proposta foram superiores aos resultados tradicionais desses algoritmos. Destaca-se do projeto, sua importante contribuição para previsões de curto prazo.

As séries temporais do mercado financeiro geralmente vêm acompanhadas de dados auxiliares à cotação diária, e são muito utilizados para a análise das negociações ou *tradings* diários. Zhang L., Aggarwal C., Qi G. (2017) criaram um algoritmo para previsão do preço de ações baseado na frequência das operações de compra e venda. Os autores buscaram incluir a observação da frequência de trading para obter um melhor resultado de predição, utilizando uma técnica de *State Frequency Memory*, ou memória de frequência de estado. Os resultados mostraram que esse algoritmo obteve melhores resultados na observação de padrões de multifrequência subjacentes à série temporal de preços, em comparação aos algoritmos autorregressivos e ao LSTM.

Zhao L., Wang L. (2015) propuseram uma abordagem diferente para a predição de tendências de ações no mercado chinês. Os autores optaram por estudar as anomalias que podem ser identificadas dentro das operações, das *tradings*, ao invés de se concentrar sobre o comportamento dos preços ao longo do tempo. Essas anomalias foram classificadas dentro de um *cluster* através da técnica de *K-means*. A classificação ajudou na identificação das ações que seguiram determinada tendência, tendo maior destaque as tendências de alta dos preços. Os resultados mostraram que essa técnica teve uma precisão maior que o algoritmo SVM.

Diferentes janelas temporais também podem ser consideradas para a extração de informações ou padrões para a descoberta de tendências futuras. Gonzalez R. T., Padilha C. A., Barone D. A. C. (2015) propuseram uma nova solução para predição de cotação semanal das ações do Ibovespa. Para isso, eles decidiram usar um conjunto de algoritmos combinando algoritmos genéticos com SVM. Dessa forma, um conjunto de SVM será responsável por prever a direção do movimento do preço das ações, enquanto os algoritmos genéticos atuarão de forma a otimizar cada um desses algoritmos do conjunto. Os pesquisadores obtiveram resultados promissores, apresentando desempenho superior aos algoritmos SVM, RF, AdaBoost e Bagging, ao custo de maior tempo de processamento.

Xia Y., Liu Y., Chen Z. (2013) propuseram a predição dos preços de ações utilizando a técnica de SVR. Os autores resolveram utilizar a variação diária dos preços e também as variações minuto a minuto como entrada para esse algoritmo. Foram realizados testes com empresas de diferentes tamanhos e diferentes mercados (EUA, Brasil e China). Os autores

concluíram que a utilização de SVR consegue predizer os preços, principalmente quando o modelo passa por contínuas atualizações.

Kulaglic A., Üstündağ B. (2018) propuseram um método de predição de preço do mercado de ações através da decomposição das séries temporais em subséries de diferentes tamanhos com a técnica DWT (*Discrete Wavelet Transform*). A partir dessas subséries, foram utilizados dois algoritmos de redes neurais (NN) com uma e duas camadas ocultas para a predição dos valores. Através de simulações feitas com tamanhos distintos de janelas, os autores avaliaram que a utilização dessas janelas contribuiu para a redução da taxa de erro das previsões.

Alguns autores costumam testar uma série de algoritmos de ML para determinar o mais apropriado para um conjunto de dados. A proposta de previsão de tendências de preço dos índices de ações de Bangladesh de Majumder M., Hossain M., Hasan M. (2019) constituiu numa análise de cinco diferentes técnicas de predição, de modo a encontrar aquela que gerasse o melhor resultado para o conjunto de dados disponível. Os algoritmos utilizados foram Holt Winters, *Feed-Forward Neural Network* (FFNN), ARIMA e regressão linear. Os autores encontraram melhores resultados com o algoritmo FFNN.

A utilização de séries temporais de outros conjuntos de dados ou índices do mercado financeiro se mostrou promissora. Ryota K., Tomoharu N. (2012) montaram uma estratégia para previsão do mercado de ações através da inter-relação entre diversas séries temporais. Nesse trabalho os autores buscam entender como o comportamento de outras séries temporais como índices de ações, câmbio e preço de *commodities* como petróleo se relaciona com o índice de ações. O algoritmo proposto se baseia numa fase de descoberta de padrões seguida por uma fase de predição que utiliza esses padrões como entrada. Os resultados mostraram que a proposta consegue prever a direção dos preços do mercado, seja de alta ou de baixa.

Uma estratégia adotada por alguns pesquisadores constituiu-se da análise estatística dos dados, de forma a se obter uma melhor compreensão sobre o domínio a ser trabalhado. J. Christy Jackson, J. Prassanna, Md. Abdul Quadir, V. Sivakumar (2021) visaram compreender os dados históricos dos preços das ações através de um estudo estatístico e do cálculo de algumas métricas presentes no mercado financeiro, como CAGR, índice Sharpe e médias móveis simples, retorno acumulado e correlação. Após analisarem essas principais métricas, os autores buscaram prever o comportamento dos preços futuros utilizando os métodos ARIMA, simulações de Monte Carlo e o algoritmo de previsão Prophet do Facebook. Os autores

observaram que investimento em ações menos arriscadas, e, portanto, menos voláteis, tende a atrair bons retornos.

A Tabela 3.2 apresenta um resumo comparativo dos trabalhos mencionados nesta seção sobre algoritmos de previsão de preços em bolsa de valores. A Tabela destaca as principais técnicas de ML utilizadas e também as maiores contribuições de cada um dos trabalhos.

Tabela 3.2 – Comparativo dos trabalhos sobre algoritmos de previsão.

Trabalho	*Técnicas*	*Destaque*
Reyhani R. et al., 2011	MLP	Modelagem de séries temporais com teoria do caos.
Han T. et al., 2020	MLP, 1NN, árvore de decisão	Modelagem de série temporal através de DTW.
Gonzalez R. T. et al., 2015	Conjunto de SVM e algoritmos genéticos	Uso de um conjunto de algoritmos SVM otimizado com um conjunto de algoritmos genéticos. Análise semanal.
Majumder M. et al., 2019	Holt Winters, ARIMA, FFNN, regressão linear	Análise de índices locais.
Ouahilal M. et al., 2016	SVR + filtro Hodrick-Prescott	Abordagem híbrida com uso de filtragem de ruído dos dados.
Zhao L. et al., 2015	*Cluster*	Uso de *cluster* para classificar anomalias encontradas em dados do *trading*.
Ojo S. et al., 2019	LSTM	Utilizar LSTM conectada a outra LSTM
Wang Z. et al., 2018	Processamento de Linguagem Natural e NN	Uso de análise de sentimento de notícias dentro de uma janela temporal.
Ryota K. et al., 2012	Inter-relação	Uso de outros índices de ações, taxa de câmbio e índices de *commodities*.
J. Christy Jackson et al., 2021	ARIMA, Monte Carlo, Prophet	Análise estatística da série temporal, calculando CAGR, retorno acumulado, índice Sharpe, correlação e médias móveis simples.
Kulaglic A. et al, 2018	NN e DWT	Criação de subséries a partir da série histórica de preços.
Henrique B. et al., 2018	SVR	Utilização de diferentes janelas temporais de preço, como diários e minuto a minuto.
Xia Y. et al., 2013	SVR	Utilização de valores de fechamento, abertura, alta, baixa, fechamento ajustado e volume.
Zhang L. et al., 2017	*State Frequency Memory* (LSTM)	Aperfeiçoamento do algoritmo LSTM Levar a frequência de operações do tipo *trading* em conta.
Huynh H. et al., 2017	BGRU – (GRU)	Aperfeiçoamento do GRU e utilização de notícias.
Tang L. et al., 2018	KNN - PCA	PCA para reduzir a redundância dos dados.

Huang B. et al., 2018	Bayesian LSTM	Junção do LSTM com otimização bayesiana. Períodos semanais.
Tang J. et al., 2018	RNN-LSTM (preços) e CNN (notícias)	Modelo híbrido para análise de notícias e preços.
Khattak A. et al., 2019	KNN	Redução de dispersão de dados com remoção de dados fora da curva
Joosery B. et al., 2019	ARIMA e LSTM	ARIMA melhor para *datasets* longos e LSTM para *datasets* menores.
Patil P. et al., 2020	Grafos, *Deep Learning*, regressão linear e ARIMA	Modelagem através dos dados através de grafos e construção de um sistema complexo.
Wang H. et al., 2020	Regressão linear e SVM	Utilização para estratégia de alocação no mercado chinês.
Luo R., 2020	RF, *Adapting Boost Algorithm*, NB, ARIMA, *Prophet model* (fpprophet), LSTM, GRU e TCN	Descrição das vantagens e desvantagens de cada um dos algoritmos e utilização de variáveis macroeconômicas para predição de tendências.
J. de Mello Assis et al., 2018	Conjunto de ANN	Modelagem de diferentes ANN e utilização de um conjunto delas para definição de estratégias de negociação (compra e venda).

Fonte: Dos Autores.

3.3 Considerações

Observou-se que a maioria dos trabalhos utiliza apenas os dados das cotações para a descoberta de tendências aliada a complexos algoritmos. Ao mesmo tempo, a maioria realiza previsões de curto prazo diárias, propensas a apresentarem mais erros. Alguns poucos trabalhos realizaram uma análise de dados ampla sobre os índices utilizados e poucos se preocuparam em observar se existe relação entre os indicadores macroeconômicos e o movimento da bolsa de valores. Da mesma forma, a análise do comportamento dos investidores, oriunda das finanças comportamentais, não está muito presente quando se busca obter cotações futuras. Muitos trabalhos tiveram o foco na análise das bolsas locais, sem levar em conta outras opções de investimento global e outras classes de ativos que podem contribuir para um portfólio mais diversificado uma vez que tendem a ter relação ou algum tipo de correlação com as bolsas locais.

Para essa dissertação, com o objetivo de validar o modelo de análise, um algoritmo de previsão é proposto, em que apenas os valores das cotações serão considerados, em conjunto com os dados obtidos nas análises. Dessa forma, não há a utilização de notícias, e sim a

utilização de outros indicadores, vistos na primeira seção deste capítulo. Como o foco deste trabalho é a construção do algoritmo como um todo, não serão realizadas num primeiro momento as combinações de técnicas de ML, assim como a otimização das mesmas. O objetivo será avaliar se de uma forma geral as técnicas mais comuns vistas nos trabalhos estudados apresentam um resultado satisfatório quanto à classificação mensal do Ibovespa em meses com retorno positivo ou negativo.

Conforme visto na literatura, o algoritmo LSTM é um dos modelos mais utilizados para a previsão de séries temporais e tem sido comumente utilizado em predições de preços futuros do mercado de ações. Contudo, o algoritmo se baseia no estudo dos dados passados, utilizando as cotações próximas (de curto prazo) para a previsão da cotação do próximo dia. O LSTM acaba por ser um algoritmo muito mais de previsão do que de classificação, sendo esta última o foco do presente trabalho. Da mesma forma, o algoritmo ARIMA tem como característica a previsão estatística de curto prazo, que está fora do escopo do projeto. O algoritmo proprietário Prophet do Facebook que aparece em alguns trabalhos também não será considerado. Portanto, as técnicas escolhidas são RF, NB, SVM, SVR, KNN, RLOG e MLP.

No capítulo seguinte, será apresentado o modelo proposto neste trabalho, assim como os algoritmos utilizados em sua construção, as análises iniciais que serão realizadas sobre os dados e também a forma na qual esta análise será conduzida. O modelo irá destacar as etapas de tratamento de dados, de seleção de entradas e de escolha de algoritmos de aprendizado de máquina.

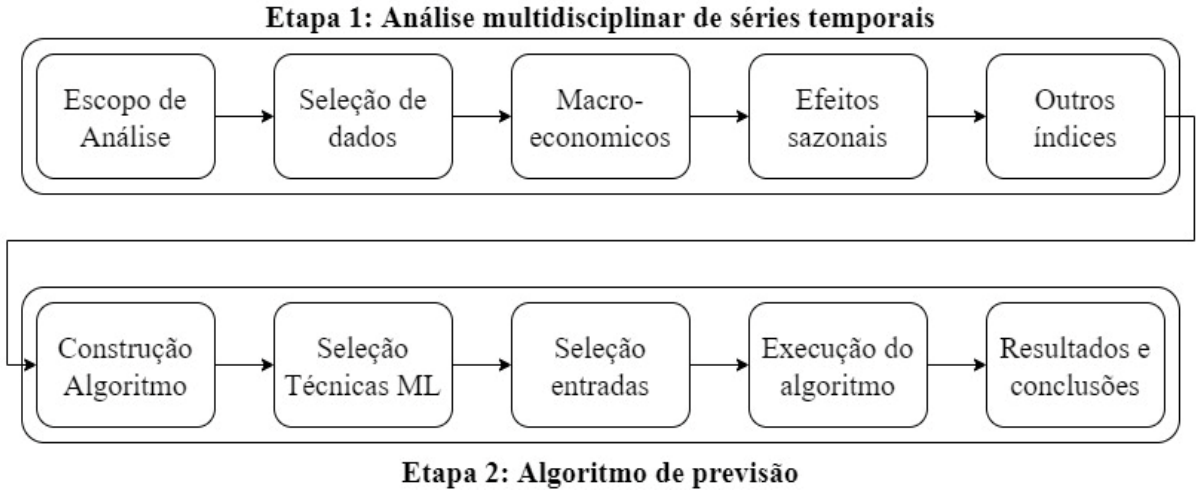
4. MODELO DE ANÁLISE MULTIDISCIPLINAR

Neste Capítulo é apresentado o modelo de análise multidisciplinar para previsão de tendência mensal na bolsa de valores brasileira. Inicialmente, mostra-se uma visão geral das etapas desse modelo, seguida pela base de dados utilizada. Por fim, o funcionamento das duas etapas que compõem esse modelo é detalhado assim como a sua implementação.

4.1 Visão Geral

O presente trabalho é constituído de duas etapas, uma de análise exploratória de séries temporais e outra de criação de um algoritmo de previsão do comportamento mensal da bolsa de valores brasileira. O modelo de análise multidisciplinar é composto de três fases para o estudo de séries temporais e o Ibovespa. Partindo dessa análise, elaborou-se uma proposta de algoritmo utilizando as variáveis estudadas nas três fases da primeira etapa para prever o comportamento mensal da bolsa de valores. A Figura 4.1 ilustra as duas etapas e seus passos.

Figura 4.1 – Visão geral do modelo.



Fonte: Dos Autores.

4.2 Base de dados

Foram utilizados dados de séries temporais de indicadores macroeconômicos e índices do mercado financeiro. Os dados macroeconômicos presentes na análise foram obtidos a partir

dos dados públicos do BCB[20], extraídos por uma API em Python. As cotações diárias dos índices de ações foram obtidas através da base de dados das plataformas de finanças online, investing.com[21] e Yahoo! Finance[22] com a utilização de suas respectivas APIs. Ambos conjuntos de dados são formados por um par de valores (cotações, preços, pontos) e índices no formato datetime.

Buscou-se coletar os dados cuja data inicial fosse pelo menos referente ao ano 2000, de forma a se conduzir um estudo ao longo das duas primeiras décadas do século XXI. As bases de dados que não existiam em 2000 foram obtidas a partir do primeiro dia em que ficaram disponíveis. Séries temporais mais antigas tiveram seus dados anteriores ao ano 2000 descartados. Na Tabela 4.1 estão apresentadas as características da base de dados utilizada. Todos os dados possuem um índice no formato datetime e o valor da cotação ou do índice em float.

Tabela 4.1 – Conjunto de dados utilizados.

Dataset	*Fonte*	*Periodicidade*	*Início*	*Fim*	*Registros*
Ibovespa	Investing.com	Dia útil	27/12/2000	30/12/2020	4953
IFIX	Investing.com	Dia útil	10/01/2013	30/12/2020	1973
S&P 500	Yahoo! Finance	Dia útil	30/12/1927	31/12/2020	23468
IAU	Investing.com	Dia útil	31/01/2005	31/12/2020	4030
SMLL	Investing.com	Dia útil	02/09/2005	30/12/2020	3788
EEM	Investing.com	Dia útil	14/04/2003	31/12/2020	4488
SELIC	BCB	Dia útil	05/03/1999	31/12/2020	7973
Dólar	BCB	Dia útil	02/01/2001	30/12/2020	5275
CDI	BCB	Diária	02/01/2001	31/12/2020	5025
PIB	BCB	Anual	01/01/1962	01/12/2021	59
PNADC	BCB	Mensal	01/03/2012	01/01/2020	95
IPCA	BCB	Mensal	01/01/2001	01/12/2020	240
IMA-B	BCB	Dia útil	30/04/2004	31/12/2020	4181

Fonte: Dos Autores.

[20] www3.bcb.gov.br/sgspub

[21] www.investing.com

[22] finance.yahoo.com

4.3 Análise Multidisciplinar

Nesta dissertação foi desenvolvido um modelo de avaliação do comportamento do mercado acionário brasileiro, representado pelo índice Bovespa, composto pelas mais importantes ações. Os trabalhos relacionados apresentam, em sua maioria, o foco da análise em apenas um domínio. O objetivo do modelo proposto nesta dissertação é não se ater a um domínio específico e sim realizar um conjunto de análises de diferentes domínios. Dessa forma, o modelo torna-se multidisciplinar e engloba três fases de análise, definidas a seguir:

- Análise Macroeconômica: visa identificar a forma pela qual alguns indicadores macroeconômicos do país afetam ou influenciam os movimentos da bolsa de valores.
- Análise Comportamental: visa identificar comportamentos presentes na tomada de decisão dos investidores e o impacto no índice acionário causado por essas escolhas.
- Análise Benchmarks: visa identificar como o índice de ações se comporta frente a diferentes índices de referência, tanto locais quanto internacionais.

Na primeira fase, é estudado e avaliado o comportamento do índice de ações em relação a algumas importantes variáveis macroeconômicas como taxa de câmbio do dólar, taxa de juros, produto interno bruto, desemprego, índices de renda fixa e inflação.

Entender como os ciclos econômicos e as decisões de política monetária afetaram a bolsa de valores ao longo das duas primeiras décadas do século XXI é o principal foco dessa primeira análise. Para isso são estudados diferentes intervalos de tempo, cada um correspondendo a importantes variações nos indicadores escolhidos. Eventos políticos de grande relevância, assim como crises, bolhas e *booms* econômicos também serão utilizados nessa análise para a definição dos intervalos.

A segunda etapa visa investigar alguns comportamentos dos investidores e como ele pode ser observado em séries temporais. Essas investigações se baseiam em observações apresentadas nos trabalhos relacionados sobre alguns efeitos comportamentais na bolsa de valores, em especial efeitos de sazonalidade.

A última parte do estudo pretende observar a relação entre alguns importantes índices e ativos do mercado financeiro com a bolsa de valores. Essa observação concentra-se nas diferenças existentes entre os desempenhos em determinados períodos assim como num todo, medindo as diferentes rentabilidades e correlações.

4.3.1 Análise Macroeconômica

Na análise macroeconômica, o modelo proposto divide as análises entre períodos de tempo conforme o comportamento da variável observada. A primeira variável a ser estudada é a taxa de juros da economia brasileira, a SELIC. A seleção dos períodos irá acompanhar os pontos de máxima e mínima mais relevantes do gráfico. Dessa forma, chega-se aos quatro períodos descritos abaixo:

- Período 1 (P1): início da série histórica até a alta de 26,5% (de 01-01-2001 até 19-02-2003).
- Período 2 (P2): período de queda de juros de 26,5% até 7,25% (de 20-06-2003 até 10-10-2012).
- Período 3 (P3): período de alta de juros de 7,25% até 14,25% (de 18-04-2013 até 29-07-2015).
- Período 4 (P4): período de queda de juros de 14,25% até final da série histórica (de 20-10-2016 até 31-12-2020).

Para cada um dos períodos é avaliado o desempenho do índice Bovespa *versus* o desempenho dos títulos atrelados à renda fixa, nesse caso o índice DI que acompanha a rentabilidade da taxa SELIC.

A segunda variável estudada é a cotação do dólar. Novamente, os períodos são definidos de acordo com a análise gráfica de máximas e mínimas, acrescido do pico ocasionado pela crise financeira mundial de 2008. Chega-se dessa forma aos sete períodos listados abaixo:

- Período 1 (P1): do início do século até a primeira máxima histórica (de 2001 a 22-10-2002).
- Período 2 (P2): da máxima histórica à primeira mínima histórica (de 23-10-2002 a 01-08-2008).
- Período 3 (P3): da mínima histórica até a alta da crise financeira de 2008 (de 01-08-2008 a 05-12-2008).
- Período 4 (P4): da alta na crise financeira até a nova mínima histórica (de 06-12-2008 a 26-07-2011).
- Período 5 (P5): da mínima até a alta da crise financeira brasileira (de 27-07-2011 a 24-09-2015).
- Período 6 (P6): do alto da crise financeira de 2015-2016 até a nova baixa (de 25-09-2015 a 16-02-2017).

- Período 7 (P7): da baixa pós crise até a máxima histórica na pandemia (de 17-02-2017 a 14-05-2020).

Em cada um desses períodos é avaliado o desempenho do índice Bovespa *versus* o desempenho do investimento em dólar. Como o dólar é utilizado como moeda internacional, além de ser uma moeda forte por pertencer a maior economia do mundo, acaba sendo uma forma de reserva de valor para o investidor.

A terceira variável a ser analisada é a inflação. Para esse indicador são avaliados os desempenhos gerais e anuais entre o Ibovespa e o índice IMA-B, que acompanha a rentabilidade dos títulos públicos atrelados à inflação.

A variável do produto interno bruto será avaliada em dois períodos, a primeira e a segunda década do século XXI. Ambas apresentaram comportamentos bastantes distintos e essas evidências serão avaliadas juntamente com o desempenho do índice de ações e os dados do desemprego, este último apenas em parte da segunda década devido à limitação do período de dados disponíveis.

4.3.2 Análise Comportamental

A segunda fase da análise baseia-se na investigação de algumas situações relativas ao comportamento dos investidores, que são verificadas através do histórico da bolsa de valores. Nessa etapa, algumas análises estatísticas também são levantadas para ajudar a compreender o comportamento do investidor e do índice de ações ao longo do tempo.

As seguintes investigações foram formuladas:

- Investigação 1: anos com retorno positivo apresentam mais retornos diários e mensais positivos, assim como anos com retorno negativo apresentam mais retornos diários e mensais negativos.
- Investigação 2: os investidores tendem a comprar mais ações ao final do ano, estando mais otimistas, fazendo o índice subir nos meses finais.
- Investigação 3: os investidores tendem a vender ações ao final do mês e comprá-las novamente no início do mês seguinte como forma de abater imposto de renda.
- Investigação 4: o efeito SMGA se aplica também na bolsa brasileira.
- Investigação 5: os investidores tendem a vender mais ações na sexta-feira e comprá-las na segunda-feira.

- Investigação 6: não estar exposto ao mercado nos dias de maiores altas e maiores quedas acaba afetando negativamente o desempenho geral.

4.3.3 Análise de Benchmarks

A parte final do processo de análise das séries temporais visa comparar a série do Ibovespa com alguns importantes benchmarks, verificando-se a rentabilidade total, o crescimento anual composto (CAGR) e a correlação entre os ativos. Nesta dissertação, optou-se por utilizar alguns dos principais índices brasileiros como o IFIX e o SMLL de forma a se avaliar ativos diferentes dos do Ibovespa. Também se considerou os índices S&P 500 para observar a relação com a maior economia do mundo, assim como o EEM para se comparar a um conjunto de países emergentes. Por fim, o IAU foi utilizado para avaliar a performance em relação a um ativo que representa uma reserva de valor.

4.4 Algoritmo de Previsão

Após o desenvolvimento de uma análise multidisciplinar composta de três fases, que trouxe uma melhor compreensão sobre a dinâmica da bolsa brasileira em relação a outras variáveis, buscou-se modelar um algoritmo de previsão para o Ibovespa que utilizasse as variáveis estudadas.

Observou-se a importância que as variáveis macroeconômicas como taxa de juros, cotação do dólar e inflação possuem em relação ao Ibovespa. Também foi considerado o estudo do comportamento mensal dos investidores, na qual descobriu-se uma tendência de sazonalidade na bolsa brasileira e como isso pode ajudar a prever um comportamento futuro. Alguns dos índices analisados como EEM, IFIX, S&P 500, SMLL e IAU podem contribuir para o modelo através do comportamento de suas correlações com o Ibovespa. Dessa forma, o algoritmo proposto nesta seção visa unir os conhecimentos obtidos das diferentes fases de análise, visando projetar o resultado mensal do Ibovespa, em períodos de tempo contidos nas duas primeiras décadas do século XXI.

O foco do algoritmo proposto é de classificar os comportamentos mensais em alta e baixa com base nos parâmetros utilizados, diferentemente dos trabalhos relacionados às previsões em bolsa de valores vistos que se dedicam a encontrar com precisão um valor para o

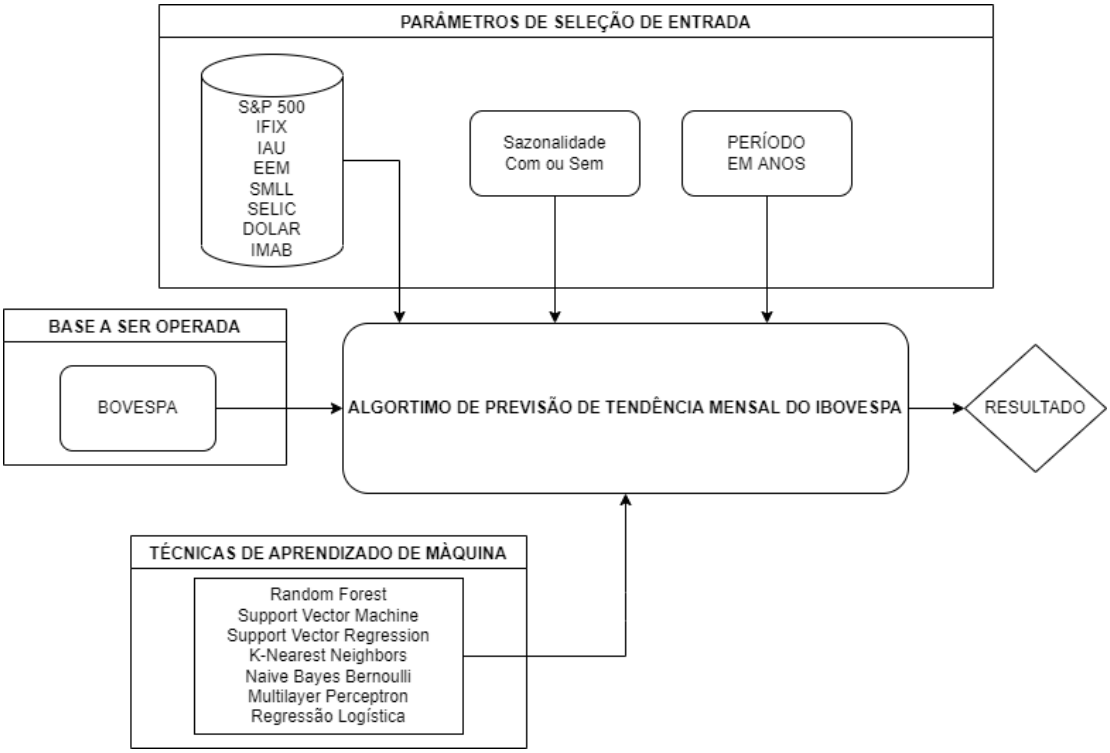
preço futuro do Ibovespa. Nesta dissertação, busca-se encontrar um conjunto de variáveis de entrada que, aliadas a uma técnica de ML, obtenha o melhor desempenho em indicar se o resultado mensal do Ibovespa será positivo ou negativo. A utilização de um conjunto de técnicas de ML se torna necessário para evitar distorções na avaliação desse conjunto de entradas, de forma que o resultado final não seja dependente de uma técnica específica.

Essa seção apresenta a descrição do algoritmo, com suas considerações e adaptações. O algoritmo faz uso de diferentes técnicas de ML, de modo a se comparar seus desempenhos visando encontrar as configurações que apresentam maior precisão.

4.4.1 Descrição do Algoritmo

A Figura 4.2 apresenta os principais componentes utilizados no algoritmo de previsão de tendência mensal proposto, *Monthly Trend Forecast Algorithm* (MTFA). O algoritmo é composto de parâmetros de entrada, base a ser operada e técnicas de aprendizado de máquina, além de uma lógica interna de codificação.

Figura 4.2 – Representação do MTFA



Fonte: Dos Autores.

4.4.1.1 Parâmetros de seleção de entrada

Os parâmetros de entrada são um conjunto de variáveis que representam a configuração do algoritmo. A seguir, estão especificadas cada um desses parâmetros.

Base de dados: Conjunto de dados de entrada, composto pelos índices e variáveis macroeconômicas estudados na análise disciplinar.

Sazonalidade ou sentimento: Opção para incluir um parâmetro que representa uma análise do sentimento do investidor em relação ao Ibovespa no período de 20 anos. Esse parâmetro terá um valor positivo (1) se o mês em questão tiver obtido na análise ao longo de vinte anos mais de 50% de meses com retornos positivos. Da mesma forma, um valor negativo (-1) será atribuído caso o mês apresentou mais anos com performance negativa do que positiva.

Período: deve ser selecionado o ano inicial para a análise. Esse parâmetro delimita a base de dados considerando apenas os dados disponíveis entre o primeiro dia do ano informado e 31 de dezembro de 2020. Os anos utilizados no experimento seguem a disponibilidade dos índices utilizados, de forma a sempre utilizar bases de tamanhos iguais. O maior período é aquele definido pelo ano inicial de 2001, resultando em vinte anos e um total de 240 amostras mensais. Estipulou-se que o mínimo a ser utilizado no algoritmo são 36 amostras, dessa forma o ano inicial máximo será o ano de 2018.

O algoritmo também apresenta uma seleção de técnicas de aprendizado de máquinas. Essa seleção visa observar e comparar a precisão final para cada uma das diferentes técnicas.

A base a ser operada, e por sua vez prevista, será o Ibovespa. O resultado mensal conhecido será avaliado em relação ao resultado previsto pelo algoritmo através de um conjunto de métricas.

4.4.1.2 Funcionamento do algoritmo

O algoritmo reúne os dados de entrada com as configurações mencionadas acima e a técnica de AM selecionada para obter o comportamento mensal do Ibovespa. As bases de dados passam por uma transformação para se adequarem aos modelos de classificação. Dessa forma, as variações mensais serão codificadas através de dois tipos de codificação propostos pelo algoritmo: Positivo-Negativo (PN) e N-valor (NV).

Positivo-Negativo

Essa é uma codificação mais simples, em que se está interessado no resultado mensal das variáveis analisadas para identificar, de forma binária, o comportamento mensal do Ibovespa, positivo ou negativo. Com essa codificação, ignora-se a magnitude do resultado mensal, se interessando apenas no sinal.

Para a variável da SELIC, optou-se por utilizar o sinal do movimento corrente, em situações em que ela ficou estável no mês. Dessa forma a SELIC assume o sinal imediatamente anterior, indicando tendência de alta se a SELIC está num movimento crescente ou baixa se está num movimento decrescente.

N-Valor

Essa codificação considera, além do sinal, a magnitude das variações. Dessa forma, utiliza-se o desvio padrão obtido entre as variações mensais de cada variável para se calcular o valor final. Um mês positivo que tenha entre 0 e 1 desvio padrão recebe o valor 1, entre 1 e 2 recebe 2 e assim sucessivamente. Da mesma forma, um mês negativo que tenha entre 0 e -1 desvio-padrão recebe o valor -1, entre -1 e -2 recebe -2 e assim sucessivamente. Escolheu-se como maior valor três desvios, portanto os dados poderão assumir valores de -3 a 3.

4.4.2 Técnicas de aprendizado de máquina

A contribuição dessa fase está relacionada à descoberta da melhor técnica de aprendizado de máquina para os algoritmos propostos. Nesta seção, são ilustrados os algoritmos utilizados e suas configurações. Como o foco do trabalho não está na otimização dos algoritmos, foram escolhidas as configurações padrão da biblioteca da linguagem Python, de forma a verificar a utilidade ou não do algoritmo.

4.4.2.1 Random Forest

O trabalho utiliza um algoritmo simples de *Random Forest*, conforme a Figura 4.3. Optou-se por limitar a profundidade das árvores a 2 níveis (max_depth = 2) e de se manter um parâmetro de aleatoriedade padrão para todas as execuções (random_state = 0).

Figura 4.3 – Parâmetros do algoritmo RF.

```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=2, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=0, verbose=0,
                       warm_start=False)
```

Fonte: Dos Autores.

4.4.2.2 Support Vector Machine

A técnica utilizada, dentre os algoritmos de SVM, é o SVC (*Support Vector Classification*), técnica capaz de desempenhar classificações binárias e multivaloradas. O algoritmo utiliza um pré-processamento através da transformação escalar que subtrai a média das entradas do valor de cada entrada e o divide pelo desvio padrão das entradas. Na Figura 4.4, encontram-se detalhadas as configurações padrões do SVC utilizado. A única alteração realizada foi a definição do parâmetro gamma como automático.

Figura 4.4 – Parâmetros do algoritmo SVM.

```
SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

Fonte: Dos Autores.

4.4.2.3 Support Vector Regression

O algoritmo SVR necessitou executar um tratamento em suas entradas da mesma forma que o SVM, utilizando uma transformação escalar. O resultado, contudo, do algoritmo é um conjunto de valores discretos entre um intervalo de 0 e 1. Buscou-se uma forma de traduzir esses valores em saídas que estivessem relacionadas com as entradas. Criou-se, portanto, uma conversão binária para classificar valores entre 0 e 0,5 como negativos e entre 0,5 e 1 como positivos. Dessa forma, tentou-se adaptar um conjunto de valores entre 0 e 1 em um conjunto entre -1 e 1. Notadamente, esse é o comportamento esperado quando se utiliza a configuração de conversão binária. A conversão multinível apresenta uma distorção que pode ocasionar a

perda da relevância de seus resultados. Na Figura 4.5, está presente a configuração original do SVR.

Figura 4.5 – Parâmetros do algoritmo SVR.

```
SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.2, gamma='scale',
    kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
```

Fonte: Dos Autores.

4.4.2.4 Naive Bayes Bernoulli

O algoritmo NBB utilizou a configuração padrão ilustrada na Figura 4.6. Optou-se por não priorizar nenhuma classe específica para esse experimento.

Figura 4.6 – Parâmetros do algoritmo NBB.

```
BernoulliNB(alpha=1.0, binarize=0.0, class_prior=None, fit_prior=True)
```

Fonte: Dos Autores.

4.4.2.5 K-Nearest Neighbors

O algoritmo KNN utilizou a configuração padrão ilustrada na Figura 4.7. O número de vizinhos selecionado foi de 5, por padrão, e não se atribuiu pesos diferentes a nenhum atributo.

Figura 4.7 – Parâmetros do algoritmo KNN.

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                     weights='uniform')
```

Fonte: Dos Autores.

4.4.2.6 Regressão Logística

O algoritmo RLOG utilizou a configuração padrão ilustrada na Figura 4.8. O número máximo de iterações não precisou ser elevado, uma vez que não reportou erros.

Figura 4.8 – Parâmetros do algoritmo RLOG.

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False)
```

Fonte: Dos Autores.

4.4.2.7 Multilayer Perceptron

O algoritmo MLP apresenta as configurações ilustradas na Figura 4.9. Além disso, optou-se por selecionar um estado aleatório padrão (random_state = 0) e sem necessidade de embaralhar as amostras (shuffle = false). Após alguns testes, observou-se que, para uma maior quantidade de dados, o algoritmo não convergiu, dessa forma optou-se por elevar a quantidade de iterações máximas para 2000.

Figura 4.9 – Parâmetros do algoritmo MLP.

```
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping=False, epsilon=1e-08,
              hidden_layer_sizes=(100,), learning_rate='constant',
              learning_rate_init=0.001, max_fun=15000, max_iter=2000,
              momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
              power_t=0.5, random_state=0, shuffle=False, solver='adam',
              tol=0.0001, validation_fraction=0.1, verbose=False,
              warm_start=False)
```

Fonte: Dos Autores.

4.4.3 Execução do algoritmo

O algoritmo foi automatizado e construído na linguagem Python, com suporte às bibliotecas de tratamento de dados e de técnicas de aprendizado de máquina descritas a seguir:

- Scikit-learning: algoritmos de aprendizado de máquina e ferramentas auxiliares.
- Numpy: tratamento de arrays.
- Pandas: tratamento de *data frames* (tabelas) de dados.
- Matplotlib: biblioteca para visualização de informações.
- Seaborn: biblioteca para visualização de informações.
- Yfinance: biblioteca para download de dados do Yahoo! Finance.
- Investpy: biblioteca para download de dados do Investing.com.

O processo de execução gera, para cada conjunto de configurações, uma saída contendo as entradas e os resultados das sete técnicas de aprendizado de máquina utilizadas. Esses resultados são armazenados em uma tabela que é salva em um arquivo de formato Excel, o qual posteriormente é utilizado para as análises.

5. ANÁLISES E RESULTADOS

Neste Capítulo são apresentados os resultados obtidos das análises multidisciplinares propostas no Capítulo 4 e discussões acerca das informações obtidas. O desempenho do algoritmo proposto será avaliado com os diferentes tipos de configuração e métodos de aprendizado de máquina. Por fim, um resumo das conclusões e dos resultados obtidos será discutido.

5.1 Análise Multidisciplinar de séries temporais

A presente seção mostra os resultados obtidos na primeira etapa desse modelo. As análises das séries temporais para as três fases distintas propostas neste trabalho são avaliadas e comentadas.

5.1.1 Análise Macroeconômica

Na análise macroeconômica buscou-se observar a rentabilidade ao longo das duas décadas iniciais do século XXI do índice Ibovespa, dos títulos indexados aos juros (CDI), do índice atrelado aos títulos públicos de inflação (IMA-B) e do dólar. Também foram avaliados indicadores da situação econômica do país como a evolução do PIB anual no período e da taxa de desemprego mensal a partir de 2014.

Inicialmente, verificou-se o desempenho dos principais índices e indicadores desde o início do século XXI, com exceção da série histórica do IMA-B, que teve início em 30 de abril de 2004. As figuras 5.1 e 5.2 ilustram a evolução da rentabilidade acumulada de 10000 reais investidos pelo período das séries históricas sem e com a presença do IMA-B, respectivamente.

Figura 5.1 – Retorno acumulado CDI, Ibovespa e dólar.

Fonte: Dos Autores.

Figura 5.2 – Retorno acumulado CDI, IMA-B, Ibovespa e dólar.

Fonte: Dos Autores.

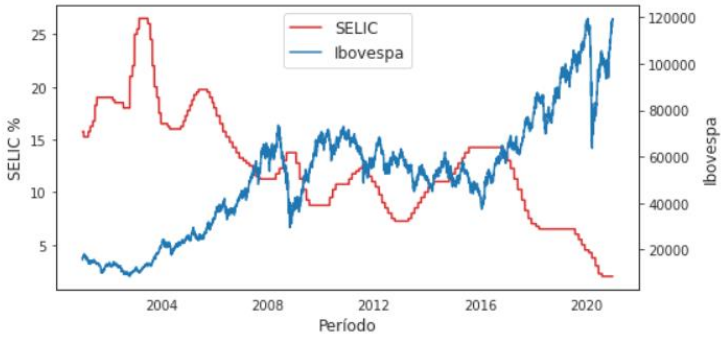
Verificou-se que os rendimentos das aplicações mais conservadoras, como os títulos indexados a juro e inflação, foram superiores, atraindo investidores devido às elevadas taxas de juros praticadas no país, com destaque para o início do século e durante os anos de 2015 e 2016. As altas taxas de juros praticadas eram justificadas pela política monetária de controle da inflação, que enquanto estava alta favorecia uma maior rentabilidade do IMA-B. Essas altas rentabilidades apresentadas em ambos os gráficos demonstram o motivo pelo qual o país tem atraído inúmeros investidores estrangeiros e nacionais para a renda fixa. Contudo, ao se observar os últimos anos, em especial, após a crise econômica brasileira, constata-se que a queda das taxas de juros e a migração de investimentos para a renda variável contribuíram para a elevação do desempenho do Ibovespa.

Nas próximas subseções são analisados os indicadores individualmente em conjunto com o comportamento do Ibovespa. Algumas análises foram divididas conforme os períodos definidos no Capítulo anterior.

5.1.1.1 Taxa de juros

A Figura 5.3 apresenta a evolução da taxa SELIC e do Ibovespa de 2001 a 2020 que teve um valor médio de 12,42%, justificando a alta rentabilidade no período. A tendência recente indica taxas em suas mínimas históricas, reduzindo a atratividade de investimentos de renda fixa, como o CDI.

Figura 5.3 – Evolução da taxa SELIC e do desempenho do Ibovespa.



Fonte: Dos Autores.

Na Tabela 5.1, estão descritos a rentabilidade do CDI e do Ibovespa para cada um dos períodos analisados e, na Tabela 5.2, um resumo comparativo entre o desempenho anual do Ibovespa e do CDI. Pelos dados observados, infere-se que o aumento da taxa de juros aumenta o rendimento da renda fixa, reduzindo a atratividade de investimentos mais arriscados como renda variável. Por outro lado, a redução das taxas de juros impacta no rendimento das aplicações de renda fixa, aumentando o apetite por risco, levando a uma procura maior por investimentos de renda variável.

Tabela 5.1 – Retorno CDI e Ibovespa nos quatro períodos e as taxas de juros dos períodos.

Período	*CDI (%)*	*Bovespa (%)*	*Juros Inicial*	*Juros Final*
1	44,11%	-33,54%	15,75%	26,5%
2	218,76%	345,18%	26,5%	7,25%
3	30,68%	5,49%	7,25%	14,25%
4	26,5%	86,44%	14,25%	2%

Fonte: Dos Autores.

Tabela 5.2 – Desempenho CDI e Ibovespa anual e as taxas de juros iniciais e finais.

Ano	*Desempenho Ibovespa*	*Desempenho CDI*	*SELIC Inicial*	*SELIC Final*
2001	-11,02	17,35	15,75	19,00
2002	-17,01	19,19	19,00	25,00
2003	97,33	23,36	25,00	16,50
2004	17,81	16,24	16,50	17,75
2005	27,71	19,08	17,75	18,00
2006	32,93	15,10	18,00	13,25
2007	43,65	11,87	13,25	11,25
2008	-41,22	12,43	11,25	13,75
2009	82,66	9,93	13,75	8,75
2010	1,04	9,79	8,75	10,75
2011	-18,11	11,64	10,75	11,00
2012	7,40	8,44	11,00	7,25
2013	-15,50	8,09	7,25	10,00
2014	-2,91	10,86	10,00	11,75
2015	-13,31	13,29	11,75	14,25
2016	38,93	14,06	14,25	13,75
2017	26,86	9,98	13,75	7,00
2018	15,03	6,45	7,00	6,50
2019	31,58	5,99	6,50	4,50
2020	2,72	2,78	4,50	2,00

Fonte: Dos Autores.

O início do século apresentou um expressivo aumento nas taxas de juros, atingindo a máxima histórica de 26,5%, ao passo que o Ibovespa registrou acentuada queda, atingindo sua mínima histórica a 8370,88 pontos. No período seguinte, a queda dos juros ocorreu em complemento à expansão da economia brasileira, numa época caracterizada pela entrada de capital estrangeiro devido ao *boom* das *commodities*. Contudo, a crise financeira mundial de 2008 fez a bolsa perder quase 60% do seu valor, caindo de 73516,8 para 29435,11 em cinco meses, entre 20 de maio de 2008 e 27 de outubro de 2020. O mercado só voltaria a superar a máxima do período, quase dez anos depois, em 11 de setembro de 2017. Considerando-se

apenas o período pré-crise, o rendimento do CDI teria sido de 105,35% contra 459,87% do Ibovespa.

A elevação dos juros no terceiro período veio como reação ao aumento da inflação e ao início da crise econômica brasileira. Após a crise, as taxas de juros voltaram a cair até 4,5% no início de 2020, contribuindo para a valorização do Ibovespa. Passado o auge da crise, a economia do país voltou a crescer, ainda que de forma modesta, e os juros foram caindo até chegar em novas mínimas históricas no início de 2020, a 4,5%. O cenário ocasionado por juros baixos pela primeira vez na história do país, fez o investimento em renda variável apresentar novos recordes, ao mesmo tempo em que atraiu diversos novos investidores.

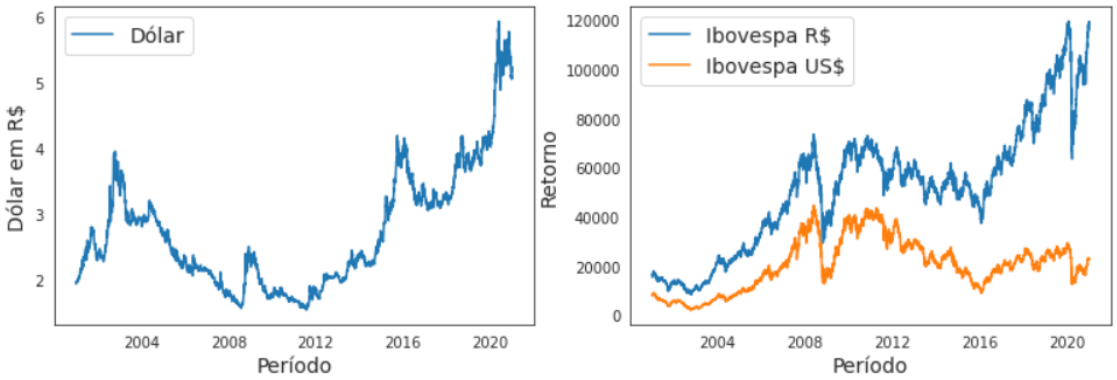
No primeiro trimestre de 2020, os temores relativos aos impactos econômicos da pandemia causaram uma forte queda na bolsa, de 119527,3 pontos em 23 de janeiro para 63569,62 em 23 de março de 2020, uma queda de 46,82% em 2 meses. Os efeitos da pandemia fizeram o banco central abaixar a taxa de juros para 2% como forma de incentivar a retomada econômica. Devido ao baixo nível de juros, houve uma forte procura por exposição à bolsa de valores, marcado pelo recorde de entrada de investidores pessoas físicas. Ao final do ano, a bolsa fecharia positiva.

De posse das informações da Tabela 5.2, se identifica a tendência de queda do desempenho da renda variável e alta do rendimento do CDI quando as taxas de juros sobem e o oposto ocorre quando elas caem. Consegue-se identificar as quedas na bolsa nos primeiros anos do século XXI com altas taxas de juros, seguidas por uma forte valorização de 5 anos consecutivos, aliados a uma queda de juros. Os impactos da crise de 2008 e de sua forte recuperação também podem ser observados, assim como a desvalorização causada na crise de 2015. A recuperação a partir de 2016 com 5 anos de altas consecutivas na bolsa e cortes nas taxas de juros, ajudam a justificar tais valores.

5.1.1.2 Dólar

O comportamento do dólar ao longo dos 20 anos do século XXI é apresentado na Figura 5.4a. O valor médio da cotação foi de 2,68 reais, tendo 1,53 real de mínima e 5,93 reais como máxima durante esse período. Esse aumento teve influência de fatores como a desvalorização do real e a saída de investimento estrangeiro devido às crises e à aversão aos riscos presentes no Brasil.

Figura 5.4 – Desempenho do dólar (a) e retorno acumulado do Ibovespa em real e dólar (b).



Fonte: Dos Autores.

Quando se considera apenas o retorno do dólar, se parte do princípio que esse seria apenas mais uma classe de investimento. Contudo, o dólar tem uma importância fundamental para a economia, já que é utilizado como moeda na maioria das transações internacionais, além de impactar no preço de diversas matérias-primas. Um dólar forte beneficia a indústria exportadora, ao passo que encarece as importações. Nos investimentos há um impacto perceptível que é a visão do investidor estrangeiro.

O investidor estrangeiro que deseja investir na Bolsa brasileira deverá converter dólares para a moeda local, portanto, para esse investidor, o comportamento da bolsa deve ser analisado em relação ao dólar. Dividindo-se o valor do Ibovespa pela cotação do dólar, se obtém o Ibovespa em dólares. Está representado na Figura 5.4b o retorno do Ibovespa tanto em reais quanto em dólares. Percebe-se que, nos últimos anos, o aumento da diferença entre o desempenho do Ibovespa em dólares e o desempenho do índice em reais. Se ao final de 2020, a bolsa está em suas máximas em reais, podendo indicar que ela estava cara, para um investidor estrangeiro ela estaria inferior às maiores altas registradas entre 2008 e 2012.

Analisando-se os dados, conclui-se que a elevação do câmbio se relaciona com uma queda no desempenho da bolsa de valores. A valorização do dólar frente a moeda local tem como causa a retirada de dólares do país, impactada principalmente pela saída de investidores preocupados com os riscos fiscais, incertezas políticas, crises internas e busca por ativos com menor risco.

A Tabela 5.3 apresenta um resumo da análise dos períodos propostos. O primeiro período, P1, foi marcado por uma expressiva alta do dólar, mais do que dobrando de valor no período. Ao mesmo tempo, o Ibovespa apresentou uma forte desvalorização, chegando em sua mínima histórica no século de 8370,88 pontos. Observa-se, portanto, que um dólar alto aliado

com uma alta taxa de juros, conforme mostrado em parágrafos anteriores, tiveram forte impacto nas cotações do Ibovespa. O período seguinte apresentou forte valorização da bolsa aliado a uma queda de mais de 50% da cotação do dólar.

Tabela 5.3 – Desempenho Dólar, Ibovespa e cotação do dólar durante os sete períodos

Período	*Dólar (%)*	*Ibovespa (%)*	*Dólar Inicial (R$)*	*Dólar Final (R$)*
1	104,05	-39,51	1,94	3,96
2	-59,71	485,67	3,96	1,56
3	59,68	-36,44	1,56	2,5
4	-37,85	55	2,5	1,53
5	168,23	-22,3	1,53	4,19
6	-22,72	51,26	4,19	3,05
7	91,83	16,62	3,05	5,93

Fonte: Dos Autores.

A crise financeira de 2008 nos Estados Unidos ocasionou uma disparada do dólar, uma vez que investidores buscaram ativos mais seguros como o tesouro americano para se protegerem das fortes quedas do mercado, dessa forma acabaram por retirar seus dólares do país influenciando na cotação.

A chegada da crise econômica brasileira também foi uma forte responsável pela elevação da cotação do dólar. Períodos de recessão econômica, altas taxas de juros e inflação elevada aumentam a percepção de risco do país, afastando investidores estrangeiros da economia e do mercado financeiro brasileiro.

A cotação do dólar terminou o ano de 2020 em alta, acima dos 5 reais, ao mesmo tempo em que o índice Bovespa batia sua máxima histórica. Diferentemente dos outros períodos analisados, o sétimo apresentou valorização tanto para o dólar quanto para a bolsa. Se por um lado o temor dos investidores com a situação econômica e fiscal do país elevou o dólar, as baixas taxas de juros, mantidas em 2%, influenciaram a entrada de investidores nacionais na bolsa de valores em busca de retornos maiores.

O desempenho do dólar frente ao real tornou a bolsa brasileira muito mais atrativa para investidores estrangeiros. Na Tabela 5.4, pode-se observar o desempenho da bolsa em reais, em dólar e também do próprio dólar. Em períodos de forte valorização do dólar, o desempenho da bolsa em reais e, principalmente da bolsa em dólares, foi fraco. Em períodos de forte desvalorização do dólar, a bolsa performou bem, mas a bolsa em dólares obteve mais retorno

ainda, sendo responsável pelos dois melhores desempenhos anuais, em 2003 e em 2009, ambos acima de 140%.

Tabela 5.4 – Desempenho Ibovespa em reais, em dólares e do dólar.

Ano	*Desempenho Ibovespa em Reais*	*Desempenho Ibovespa em Dólares*	*Desempenho Dólar*
2001	-11,02	-25,02	18,67
2002	-17,01	-45,50	52,27
2003	97,33	141,32	-18,23
2004	17,81	28,23	-8,13
2005	27,71	44,83	-11,82
2006	32,93	45,54	-8,66
2007	43,65	73,39	-17,15
2008	-41,22	-55,45	31,94
2009	82,66	145,16	-25,49
2010	1,04	5,59	-4,31
2011	-18,11	-27,26	12,58
2012	7,40	-1,42	8,94
2013	-15,50	-26,28	14,64
2014	-2,91	-14,37	13,39
2015	-13,31	-41,03	47,01
2016	38,93	66,46	-16,54
2017	26,86	24,98	1,50
2018	15,03	-1,79	17,13
2019	31,58	26,49	4,02
2020	2,92	-20,18	28,93

Fonte: Dos Autores.

5.1.1.3 IMA-B

Os títulos federais atrelados à inflação, como o IMA-B, possuem um rendimento baseado na inflação acumulada em um determinado período acrescido de uma taxa extra. Diferentemente do CDI, o índice IMA-B sofreu algumas quedas devido à queda dos preços dos

títulos de inflação num cenário de forte queda de juros e de inflação. O retorno total do Ibovespa no período foi de 507,01% versus 809,46% do IMA-B. Com isso, o IMA-B apresentou um CAGR de 14,18% contra 10,05% no período de 16 anos, não considerando 2004. Na Tabela 5.5, pode-se avaliar a evolução dos desempenhos anuais de ambos os índices.

Tabela 5.5 – Desempenho Ibovespa versus IMA-B.

Ano	*Desempenho Ibovespa*	*Desempenho IMA-B*
2004	17,81	8,89
2005	27,71	13,89
2006	32,93	22,09
2007	43,65	14,04
2008	-41,22	11,03
2009	82,66	18,95
2010	1,04	17,04
2011	-18,11	15,11
2012	7,40	26,68
2013	-15,50	-10,02
2014	-2,91	14,54
2015	-13,31	8,88
2016	38,93	24,81
2017	26,86	12,79
2018	15,03	13,06
2019	31,58	22,95
2020	2,92	6,41

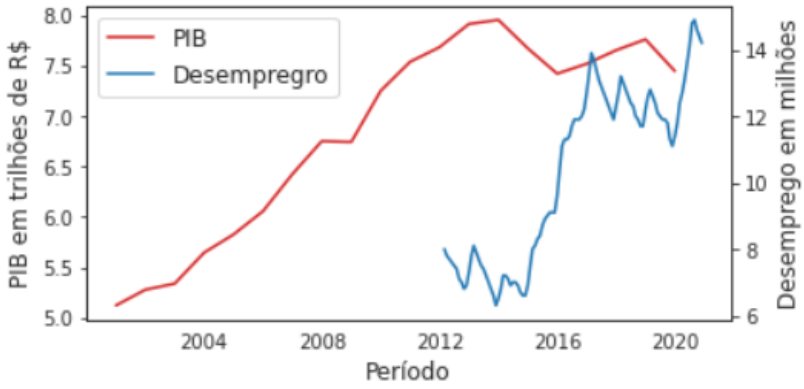
Fonte: Dos Autores.

O desempenho do IMA-B foi superior ao do Ibovespa de 2010 a 2015, período em que houve um aumento da pressão inflacionária que culminou na crise financeira brasileira, reduzindo o desempenho dos investimentos em bolsa. O investimento nesses títulos se mostrou uma oportunidade, tanto para proteção contra os efeitos inflacionários, quanto para ganhos reais superiores a investimentos mais arriscados, visto que os títulos públicos indexados à inflação possuem a garantia do tesouro nacional.

5.1.1.4 PIB, PNADC e Ibovespa

Finalizando a análise dos fundamentos macroeconômicos, foram avaliados o Produto Interno Bruto e a taxa de desemprego, este último obtido através do PNADC. Tanto o PIB quanto a taxa de desemprego tentam ajudar a explicar a situação econômica e social de um país, às vezes se distanciando do que é observado na bolsa de valores.

Figura 5.5 – PIB e desemprego nas duas primeiras décadas dos anos 2000.



Fonte: Dos Autores.

Na Figura 5.5, observa-se um crescimento inicialmente lento até 2003, seguido de uma aceleração até 2008, beneficiado pelo *boom* das *commodities*, interrompido em 2008 pela crise financeira mundial. Após a crise, o PIB continuou subindo até 2014, quando registrou sua máxima. Em 2015 e 2016, o Brasil viveu uma recessão, com crescimento negativo do PIB, sendo retomado apenas em 2016 num ritmo menor. O ano de 2020 foi marcado pela pandemia que afetou a economia nos primeiros semestres, acarretando numa queda de 4,06% do PIB. O crescimento total do PIB, ao longo de 20 anos, foi de 45,41%. Comparando-se as duas décadas, na primeira o PIB teve um crescimento de 41,59% e o Ibovespa de 349,3% contra um crescimento de -1,23% no PIB e 65,3% do Ibovespa na segunda. Na primeira década, o CAGR do PIB foi de 3,54% e o do Ibovespa de 16,21% contra de -0,12% do PIB e 5,74% do Ibovespa na segunda.

As empresas listadas na bolsa representam apenas uma parcela muito pequena das empresas do país, ao mesmo tempo em que estão entre as maiores e com desempenhos superiores à média da economia. Contudo, as grandes empresas não são responsáveis pela maior parte da força de trabalho. A recessão de 2015 impactou a situação social da população brasileira, acentuando a desigualdade, elevando o número de desempregados.

O PNADC apresenta uma pesquisa mensal sobre o número oficial de desempregados do país, sendo disponibilizado a partir de 2012. No trabalho utilizaram-se análises a partir de 2013 para obtenção de dados anuais completos. Observa-se um aumento do número de desempregados em 93,01%, devido à recessão e aos efeitos da pandemia, ao passo que o PIB caiu 5,92% no período. A bolsa, por outro lado, apresentou ganhos de 90,28%, projetando um futuro otimista de recuperação.

O CAGR do PIB, do número de desempregados e do Ibovespa foram respectivamente -0,76%, 8,57% e 8,37%. A Tabela 5.6 ilustra o desempenho anual do Ibovespa, do PIB e do número de desempregados. Observa-se que o mercado costuma fazer previsões acerca do futuro, colocando, nos preços dos ativos, expectativas futuras. O mercado apresentou quedas antes do PIB e da explosão do número de desempregados, entretanto, foi o primeiro a se recuperar.

Tabela 5.6 – Variação anual (%) do Ibovespa, PIB e número de desempregados.

Ano	*Ibovespa*	*PIB*	*Número de Desempregados*
2013	-15,5	3,00	10,14
2014	-2,91	0,50	4,84
2015	-13,31	-3,55	36,92
2016	38,93	-3,28	34,83
2017	26,86	1,32	-1,67
2018	15,03	1,78	-1,69
2019	31,58	1,41	-5,17
2020	2,92	-4,06	26,36

Fonte: Dos Autores.

5.1.2 Análise comportamental

Nesta seção são discutidas as investigações propostas no modelo acerca do comportamento dos investidores. Inicialmente, um pequeno resumo sobre o desempenho do Ibovespa no período é apresentado. O Ibovespa apresentou entre 2001 e 2020 uma valorização de 679,97%. Nesses 20 anos, teve retorno positivo em 14 deles, representando 70%. O maior retorno anual ocorreu em 2003 aos 97,33%, após dois anos seguidos de quedas em que houve o estouro da bolha das empresas de internet americanas. O pior desempenho anual da bolsa ocorreu em virtude da crise de 2008, com uma queda de 41,22%.

Por duas vezes a bolsa apresentou 5 anos seguidos de desempenho positivo. O primeiro foi de 2003 a 2007, na qual o Ibovespa valorizou 466,95%; o segundo período foi de 2016 a 2020, com uma valorização de 174,55%.

5.1.2.1 Investigação 1:

Na primeira investigação busca-se verificar se o desempenho anual, positivo ou negativo, reflete num maior número de dias ou meses com o mesmo desempenho. Observou-se que em 90% dos anos analisados, a maioria dos desempenhos diários acompanhou o desempenho anual. Não se pode concluir que essa é uma tendência definitiva, uma vez que houve anos em que a diferença entre o número de dias com retornos positivos e negativos foi muito próxima, ao passo que o retorno anual teve grande magnitude. O ano de 2008 ilustra essa situação com 125 dias com retornos negativos e 124 com retornos positivos, mas apresentando um desempenho anual negativo de 41,22%.

Analisando-se as variações mensais, percebe-se uma tendência de que anos que tenham a maioria dos desempenhos mensais positivos venham a ter um desempenho anual positivo, da mesma forma para desempenhos mensais negativos. Essa relação foi verdadeira em 65% dos anos. Em 30% dos anos, metade dos desempenhos mensais foram positivos e metade negativos e, em apenas um ano (5%) houve mais meses positivos para um retorno anual negativo de 2,91%.

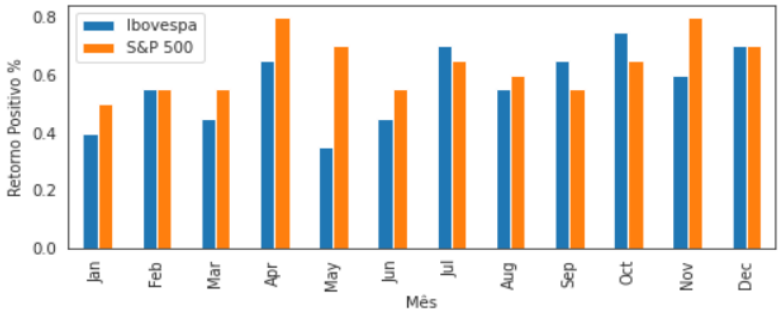
A respeito da primeira investigação conclui-se que a tendência de apresentar maior número de meses com retornos de um tipo pode levar a um retorno anual semelhante. Considera-se, contudo, que nessa análise a bolsa teve um movimento crescente ao longo dos anos, não fornecendo indícios de como seriam os retornos em períodos de tempo com movimentos decrescentes ou lateralizados.

5.1.2.2 Investigação 2

Analisando-se os retornos mensais, confirma-se a investigação de que no Brasil há uma tendência de alta na bolsa nos últimos meses do ano, em especial nos meses de outubro e dezembro que obtiveram 75% e 70% de meses em alta, respectivamente, conforme a Figura 5.6. Esse padrão observado tem como possíveis causas os movimentos de euforia e otimismo com as festas de final de ano, o recebimento de décimo terceiro salário ou bônus, assim como

a oportunidade para os fundos de investimentos baterem suas metas e recolherem taxa de performance dos cotistas.

Figura 5.6 – Percentual de meses com retorno positivo no Ibovespa e no S&P 500.



Fonte: Dos Autores.

Por outro lado, destacam-se as quedas ocorridas no mês de janeiro, em 60% dos anos. Por janeiro ser um mês de férias de verão no hemisfério sul, gestores podem optar por venderem suas posições antes de viajar. Outra conclusão é a existência de movimentos de correção em resposta às altas de final de ano.

5.1.2.3 Investigação 3

Foi observado no Ibovespa, o comportamento "*Sell in May and go Away*". Utilizada por investidores estrangeiros, essa frase faz referência às vendas que ocorrem no mês de maio, período anterior às férias de verão no hemisfério norte. Apesar de ser um efeito comentado em outros países, ele teve repercussão na bolsa brasileira. No período estudado, observou-se que o mês com maior número de quedas é o mês de maio, com 65% dos anos. Comparando-se com o índice americano S&P 500 no mesmo período, conforme mostrado na Figura 5.6, observa-se que o mês de maio apresentou o resultado inverso, sendo positivo em 70% do período. Conclui-se que a percepção de risco dos investidores estrangeiros é maior ao se investir num país emergente como o Brasil.

5.1.2.4 Investigação 4

No Brasil, há um benefício fiscal para vendas com lucro de valores inferiores a 20000 reais em ações por mês. Dessa forma, investidores podem vender suas ações ao final de cada mês até esse valor para evitar o pagamento de imposto de renda e recomprá-las no início do mês seguinte. Para esta investigação, calculou-se a variação entre o último dia de cada mês e o

primeiro dia do mês subsequente. O resultado geral obtido foi uma alta de 63,33%, contribuindo para a confirmação desta investigação, embora seja difícil torná-la de fato absoluta, uma vez que o valor para isenção de imposto é baixo, *versus* o volume negociado diariamente na bolsa.

Para o investidor que planeja utilizar esse tipo de estratégia, identifica-se a tendência de elevação de preço, obrigando o investidor a pagar um valor mais elevado, na hora de recomprar suas ações. Cabe ao investidor avaliar se esse risco é válido em troca da isenção tributária.

5.1.2.5 Investigação 5

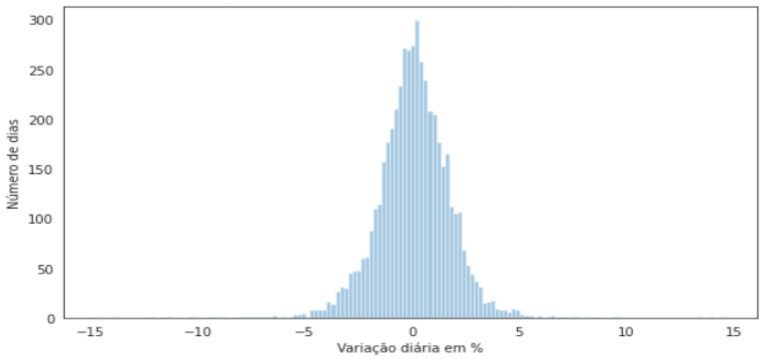
Analisando-se as variações diárias, identificou-se que somente nas segundas-feiras houve mais desempenho diário negativo (50,15%) do que positivo (49,85%). Para os outros dias da semana, o maior percentual encontrado foi para as terças-feiras, com 53,7% de altas. Durante o período, 52,19% dos dias foram de alta, então se conclui que as terças-feiras estão um pouco acima da média, do mesmo modo que as segundas-feiras estão um pouco abaixo. Não se pode concluir que é uma tendência clara, uma vez que os percentuais são próximos à média.

5.1.2.6 Investigação 6

Uma discussão interessante a ser abordada se refere a estar presente no mercado financeiro. Na maioria dos dias as variações positivas ou negativas são de baixa magnitude, contudo, com menor frequência, ocorrem dias com eventos de significativo impacto. Estar exposto ao mercado nesses dias pode significar altos lucros ou prejuízos como será visto em mais detalhes.

A Figura 5.7 mostra a distribuição dos retornos diários ao longo de vinte anos do Ibovespa. É fácil de identificar que na maioria dos dias os retornos estão concentrados numa faixa que varia entre -2,5% e +2,5%. Contudo, pela Figura 5.7 não é possível identificar os dias de grandes altas ou quedas, uma vez que são raros e acabam por não aparecer nessa distribuição.

Figura 5.7 – Distribuição da variação diária do Ibovespa.



Fonte: Dos Autores.

Na Tabela 5.7, é apresentado os valores das maiores altas e baixas da bolsa. Observa-se que as maiores quedas foram ligeiramente superiores às maiores altas, mostrando que sentimentos de aversão a perdas podem ser mais fortes que sentimentos otimistas.

Tabela 5.7 – Maiores altas e quedas do Ibovespa.

Posição	*Maiores Altas*	*Maiores Baixas*
1	14,66	-14,78
2	13,91	-13,92
3	13,42	-12,17
4	9,69	-11,39
5	9,57	-10,35
6	9,40	-10,18
7	8,36	-9,36
8	8,31	-9,18
9	7,63	-8,80
10	7,61	-8,09

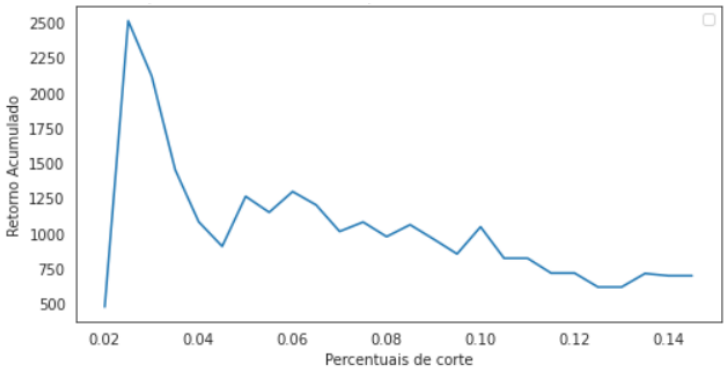
Fonte: Dos Autores.

Em vinte anos o índice Bovespa rendeu 697,98%, mas se fossem desconsideradas as cinco maiores altas, o rendimento cairia para 338,09% e, 194,58%, se fossem retiradas as dez maiores altas. Por outro lado, se fossem retiradas as cinco maiores quedas o rendimento subiria para 1596,71% e, 2358,87%, se fossem retiradas as dez maiores quedas. Esse resultado era esperado, mas surpreende sua magnitude quando da retirada das grandes altas. Um resultado importante pode ser obtido ao se retirar ambos os dias de grandes altas e grandes quedas. Retirando-se os cinco dias com as maiores altas e os cinco com as maiores quedas, o rendimento do Ibovespa seria de 853% e se fossem retirados os 10 dias, seria de 828,68%. Ambos os

retornos, obtidos eliminando as variações de maior magnitude, mostraram-se superiores ao rendimento normal da bolsa. Isso reforça que matematicamente a recuperação de grandes quedas exige um retorno maior. Um exemplo é uma queda de 50%, na qual é necessária uma recuperação de 100% para se voltar ao patamar inicial.

No gráfico 5.8 é possível observar o comportamento do desempenho do Ibovespa quando são retirados os dias com variações acima de determinado valor em módulo. Percebe-se que o desempenho aumenta inicialmente quando se aumenta o valor de corte até um desempenho máximo ao se eliminar as variações superiores a 3% em módulo. Logo após esse ponto o desempenho começa a cair e ao final tem-se o comportamento inicial.

Figura 5.8 – Percentuais de corte e seu respectivo retorno acumulado.



Fonte: Dos Autores.

O aprendizado que se obtém com essa análise está relacionado à importância de se manter no mercado por mais dias para aproveitar as pequenas variações que impactam fortemente o resultado. Também é perceptível que estar exposto a movimentos extremos, onde o risco é maior, pode ao mesmo tempo gerar excelentes retornos como graves perdas. Portanto, é importante saber controlar a exposição e aproveitar sequências de quedas para novos investimentos, uma vez que quando o movimento se inverter, os papéis comprados a um menor preço têm mais chance de auferir maiores ganhos. Ao mesmo tempo, é importante estar atento às excessivas altas que podem causar movimentos de correção, ou até estouro de bolhas especulativas, fazendo os papéis comprados em períodos de alta terem que se recuperar de grandes quedas para voltar a seus valores originais.

5.1.3 Análise de Benchmark

Nesta seção, são analisados os desempenhos dos índices S&P 500, IAU, EEM, SMLL e IFIX em relação ao Ibovespa. Os índices S&P 500, IAU e EEM são cotados em dólar, então,

resolveu-se incluir também na análise esses índices convertidos para reais. A Tabela 5.8 apresenta o desempenho de cada índice, o desempenho do Ibovespa no mesmo período, a correlação frente ao Ibovespa e o CAGR.

Tabela 5.8 – Desempenho dos índices analisados em relação ao Ibovespa.

Índice	*Ganho*	*Desempenho Ibovespa*	*CAGR*	*Correlação*	*Anos considerados*
S&P US$	184,49	679,97	5,52	0,62	20
S&P R$	656,07	679,97	10,85	0,31	20
IFIX	80,52	92,96	11,07	0,44	7
SMLL	443,66	320,27	15,03	089	15
IAU US$	328,61	388,72	8,52	0,13	15
IAU R$	748,57	388,72	14,43	-0,16	15
EEM US$	357,66	902,33	6,22	0,73	17
EEM R$	663,41	902,33	9,96	0,58	17

Fonte: Dos Autores.

Observa-se que a cotação do dólar influenciou de forma considerável o retorno dos índices dolarizados, impulsionando seus desempenhos versus as cotações em moeda estrangeira. Mesmo assim, o Ibovespa dentro do período analisado para cada índice, acabou por apresentar resultado melhor contra a grande maioria dos índices dolarizados, contra o IFIX e até contra alguns ativos convertidos em moeda local, perdendo apenas para o SMLL e IAU em reais.

A correlação é bem fraca em relação aos ativos dolarizados em reais como S&P 500. Por sua vez, desconsiderando-se a variação do dólar, a correlação entre as bolsas americana e brasileira mostra uma tendência de movimentos próximos. Os fundos imobiliários presentes no IFIX apresentam uma dependência fraca em relação ao Ibovespa, tornando esse índice uma opção para diversificação. De todos os ativos presentes, o ouro é o que apresenta menor correlação, chegando a ser negativa quando convertido para moeda local. O fato de ser considerado uma reserva de valor e atrair os investidores em momentos de estresse do mercado ajuda a exemplificar o porquê de um movimento oposto ao Ibovespa.

Quando se observa o CAGR, identifica-se um bom desempenho médio dos índices de ações brasileiros, perdendo apenas para os índices dolarizados convertidos para reais. O valor do CAGR mais que dobrou os valores originais desses índices em virtude do dólar. Como uma economia emergente, o Brasil apresentou elevadas taxas de desempenho se comparado aos

países desenvolvidos, contudo, ao nível global, a desvalorização da moeda acabou por enfraquecer o desempenho.

5.1.4 Considerações sobre Análise Multidisciplinar

Nesta seção é mostrado um resumo sobre os resultados apresentados nas seções anteriores. De uma forma mais objetiva, os pontos citados a seguir destacam os principais resultados obtidos na etapa de análise multidisciplinar.

- Importância de se visualizar o comportamento das diferentes variáveis em relação ao Ibovespa, em destaque aquelas de opostos como dólar e SELIC.
- O mercado não acompanha a situação social do país.
- Inflação e CDI tiveram desempenho mais fracos em períodos de alta na bolsa.
- Índices dolarizados apresentaram desempenho muito superior, devido aos períodos de câmbio valorizados.
- Verifica-se uma tendência de sazonalidade na bolsa brasileira, representada pelo efeito SMGA e pela euforia de final de ano.

5.2 Análise e Resultados MTFA

Esta seção apresenta os resultados das simulações realizadas com o algoritmo MTFA. O objetivo dessas simulações foi encontrar a melhor configuração de entradas que obtivesse a maior precisão na predição mensal do Ibovespa. Utilizou-se um conjunto de técnicas de aprendizado de máquina (RF, SVM, SVR, NBB, MLP, RLOG e KNN) para classificação e que fossem comuns na literatura e de fácil implementação e replicação.

A partir de uma configuração de entrada, as simulações produziram como resultado a precisão da previsão mensal do Ibovespa para cada uma das técnicas de ML. A presente análise visa identificar quais algoritmos tiveram a maior precisão, quais configurações obtiveram melhores resultados e como as diferentes configurações impactaram os resultados.

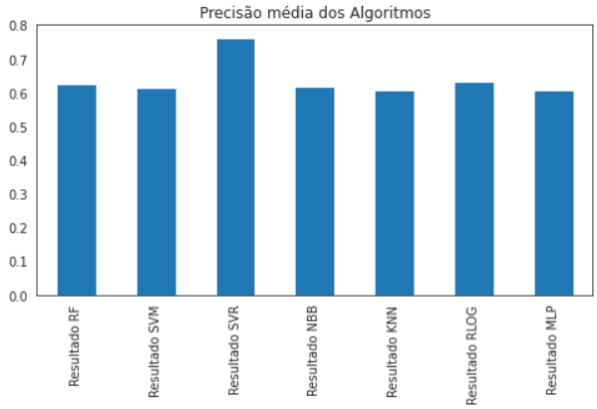
A seguinte seção está organizada em uma análise mais geral dos resultados como um todo e em uma análise de acordo com a quantidade de índices utilizados na configuração. Para esta última, os resultados estão divididos de acordo com a quantidade de índices de entrada, resultando em oito conjuntos de dados, contendo de dois a nove índices.

5.2.1 Análise Geral

As 9432 configurações de entrada foram compostas por variações de um conjunto de índices, indo de apenas dois elementos até nove e também de uma variação da data inicial da amostra, indo de 2001 a 2018. A data inicial tende a acompanhar a disponibilidade de determinado índice, conforme destacado na tabela de dados. Anos iniciais menores produziram menos configurações, uma vez que menos índices possuíam dados no período. A data inicial é utilizada como limitador da amostra de dados. Exemplificando, uma data inicial de 2018 apresenta 36 meses de amostra, visto que a amostragem inclui todos os meses desde o ano inicial até o mês de dezembro de 2020. De forma semelhante, a data inicial de 2001 apresenta a maior amostra com 240 meses.

A primeira análise realizada foi a das médias de desempenho de cada um dos algoritmos e está ilustrada na Figura 5.9. Observa-se que o resultado do algoritmo SVR se destaca em relação aos demais, apresentando precisão de 76,26%, enquanto os outros apresentam precisão um pouco superiores a 60%. O resultado do algoritmo SVR é muito influenciado pelo algoritmo de avaliação utilizado, que arredonda os valores obtidos entre o intervalo de 0 e 1 para esses valores. Dessa forma, a aproximação acaba impactando de forma decisiva a performance desse método em todos os cenários. Excluindo-se o algoritmo SVR, o método de Regressão Logística apresentou o melhor resultado médio com 63,35%. O pior resultado ficou com o algoritmo de MLP, com 60,68%.

Figura 5.9 – Precisão média dos algoritmos.

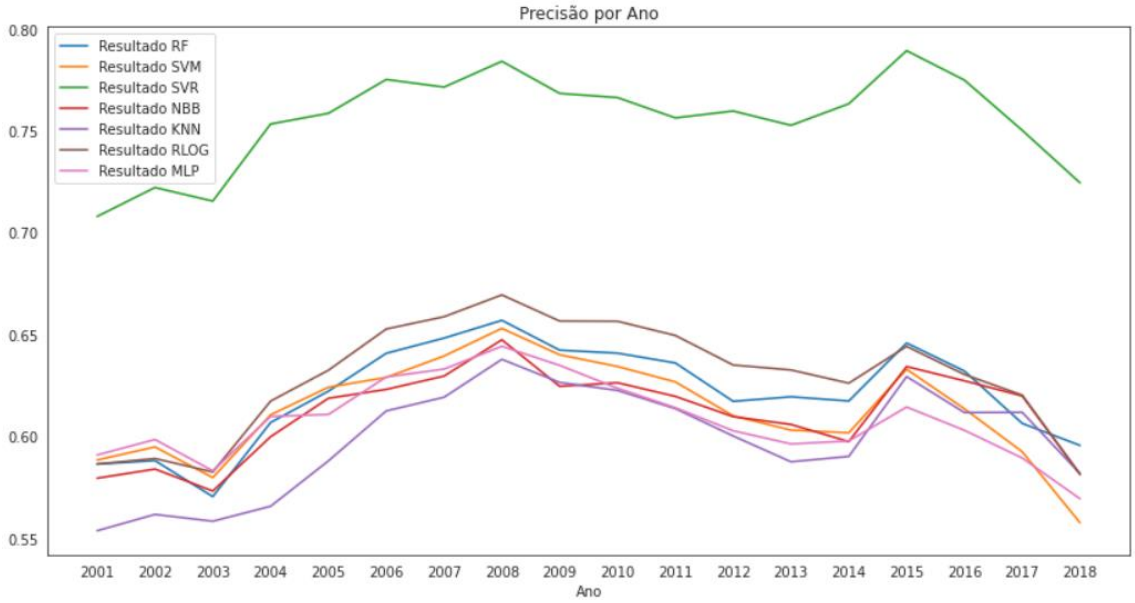


Fonte: Dos Autores.

Essa primeira análise não ajuda a se obter uma maior compreensão sobre como as diferentes configurações impactam na precisão dos algoritmos. Em busca de uma melhor visualização dos resultados, buscou-se avaliar isoladamente as diferentes configurações, em especial, os tipos de conversão, a utilização de sentimento e o ano inicial.

Na Figura 5.10 observa-se a evolução do desempenho de todos os algoritmos de acordo com o ano inicial e por consequência, com o tamanho da amostra. A partir da Figura 5.10, pode-se identificar alguns fatos relevantes. Os resultados médios mais elevados estão relacionados com amostras medianas, sendo o ano inicial de 2008 e suas 130 amostras aquele que obteve melhores resultados para a grande maioria dos algoritmos. Destaca-se também que as maiores amostras (2001-2003) acabaram por obter um resultado mais fraco, assim como o ano de menor amostra (2018).

Figura 5.10 – Precisão média dos algoritmos anual.

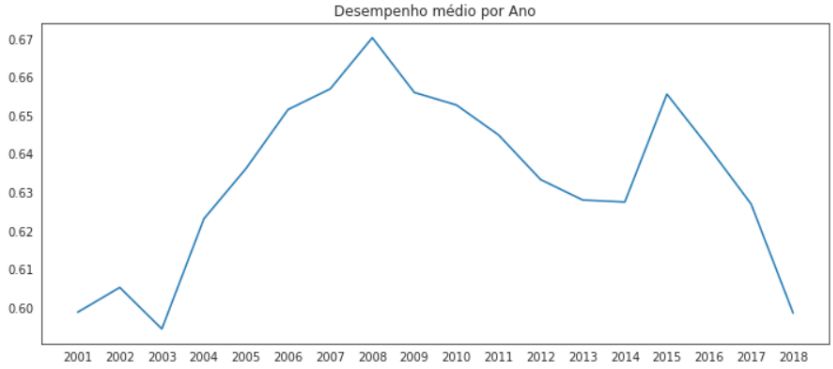


Fonte: Dos Autores.

Em relação aos algoritmos, devido à utilização do método de aproximação, o SVR obteve expressivos resultados em relação aos demais. Mas cabe destacar o algoritmo de RLOG que obteve a segunda melhor performance em grande parte dos anos, assim como o de RF que obteve a terceira. Diferentemente do que os resultados anteriores mostravam, o MLP não teve o pior desempenho em grande parte do período, sendo o KNN responsável por esta posição. No entanto, a diminuição do número de amostras afetou o MLP de forma mais impactante que o KNN.

Uma nova forma de olhar para esses dados é através da média calculada a partir das médias dos algoritmos por ano, conforme está ilustrado na Figura 5.11. Dessa forma, observa-se que os melhores períodos para início de amostra foram os anos de 2007 a 2009 e 2015. Esses períodos capturam o início das crises financeiras internacional e nacional e um padrão inicial de comportamento pode ter sido fundamental para que os algoritmos encontrassem semelhanças que permitissem uma maior precisão na classificação dos dados.

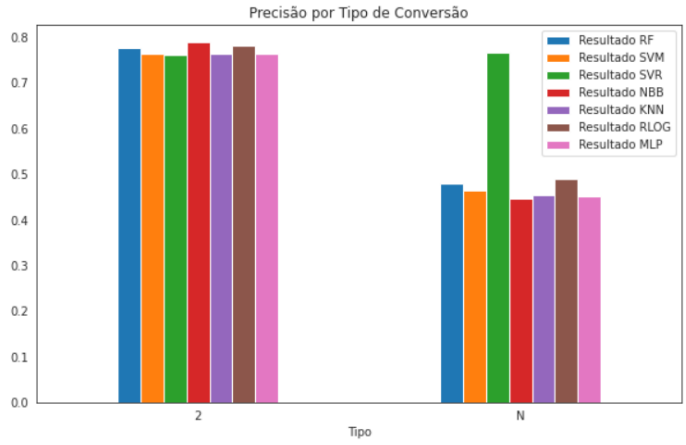
Figura 5.11 – Precisão média anual de todos os algoritmos.



Fonte: Dos Autores.

Uma nova observação é feita colocando-se o tipo de conversão como destaque como está destacado na Figura 5.12. De imediato, nota-se a grande diferença entre a precisão média dos algoritmos para cada um dos tipos de conversão. A conversão binária apresenta um desempenho superior em quase todos os algoritmos, sendo apenas o SVR que apresenta um resultado um pouco superior com a conversão multinível.

Figura 5.12 – Precisão média dos algoritmos em relação ao tipo de conversão.



Fonte: Dos Autores.

Em relação à conversão binária destacam-se os seguintes aspectos:

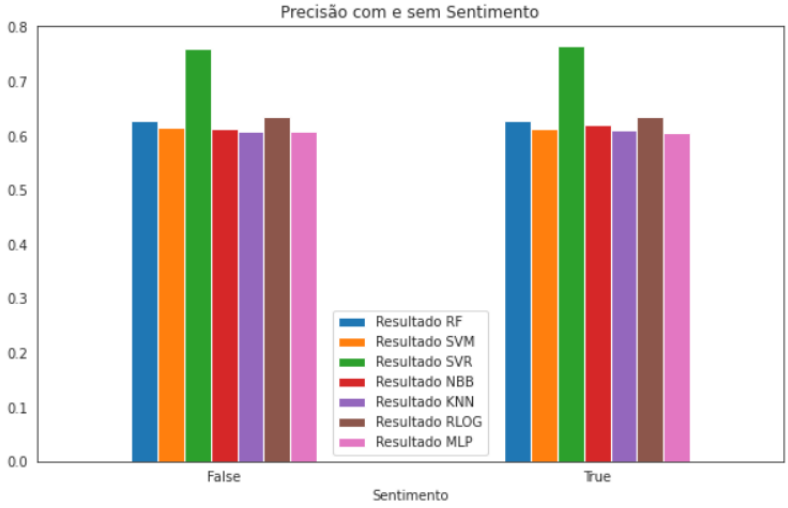
- NBB apresenta o melhor desempenho, superando inclusive SVR;
- SVR apresenta uma precisão levemente inferior comparada a todos os seus pares;
- A média de precisão fica acima de 75% em todos os algoritmos.

Em relação à conversão multinível, destacam-se:

- SVR teve o maior desempenho e único acima de 50%;
- RLOG obteve o segundo melhor desempenho, mesmo que insuficiente;
- NBB acabou tendo a pior performance.

Observa-se através da Figura 5.13 que a utilização ou não de uma entrada com o parâmetro de sentimento não resultou em um importante impacto no desempenho dos algoritmos. Os algoritmos RF, SVR, NBB, KNN e RLOG tiveram um leve acréscimo de precisão, enquanto SVM e MLP tiveram uma leve piora.

Figura 5.13 – Precisão média por algoritmo em função do uso de sentimento.



Fonte: Dos Autores.

Uma forma de medir o impacto da presença de um determinado índice como entrada é através da diferença entre as médias de precisão de um conjunto de entradas contendo o índice e outro conjunto não contendo o índice avaliado. Na Figura 5.14 (a) observa-se a precisão média com a remoção de cada um dos índices indicados na coluna da esquerda. A precisão média nesse caso é a média das precisões médias de cada um dos algoritmos. Na Figura 5.14 (b), por sua vez, observa-se a precisão média para todas as entradas em que o índice da coluna da esquerda está disponível. Na Figura 5.14 (c) temos a diferença entre ambas as médias.

Figura 5.14 – Precisão média de acordo com a (a) ausência e (b) presença de um determinado índice e a diferença entre estas (c).

Small	60.32	Ifix	63.47	Ouro	-0.60
Eem	62.53	Ouro	63.51	Ifix	-0.47
Dolar	63.24	Selic	63.66	Selic	-0.31
SP500	63.72	Imab	63.79	Imab	-0.05
Imab	63.84	SP500	63.92	SP500	0.20
Ifix	63.95	Dolar	64.39	Dolar	1.15
Selic	63.97	Eem	65.11	Eem	2.58
Ouro	64.11	Small	67.48	Small	7.15

(a) (b) (c)

Fonte: Dos Autores.

Algumas conclusões importantes podem ser levantadas a partir dessas observações:

- O índice SMLL apresentou a maior diferença, uma vez que teve o melhor desempenho quando foi incluído nos resultados ao mesmo tempo que teve o pior desempenho quando foi excluído. A relevância do SMLL está ligada à sua proximidade e correlação com o índice Ibovespa, uma vez que existem empresas que compõem ambos os índices.
- Enquanto metade dos índices causaram impacto em sua exclusão, observa-se também que a outra metade levou a um ganho de performance quando da sua exclusão. Analisando-se os percentuais, não se pode ser objetivo a ponto de se afirmar que esses índices acabam piorando o desempenho dos algoritmos.
- Também se destaca o índice EEM que obteve a segunda melhor performance quando incluído e a segunda pior quando excluído. Esse índice possui uma boa correlação com o Ibovespa, uma vez que o Ibovespa tende a seguir a tendência dos mercados de países emergentes.

Na sequência, analisou-se os dois extremos dos resultados, aqueles com precisão nula ou máxima. Das 9432 configurações utilizadas, 131 obtiveram pelo menos um algoritmo com precisão máxima, enquanto que 10 obtiveram todos os algoritmos com precisão máxima. A seguir são apresentadas algumas informações relevantes sobre estes resultados:

- Dos 131 que tiveram pelo menos um algoritmo com 100%:
    - 105 utilizaram a configuração de conversão binária e 26 a de conversão multinível;
    - Todos utilizaram amostras pequenas, com data inicial mínima de 2014 gerando a maior amostra (72);
    - 51,91% dos resultados (68 de 131) utilizaram a amostra mínima (36) com ano inicial em 2018;
    - A utilização da configuração de sentimento foi responsável por 83 das 131 amostras (63,36%).
- Dos 10 que obtiveram todos os algoritmos com 100%:
    - 90% utilizaram o índice Small;
    - 100% utilizaram a configuração de conversão binária;
    - Todos utilizaram amostras pequenas, com data inicial mínima de 2015 gerando a maior amostra (60);
    - Houveram dois resultados para o ano inicial 2015, 4 para 2016 e 4 para 2018;
    - 40% utilizaram a configuração de sentimento.

Pode-se novamente verificar que o índice SMLL acaba sendo um índice com grande relevância para o resultado do Ibovespa. Em situações com pouca amostragem, o resultado pode ser facilmente influenciado por uma situação macroeconômica favorável, onde todos os principais índices acabam apresentando uma mesma tendência definida.

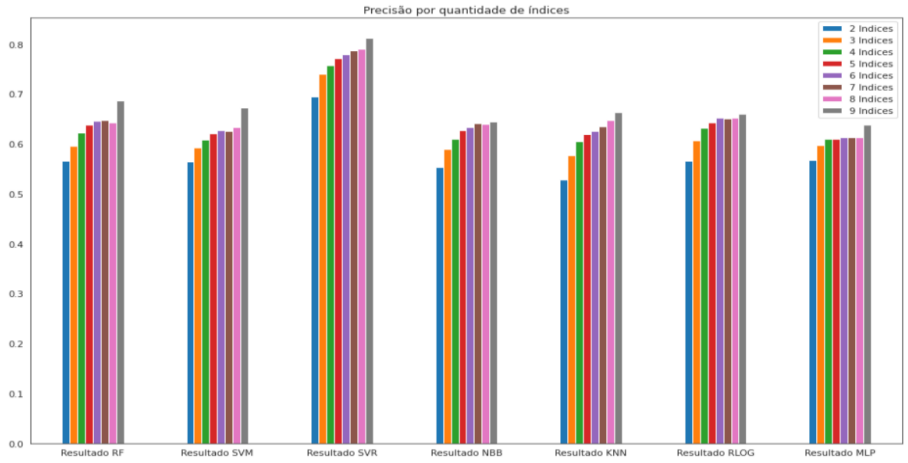
Também foi constatado que oito conjuntos de configurações de entrada tiveram algoritmos de ML que apresentaram precisão de 0%. Dentre estas cabe destacar que:

- Todas utilizaram a configuração de conversão multinível;
- 50% utilizaram a configuração de sentimento;
- Todas utilizaram a menor amostra possível (36), com data inicial de 2018;
- 75% das entradas tiveram com um de seus índices o IFIX e o IMA-B;
- Não houve predomínio de uma certa quantidade fixa de índices como entrada.

A configuração multinível tem se mostrado prejudicial para a obtenção de algoritmos mais precisos, uma vez que inclui diferentes possibilidades de classificação. Aliado com pequenas amostragens e índices que muitas vezes tem correlação opostas, obtêm-se resultados desfavoráveis para essa configuração.

A quantidade de índices escolhidos pode impactar no desempenho dos algoritmos. Nesta seção, são detalhados os resultados para cada tamanho de conjunto de índices de entrada, desde um único índice com o Ibovespa, até os 8 índices juntos. Na Figura 5.15 tem-se a média dos algoritmos para cada conjunto de índice. Observa-se de imediato que a precisão aumenta conforme a quantidade de índices aumenta. Observa-se também que em alguns algoritmos a utilização de uma quantidade moderada de índices como 6 ou 7 acaba resultando em um desempenho superior do que com 8 índices. Será visto em mais detalhes as estatísticas para cada conjunto.

Figura 5.15 – Precisão média dos algoritmos por conjunto.



Fonte: Dos Autores.

5.2.1.1 Conjunto com 2 Índices

A Figura 5.16 apresenta a precisão média para cada um dos algoritmos quando apenas 2 índices são utilizados como entrada. Conforme já observado, o algoritmo SVR apresenta desempenho superior. Filtrando-se os resultados de acordo com a utilização de sentimento verifica-se que a não utilização apresentou um resultado levemente superior. Quanto ao tipo de conversão, observa-se que o desempenho é impactado negativamente quando se aplica a conversão multinível. A Tabela 5.9 ilustra essas duas observações, agrupando os resultados conforme o tipo de conversão e o sentimento.

Figura 5.16 – Desempenho médio dos algoritmos para 2 índices.

```
Resultado RF      56.55
Resultado SVM     56.37
Resultado SVR     69.45
Resultado NBB     55.31
Resultado KNN     52.84
Resultado RLOG    56.53
Resultado MLP     56.77
```

Fonte: Dos Autores.

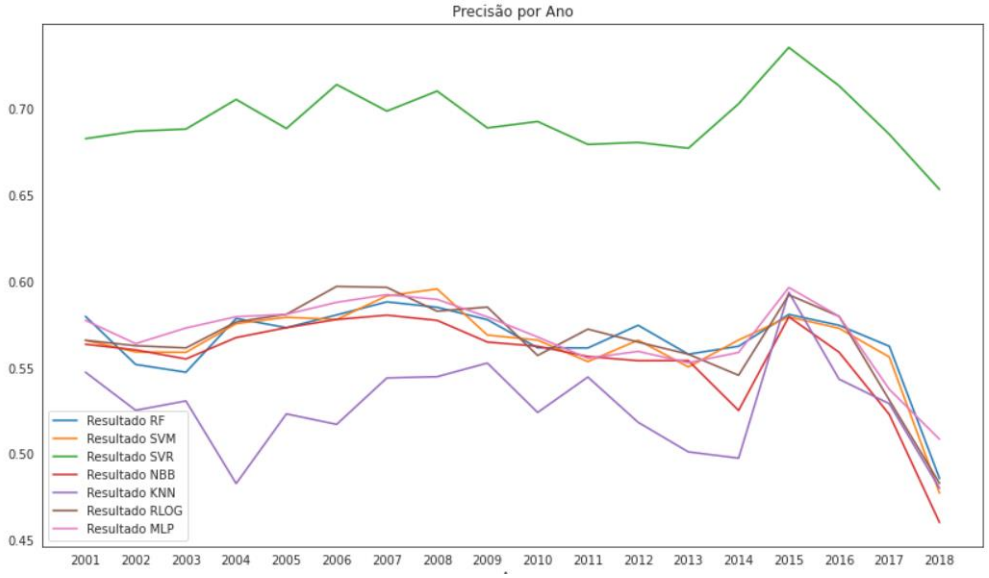
Tabela 5.9 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 2 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	70,05	69,71	69,69	69,76	64,5	69,44	69,71
2	True	69,66	69,08	68,75	68,78	67,05	68,30	68,66
N	False	43,83	43,83	68,12	41,3	40,42	44,22	44,23
N	True	42,65	42,86	71,25	41,39	39,41	44,14	44,47

Fonte: Dos Autores.

Observa-se na Figura 5.17 a evolução da precisão dos algoritmos de acordo com o ano utilizado como delimitador da amostra. Para o conjunto de entrada que utiliza apenas 2 índices o desempenho do algoritmo SVR é largamente superior aos demais, fato já esperado. A grande maioria dos algoritmos tem um desempenho ao longo dos anos sem grandes desvios, com exceção do KNN que se distancia dos demais durante por uma grande quantidade de anos. O desempenho tende a ser superior nos anos iniciais para quase todos os algoritmos, havendo uma pequena queda em 2014, seguido por um pico em 2015, terminando por cair de forma consecutiva nos anos seguintes até atingir um mínimo no ano 2018, de menor amostragem.

Figura 5.17 – Precisão por ano para conjunto de 2 índices.



Fonte: Dos Autores.

5.2.1.2 Conjunto com 3 Índices

Analisando-se a Figura 5.18 e a Tabela 5.10, observa-se o mesmo padrão de comportamentos da precisão média dos algoritmos e do tipo de conversão. Para o conjunto de 3 índices de entrada a utilização de sentimento trouxe um pequeno aumento na precisão.

O comportamento ao longo dos anos se assemelha ao conjunto de 2 entradas visto anteriormente. Na Figura 5.19 pode-se identificar que o algoritmo KNN apresentou uma evolução em comparação ao experimento anterior, aproximando seu desempenho dos demais índices. Observa-se desta vez que os maiores picos de desempenho estão nos anos de 2004 e 2008 e verifica-se uma queda acentuada no ano de 2018.

Figura 5.18 – Desempenho médio dos algoritmos para 3 índices.

Resultado RF	59.59
Resultado SVM	59.29
Resultado SVR	73.93
Resultado NBB	58.87
Resultado KNN	57.69
Resultado RLOG	60.61
Resultado MLP	59.63

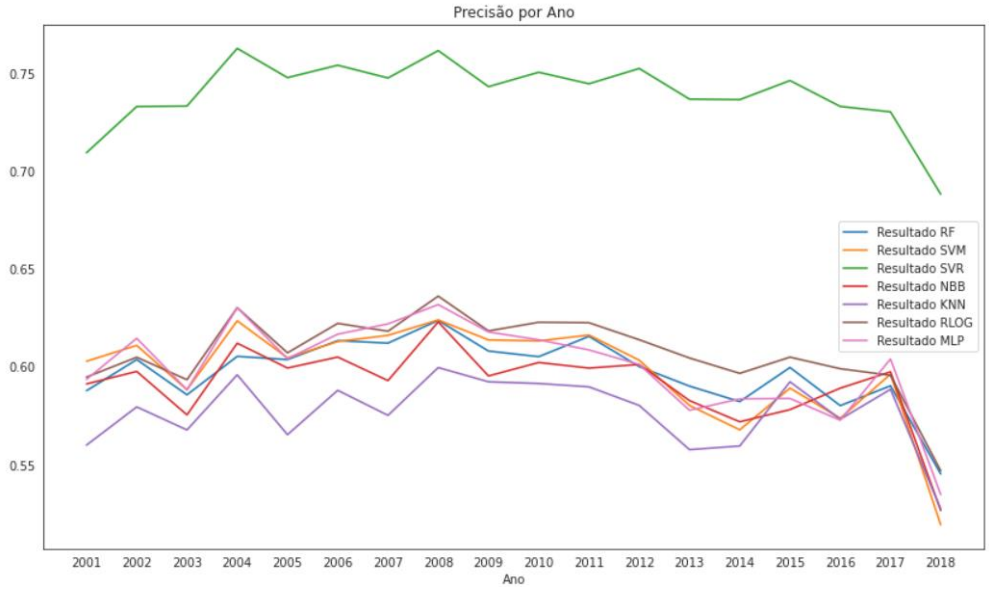
Fonte: Dos Autores.

Tabela 5.10 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 3 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	73,43	73,31	73,13	73,53	71,66	73,67	73,39
2	True	74,47	74,01	73,84	75,32	72,53	75,27	74,28
N	False	45,56	45,46	73,85	42,68	43,38	46,88	46,30
N	True	44,89	44,37	74,89	43,93	43,21	46,62	44,55

Fonte: Dos Autores.

Figura 5.19 – Precisão por ano para conjunto de 3 índices.



Fonte: Dos Autores.

5.2.1.3 Conjunto com 4 Índices

Com base na Figura 5.20 e na Tabela 5.11, verifica-se os mesmos destaques para melhor algoritmo (SVR) e a queda de desempenho com a utilização da conversão multinível. O uso da entrada de sentimento trouxe um leve ganho para o conjunto de 4 índices.

Ao se observar a Figura 5.21 pode-se identificar que o algoritmo KNN encontra-se com um comportamento bem próximo da maior parte dos algoritmos. Para esse conjunto de entrada percebe-se que o algoritmo RLOG começa a apresentar um desempenho superior aos demais por grande parte do período, mas nunca superior ao SVR. Não se identifica picos representativos para todos os algoritmos, apenas um movimento de alta no ano de 2002, 2008 e 2015, seguindo novamente pela queda no ano 2018.

Figura 5.20 – Desempenho médio dos algoritmos para 4 índices.

```
Resultado RF      62.20
Resultado SVM     60.87
Resultado SVR     75.66
Resultado NBB     61.01
Resultado KNN     60.45
Resultado RLOG    63.16
Resultado MLP     60.93
```

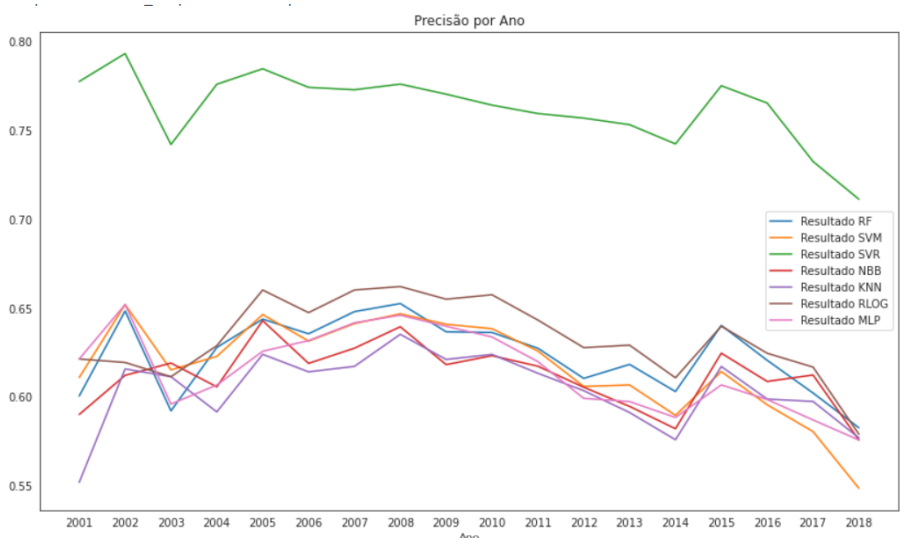
Fonte: Dos Autores.

Tabela 5.11 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 4 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	76,56	75,62	75,81	77,74	74,81	77,33	76,14
2	True	77,31	76,19	75,78	78,43	76,16	77,72	76,49
N	False	46,87	45,80	75,19	42,75	45,29	48,74	45,60
N	True	48,06	45,88	75,85	45,12	45,53	48,85	45,47

Fonte: Dos Autores.

Figura 5.21 – Precisão por ano para conjunto de 4 índices.



Fonte: Dos Autores.

5.2.1.4 Conjunto com 5 Índices

A Figura 5.22 e a Tabela 5.12 apresentam a superioridade do SVR e também o fraco desempenho da utilização da conversão multinível. O uso da entrada de sentimento não representou um movimento claro, sendo mais efetiva para alguns algoritmos enquanto para outros acaba diminuindo a precisão.

Na Figura 5.23, além do destacado KNN, o algoritmo RLOG mantém sua relevância em relação a média por grande parte do período. O ano de 2008 apresenta o maior pico de desempenho, seguido por outro pico em 2015. As mínimas continuam ocorrendo nos períodos finais.

Figura 5.22 – Desempenho médio dos algoritmos para 5 índices.

```
Resultado RF      63.77
Resultado SVM     62.00
Resultado SVR     77.03
Resultado NBB     62.66
Resultado KNN     61.84
Resultado RLOG    64.24
Resultado MLP     61.01
```

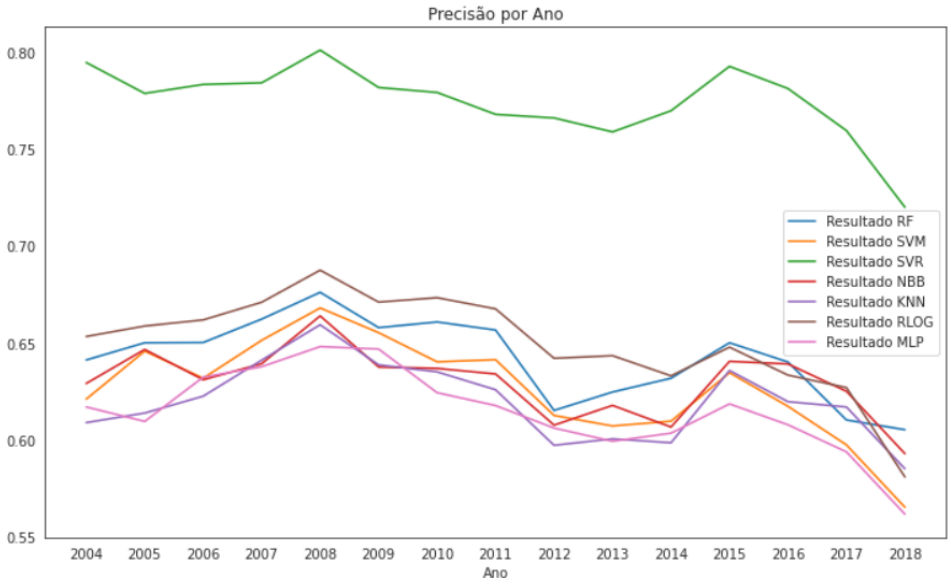
Fonte: Dos Autores.

Tabela 5.12 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 5 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	78,61	77,54	77,12	79,75	77,29	79,23	77,57
2	True	78,83	76,83	76,59	80,43	78,11	79,29	76,47
N	False	49,24	46,90	76,97	44,51	46,36	49,54	45,03
N	True	48,42	46,71	77,45	45,94	45,62	48,90	44,98

Fonte: Dos Autores.

Figura 5.23 – Precisão por ano para conjunto de 5 índices.



Fonte: Dos Autores.

5.2.1.5 Conjunto com 6 Índices

A Figura 5.24 e a Tabela 5.13 ilustram o desempenho superior do SVR e da conversão binária. A entrada de sentimento gerou um pequeno ganho para a maioria dos algoritmos.

Na Figura 5.25, percebe-se que o ano inicial de 2005 trouxe o maior desempenho para todos os algoritmos. O algoritmo RLOG conseguiu se manter como o segundo com maior desempenho ao longo de quase todo o período. Com o incremento do ano inicial, o desempenho dos algoritmos apresenta uma tendência decrescente, terminado na mínima em 2018.

Figura 5.24 – Desempenho médio dos algoritmos para 6 índices.

```
Resultado RF      64.49
Resultado SVM     62.62
Resultado SVR     77.97
Resultado NBB     63.39
Resultado KNN     62.58
Resultado RLOG    65.15
Resultado MLP     61.27
```

Fonte: Dos Autores.

Tabela 5.13 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 6 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	79,41	77,72	77,74	81,01	78,55	79,79	77,96
2	True	80,11	77,97	77,90	81,38	78,82	80,28	77,04
N	False	49,60	47,44	77,16	45,56	46,20	50,35	44,68
N	True	48,82	47,35	79,07	45,60	46,75	50,18	45,40

Fonte: Dos Autores.

Figura 5.25 – Precisão por ano para conjunto de 6 índices.

Fonte: Dos Autores.

5.2.1.6 Conjunto com 7 Índices

A superioridade do algoritmo SVR e da conversão binária estão representadas no gráfico da Figura 5.26 e na Tabela 5.14, respectivamente. A entrada de sentimento para esse conjunto de 7 índices gerou uma queda no desempenho para todos os algoritmos.

Na Figura 5.27, percebe-se a liderança do SVR em termos de precisão, acompanhado pelo algoritmo RLOG. Um comportamento decrescente está presente com o avanço do ano inicial da amostra. Os picos de desempenho estão representados nos anos 2008, 2010 e 2015.

Figura 5.26 – Desempenho médio dos algoritmos para 7 índices.

```
Resultado RF      64.74
Resultado SVM     62.53
Resultado SVR     78.72
Resultado NBB     64.05
Resultado KNN     63.54
Resultado RLOG    65.04
Resultado MLP     61.34
```

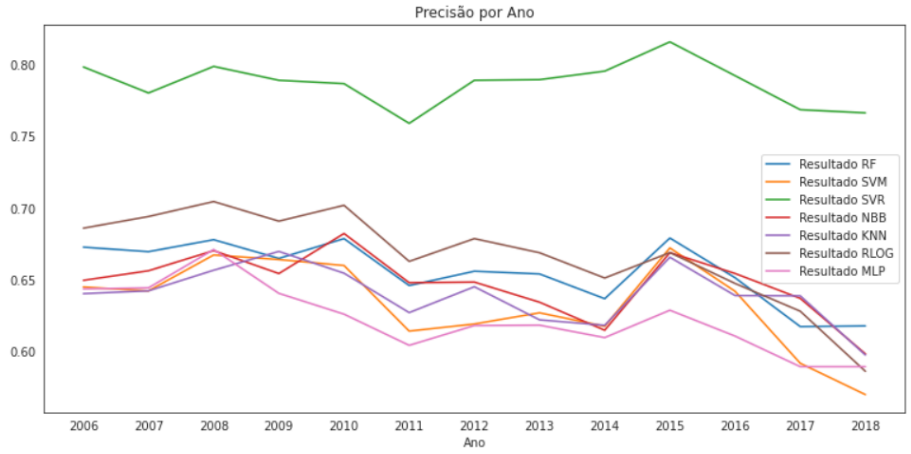
Fonte: Dos Autores.

Tabela 5.14 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 7 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	80,47	78,71	78,77	82,34	80,80	80,31	78,30
2	True	80,30	78,01	77,90	81,91	79,42	80,04	76,56
N	False	50,39	47,60	79,37	46,48	47,42	50,71	44,27
N	True	47,80	45,81	78,84	45,47	46,53	49,11	46,22

Fonte: Dos Autores.

Figura 5.27 – Precisão por ano para conjunto de 7 índices.



Fonte: Dos Autores.

5.2.1.7 Conjunto com 8 Índices

Seguindo o padrão dos outros conjuntos analisados, o algoritmo SVR e a conversão binária apresentam desempenhos superiores aos demais, conforme se verifica na Figura 5.28 e na Tabela 5.15. A utilização de sentimento não apresentou uma direção definida, elevando a precisão de alguns algoritmos enquanto diminui para outros.

Na Figura 5.29, percebe-se que os algoritmos possuem uma maior variação de desempenho, não seguindo uma trajetória harmoniosa. Com exceção do SVR que está em destaque, os outros algoritmos alternam precisões mais elevadas com outras mais baixas. Os picos e vales são variados de acordo com os algoritmos e não se encontra um ano em que a maioria convirja, com exceção de uma alta em 2009 e 2015.

Figura 5.28 – Desempenho médio dos algoritmos para 8 índices.

```
Resultado RF      64.23
Resultado SVM     63.35
Resultado SVR     78.99
Resultado NBB     63.99
Resultado KNN     64.71
Resultado RLOG    65.14
Resultado MLP     61.25
```

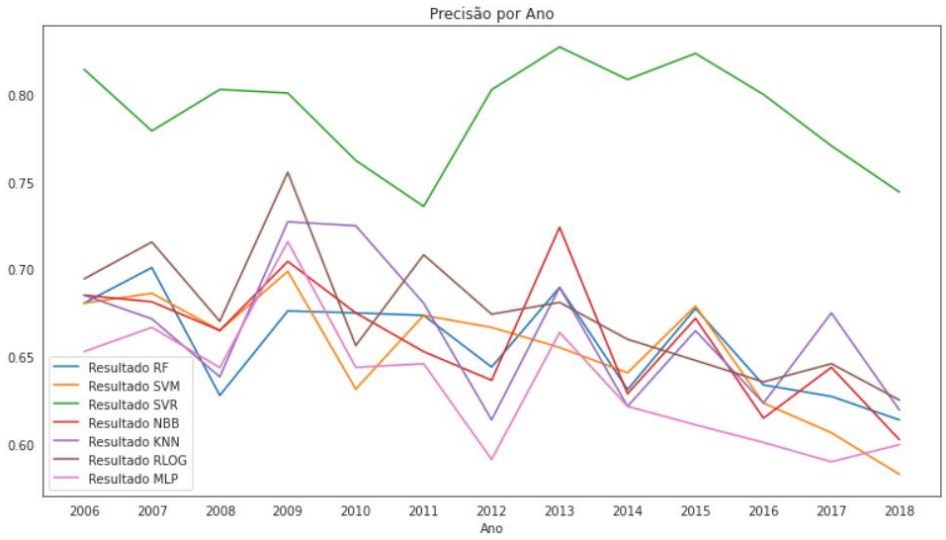
Fonte: Dos Autores.

Tabela 5.15 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 8 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	79,31	79,43	78,29	81,84	80,45	80,20	78,22
2	True	81,15	78,85	77,74	84,27	82,07	82,10	78,10
N	False	48,18	46,94	80,62	44,13	48,23	49,58	44,27
N	True	48,29	48,19	79,09	45,70	48,09	48,67	44,39

Fonte: Dos Autores.

Figura 5.29 – Precisão por ano para conjunto de 8 índices.



Fonte: Dos Autores.

5.2.1.8 Conjunto com 9 Índices

Uma vez que o algoritmo SVR e a conversão binária apresentam desempenhos superiores aos demais, como ilustra a Figura 5.30 e a Tabela 5.16, tem-se que para todas as quantidades de índices de entrada essas configurações são sempre as mais otimizadas. A utilização de sentimento novamente não apresentou uma tendência definida, alternando bons e maus resultados para a precisão dos algoritmos.

Na Figura 5.31, percebe-se que os algoritmos possuem um movimento mais suave em sua precisão, sem grandes picos ou vales. Além do destaque do SVR, o algoritmo RF tem um comportamento crescente conforme se diminui a amostra. Os picos e vales não estão tão presentes, mas diferentemente de todos os conjuntos anteriores, o desempenho não despenca no ano final de 2018.

Figura 5.30 – Desempenho médio dos algoritmos para 9 índices.

```
Resultado RF      68.59
Resultado SVM     67.22
Resultado SVR     81.17
Resultado NBB     64.43
Resultado KNN     66.32
Resultado RLOG    65.92
Resultado MLP     63.73
```

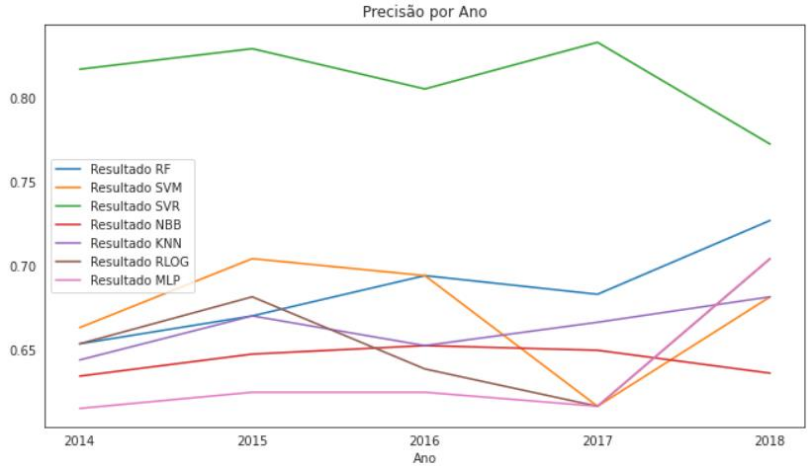
Fonte: Dos Autores.

Tabela 5.16 – Desempenho dos algoritmos de acordo com a codificação, sentimento e algoritmo para o conjunto de 9 índices.

Tipo	*Sentimento*	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*
2	False	85,24	84,28	79,24	86,92	87,69	83,91	81,89
2	True	82,54	85,47	83,79	85,55	84,14	82,05	82,84
N	False	52,20	48,40	79,95	44,02	47,05	49,69	47,37
N	True	54,37	50,73	81,69	41,23	46,40	48,00	42,83

Fonte: Dos Autores.

Figura 5.31 – Precisão por ano para conjunto de 9 índices.



Fonte: Dos Autores.

5.2.2 Análise em Anos

Outra análise foi realizada visando compreender se em cada ano utilizado como ano inicial e limitador da amostra, o maior conjunto de índices de entrada implicava em uma maior precisão. Essa análise se mostrou válida para uma grande quantidade de anos e algoritmos estudados. Os resultados detalhados para cada ano podem ser vistos no Apêndice A. Contudo, alguns resultados não seguiram essa tendência.

Na Tabela 5.17 estão representadas as ordens dos conjuntos de acordo com a precisão média para cada algoritmo. Em vermelho estão identificadas as distorções nas quais não seguem o padrão de quanto mais índices melhor o desempenho.

Em uma análise horizontal, verifica-se que nos anos de 2008, 2011, 2016 e 2018 para nenhum dos algoritmos seguiu-se o padrão de quantidade de índices. Apenas nos anos de 2002, 2003, 2006 e 2013 o padrão foi seguido para todos os algoritmos. Observa-se também que

quanto mais índices estão envolvidos maior é a probabilidade de haver para um ano alguma distorção. Dos 12 anos em que estão presentes 7 ou mais conjuntos (2007-2018), apenas no ano de 2013 não houve uma distorção para nenhum dos algoritmos.

Numa análise vertical, observa-se que em metade dos anos (9 de 18) a média do desempenho dos algoritmos também não respeitou o padrão. Com exceção do algoritmo SVR, todos os outros apresentaram em mais de 50% dos anos resultados que fogem do padrão. Destes, destaca-se o algoritmo MLP que em 13 dos 18 anos apresentou resultados divergentes.

Ao se analisar a média geral de cada algoritmo, esperava-se que o padrão ilustrado na Figura 5.32 fosse mantido. Contudo, ao se analisar as médias de todos os algoritmos para cada quantidade de índices de entrada, verifica-se que entre 6 e 8 índices o desempenho médio está bem próximo, justificando o fato de que uma parcela importante dos resultados vistos na Tabela 5.17 não respeitam o padrão.

Tabela 5.17 – Ordem decrescente de precisão dos conjuntos de índices.

Ano	*RF*	*SVM*	*SVR*	*NBB*	*KNN*	*RLOG*	*MLP*	*Média*
2001	4, 3, 2	4, 3, 2	4, 3, 2	3, 4, 2	3, 4, 2	4, 3, 2	4, 3, 2	4, 3, 2
2002	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2
2003	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2	4, 3, 2
2004	5, 4, 3, 2	3, 4, 5, 2	5, 4, 3, 2	5, 3, 4, 2	5, 3, 4, 2	5, 3, 4, 2	3, 5, 4, 2	5, 3, 4, 2
2005	6, 5, 4, 3, 2	6, 4, 5, 3, 2	6, 4, 5, 3, 2	6, 5, 4, 3, 2	6, 4, 5, 3, 2	6, 4, 5, 3, 2	6, 4, 5, 3, 2	6, 4, 5, 3, 2
2006	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2
2007	8, 6, 7, 5, 4, 3, 2	8, 6, 5, 7, 4, 3, 2	6, 5, 7, 8, 4, 3, 2	8, 6, 7, 5, 4, 3, 2	8, 6, 7, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 4, 5, 6, 3, 2	8, 7, 6, 5, 4, 3, 2
2008	6, 7, 5, 4, 8, 3, 2	6, 5, 7, 8, 4, 3, 2	6, 8, 5, 7, 4, 3, 2	6, 7, 8, 5, 4, 3, 2	6, 5, 7, 8, 4, 3, 2	7, 6, 5, 8, 4, 3, 2	7, 6, 5, 4, 8, 3, 2	6, 7, 5, 8, 4, 3, 2

2009	8, 6, 7, 5, 4, 3, 2	8, 7, 5, 6, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 5, 7, 4, 6, 3, 2	8, 7, 6, 5, 4, 3, 2
2010	7, 8, 6, 5, 4, 3, 2	7, 6, 5, 4, 8, 3, 2	7, 6, 5, 4, 8, 3, 2	7, 8, 5, 6, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	7, 6, 5, 4, 8, 3, 2	8, 4, 6, 7, 5, 3, 2	7, 8, 6, 5, 4, 3, 2
2011	8, 5, 6, 7, 4, 3, 2	8, 5, 6, 4, 3, 7, 2	6, 5, 4, 7, 3, 8, 2	8, 7, 5, 6, 4, 3, 2	8, 6, 7, 5, 4, 3, 2	8, 6, 5, 7, 4, 3, 2	8, 6, 4, 5, 3, 7, 2	8, 6, 5, 7, 4, 3, 2
2012	7, 6, 8, 5, 4, 3, 2	8, 6, 7, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	7, 8, 6, 5, 4, 3, 2	7, 6, 8, 4, 5, 3, 2	7, 8, 6, 5, 4, 3, 2	7, 6, 5, 3, 4, 8, 2	7, 8, 6, 5, 4, 3, 2
2013	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2	8, 7, 6, 5, 4, 3, 2
2014	9, 7, 5, 8, 6, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	8, 9, 7, 6, 5, 4, 3, 2	8, 9, 7, 5, 6, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2
2015	7, 8, 9, 6, 5, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	8, 7, 6, 9, 5, 4, 2, 3	9, 7, 8, 5, 6, 4, 2, 3	9, 7, 6, 8, 5, 4, 3, 2	7, 6, 9, 5, 8, 4, 2, 3	9, 7, 8, 6, 5, 4, 3, 2
2016	9, 6, 7, 5, 8, 4, 3, 2	9, 7, 6, 8, 5, 4, 3, 2	9, 8, 6, 7, 5, 4, 3, 2	7, 9, 6, 5, 8, 4, 3, 2	9, 7, 6, 8, 5, 4, 3, 2	7, 6, 9, 8, 5, 4, 3, 2	9, 6, 7, 5, 8, 4, 2, 3	9, 7, 6, 5, 8, 4, 3, 2
2017	9, 8, 7, 5, 6, 4, 3, 2	9, 8, 6, 5, 3, 7, 4, 2	9, 8, 7, 6, 5, 4, 3, 2	9, 8, 7, 6, 5, 4, 3, 2	8, 9, 7, 6, 5, 4, 3, 2	8, 6, 7, 5, 4, 9, 3, 2	9, 3, 5, 8, 7, 4, 6, 2	9, 8, 7, 6, 5, 4, 3, 2
2018	9, 6, 7, 8, 5, 4, 3, 2	9, 8, 6, 7, 5, 4, 3, 2	9, 7, 6, 8, 5, 4, 3, 2	9, 8, 6, 7, 5, 4, 3, 2	9, 8, 6, 7, 5, 4, 3, 2	9, 8, 6, 7, 5, 4, 3, 2	9, 8, 7, 6, 4, 5, 3, 2	9, 8, 6, 7, 5, 4, 3, 2

Fonte: Dos Autores.

Figura 5.32 – Desempenho médio dos algoritmos para cada um dos conjuntos.

```
9 Indices    68.20
8 Indices    65.95
7 Indices    65.71
6 Indices    65.35
5 Indices    64.65
4 Indices    63.47
3 Indices    61.37
2 Indices    57.69
```

Fonte: Dos Autores.

5.2.3 Considerações sobre os resultados das análises do MTFA

Nesta seção é apresentado um resumo sobre os resultados do MTFA. De uma forma mais objetiva, os pontos citados a seguir destacam os principais resultados obtidos nas simulações do algoritmo MTFA.

- Conversão binária é muito mais efetiva, devido à utilização de algoritmos que estão mais adaptados a esse tipo de dado.
- Semelhança nos resultados dos métodos de maneira geral, aparecendo algumas diferenças apenas quando se visualiza com outro ponto de vista.
- Algoritmo de adaptação do SVR influenciou fortemente o resultado para a conversão multinível. Somente o SVR teve desempenho multinível próximo ao desempenho do binário.
- Quanto maior a amostra, e, portanto, mais complexo o comportamento dos dados, pior o desempenho.
- Alguns períodos de análise podem ter sido muito favoráveis por serem períodos nas quais os índices utilizados eram previsíveis, ou possuíam comportamento muito próximo.
- Pequenas amostras produziram extremos, tantos resultados próximos a 0 quanto próximos a 1.
- Viu-se que um número pequeno de índices apresenta desempenho inferior e, em média, à medida em que se aumenta a quantidade, o desempenho também aumenta. Contudo, observou-se que nem sempre esse fato se manteve, havendo métodos que ofereceram valores maiores para uma menor quantidade de índices.

- Pode-se afirmar que a quantidade de índices tende a influenciar positivamente o desempenho, mas não se pode afirmar qual a quantidade ideal de índices, uma vez que a depender do ano e do algoritmo pode haver um número ideal.
- A utilização de sentimento não trouxe uma melhora considerável para o desempenho. Na maior parte das análises vistas elevou o desempenho em valores muito modestos.
- O índice SMLL se portou com o índice de maior impacto no resultado dos algoritmos, o que de certa forma reforça a ligação dele com o Ibovespa, uma vez que compartilham algumas empresas.
- Observa-se que em os dois anos que apresentaram as maiores médias dos desempenhos dos algoritmos foram os anos de 2008 e 2015, justamente os anos impactados por crises econômicas, tanto global quanto nacional. Momentos de estresse no mercado podem sugerir que os movimentos acabam por se tornar conhecidos, visto que o mercado tende a reagir de forma semelhante a esses eventos: alta de juros, alta de inflação, queda de bolsa, alta do dólar, entre outros eventos.
- Observa-se no estudo que o fraco desempenho da conversão multinível pode estar relacionada também à escolha dos algoritmos. Os algoritmos utilizados e suas configurações podem ter sido mais adequados à classificação binária.

6 CONCLUSÃO

Esse trabalho apresentou uma análise de diversos segmentos sobre os dados históricos da bolsa de valores brasileira. Conduziu-se um estudo acerca de como as variáveis macroeconômicas têm relação com o desempenho do índice de ações, em especial a perspectiva de aumento da taxa de juros que diminui a atratividade da renda variável, ao passo que a perspectiva oposta acentua a procura por investimentos mais arriscados e lucrativos. Observou-se ainda o impacto das crises econômicas nas diversas variáveis macroeconômicas, com a disparada de dólar e dos juros para conter a inflação, e a queda no PIB com o aumento do desemprego. Foi importante também observar o impacto do dólar, moeda forte que ajuda a representar o fluxo de investimentos estrangeiros no país e como essa dinâmica afeta a bolsa em diversos períodos.

Analisou-se nesse estudo algumas situações comportamentais dos investidores. Descobriu-se que, num período de vinte anos, os investidores da bolsa brasileira tiveram uma preferência compradora nos últimos meses do ano, ao passo que janeiro e maio são meses com uma pressão vendedora. Também se observou uma tendência dos investidores a comprarem mais ações no início do mês e de venderem no final do mês anterior, o que pode impactar investidores que decidam realizar lucro. Em relação aos dias com variações positivas e negativas, não se observou uma relação relevante entre o retorno anual e a quantidade de dias com retorno semelhante. No histórico analisado, verificou-se uma leve tendência dos investidores venderem mais nas segunda-feira, após refletirem possíveis notícias negativas ocorridas no final de semana. O investidor que não está fortemente exposto à bolsa em períodos de grande volatilidade ou que aproveita as grandes altas após as grandes quedas, consegue obter um rendimento muito superior à média.

Por fim, verificou-se o desempenho de diferentes índices em relação ao Ibovespa e suas correlações. Entre todos, o Ibovespa obteve o maior rendimento e crescimento, se considerado o retorno em moeda local, contudo, com o forte desempenho do dólar, índices cotados em dólar, quando convertidos para moeda local, terminaram por ter a maior rentabilidade.

A partir do entendimento obtido com a etapa de análise de dados, partiu-se para a criação de um algoritmo que pudesse prever se o desempenho mensal do Ibovespa seria positivo ou negativo. A adoção de índices de diferentes tipos possibilitou verificar quais conjuntos de índices teriam melhor ou pior impacto na qualidade da previsão. Observou-se que a partir do aumento do número de índices se obteve uma melhora no desempenho.

Buscou-se utilizar dois tipos de conversão de dados que permitiu concluir que a classificação binária em meses de alta e de baixa apresentou resultados mais elevados quando não se foi levado em conta a magnitude do resultado mensal dos outros índices. A adoção de uma variável que indicasse um efeito comportamental de sazonalidade, acabou por ter pouca influência sobre os resultados. As diferentes janelas temporais analisadas, que, por sua vez, representavam diferentes tamanhos de *datasets*, mostraram que pequenas janelas tendem a apresentar resultados piores.

A utilização de diferentes técnicas de ML mostrou ser possível obter resultados razoáveis e próximos independente da técnica, e, ao mesmo tempo, deixou aberta a possibilidade de se obter resultados ainda melhores a partir do momento em que se opte por otimizar estas técnicas. Observou-se que algumas oscilaram de desempenho em virtude das diferentes configurações apresentadas, mas os resultados em média foram satisfatórios, sendo superiores a 60% em média e próximos a 80% quando utilizada a configuração binária.

6.1 Trabalhos futuros

Em continuidade a este trabalho, vários domínios podem ser aprimorados e aprofundados. Nesta seção serão destacados cada um desses domínios e que tipos de avanços podem ser feitos para contribuir com novas descobertas.

6.1.1 Análise Setorial:

Durante a realização do trabalho buscou-se limitar o escopo das análises e índices utilizados de forma a construir uma metodologia que pudesse ser expandida. Os índices utilizados inicialmente são aqueles mais comuns no mercado e mais facilmente acessíveis ao investidor não profissional.

Pode-se buscar outras relações entre o Ibovespa e alguns índices que representam setores do mercado financeiro, como empresas da construção civil, alimentícias, exportadoras, de minérios, agrícolas, financeiras, de combustíveis fósseis, aviação entre outros. Hoje existem diversos índices setoriais que podem ser utilizados e, se estudados, podem se mostrar mais ou menos influentes em relação ao Ibovespa.

6.1.2 Outros índices

Além da análise setorial, através de índices das empresas envolvidas em cada um dos setores, outros índices estão à disposição do investidor. Embora apresentem uma dificuldade adicional para serem obtidos, índices do mercado internacional podem contribuir para um melhor entendimento do comportamento da bolsa brasileira.

Cabe destacar os índices de preços de *commodities* como minério de ferro, petróleo, soja e milho. Esses produtos por serem negociados no mercado internacional possuem forte ligação com o dólar e com o estado da cadeia de produção global. O preço das *commodities* tem uma grande influência sobre as empresas que as utilizam ou produzem. Dependendo do momento dos preços, empresas presentes na bolsa podem se beneficiar ou ter seus resultados prejudicados, impactando de alguma forma na cotação da bolsa como um todo.

Também poderiam ser utilizados outros índices que agrupam as empresas de acordo com suas características como governança corporativa, pagadora de dividendos, crescimento, ecológica entre tantas outras formas. Um índice muito utilizado no mercado financeiro é o VIX [23](Índice de Volatilidade da Chicago Board Options Exchange) que mede a volatilidade do mercado de opções, uma métrica que ajuda a representar a aversão ao risco do investidor.

No Brasil também podem ser considerados alguns outros índices fornecidos pelo Banco Central e calculados pelos órgãos de pesquisa como o índice de atividade econômica, a balança comercial, o índice de confiança do consumidor entre outros. Cada um destes pode ser expresso em uma série temporal e ter seu comportamento analisado em relação ao Ibovespa, trazendo mais informações e novas descobertas sobre seus relacionamentos.

6.1.3 Janela Temporal

O trabalho também pode evoluir na escolha de novas janelas temporais. Uma forma de se obter mais dados seria optar por análises de períodos inferiores a um mês, como análise semanal e diária, esta última a mais presente na literatura. Outra forma de se aumentar a quantidade de dados disponíveis seria utilizando algumas plataformas pagas do mercado financeiro que contabilizam bases de dados muito maiores e possuem registros mais ricos e antigos que as bases de acesso livre utilizadas neste trabalho.

[23] https://www.cboe.com/tradable_products/vix/

6.1.4 Aperfeiçoamento de Algoritmos

O foco pode ser colocado no aperfeiçoamento dos algoritmos de ML utilizados. Poderá se escolher alguns algoritmos e realizar um estudo aprofundado sobre as melhores formas de parametrizar o algoritmo é de preparar os dados para atingir melhores resultados de precisão. Outra forma é buscar uma solução contendo um conjunto de algoritmos conectados.

6.1.5 Ações Específicas

Outra sugestão de expansão do trabalho seria analisar um conjunto de ações individuais ou ainda buscar quais as ações que melhor explicam ou refletem o comportamento do índice Bovespa. A seleção de ações permite identificar em quais momentos um dado tipo de ação está tendo papel de destaque e por sua vez influenciando o Ibovespa, ao mesmo tempo que permite encontrar aquelas cujo comportamento seja completamente oposto ou alheio ao índice.

Outra forma de análise é substituir o estudo do comportamento do Ibovespa pelo estudo do comportamento de ações específicas. Poderia, portanto, ser avaliado como um seleto grupo de ações é influenciado pelos outros índices discutidos no trabalho e também propostos nesta seção.

6.1.6 Finanças comportamentais.

Além dos trabalhos relacionados apresentados, existem ainda diversos estudos sobre finanças comportamentais. Alguns trabalhos discutem vieses comportamentais presentes em muitos investidores e outros comportamentos de manada presentes nas tomadas de decisão.

Além de focar num estudo mais detalhados desses comportamentos e destes vieses, o foco do trabalho poderia ser em construir uma representação ou modelo para tais efeitos comportamentais, sendo em forma de série temporal ou não.

REFERÊNCIAS

Shiller, R. J., Project Muse. Irrational Exuberance: Revised and Expanded Third Edition. Princeton, Princeton University Press, 2015.

Afonso J. R., Araújo E. C., Bernardo Guelber Fajardo: The role of fiscal and monetary pol-icies in the Brazilian economy: Understanding recent institutional reforms and economic changes. The Quarterly Review of Economics and Finance, vol. 62, pp. 41-55, 2016.

Ahmed B.: Understanding the impact of investor sentiment on the price formation process: A review of the conduct of American stock markets. The Journal of Economic Asymme-tries, vol. 22, 2020.

An Y., Sun M., Gao C., Han D., Li X.: Analysis of the impact of crude oil price fluctua-tions on China's stock market in different periods - Based on time series network model. Physica A: Statistical Mechanics and its Applications, vol. 492, pp. 1016-1031, 2018.

Bouman, S., Jacobsen, B.: The Halloween Indicator, ''Sell in May and Go Away'': another puzzle. Am. Econ. Rev. 92, pp. 1618-1635, 2002.

Chen T., Chien C.: Size effect in January and cultural influences in an emerging stock market: The perspective of behavioral finance. Pacific-Basin Finance Journal, vol. 19, pp. 208-229, 2011.

Chen Z., Daves P. R.: The January sentiment effect in the U.S. stock market. International Review of Financial Analysis, vol. 59, pp. 94-104, 2018.

Dichtl H., Drobetz W.: Sell in May and Go Away: Still good advice for investors?. International Review of Financial Analysis, vol. 38, pp. 29-43 2015.

Giacomel F., Pereira A.C.M., Galante R.: Improving Financial Time Series Prediction Through Output Classification by a Neural Network Ensemble. In: Chen Q., Hameur-lain A., Toumani F., Wagner R., Decker H. (eds). Database and Expert Systems Appli-cations. Globe 2015, DEXA 2015. Lecture Notes in Computer Science, vol. 9262. Springer, Cham. https://doi.org/10.1007/978-3-319-22852-5_28.

Gonzalez R. T., Padilha C. A., Barone D. A. C.: Ensemble system based on genetic algorithm for stock market forecasting. IEEE Congress on Evolutionary Computation (CEC), pp. 3102-3108, 2015.

Guo B., Luo X., Zhang Z.: Sell in May and Go Away: Evidence from China. Finance Research Letters, vol. 11, pp. 362-368, 2014.

Han T., Peng Q., Zhu Z., Shen Y., Huang H., Abid N.: A pattern representation of stock time series based on DTW. Physica A: Statistical Mechanics and its Applications, vol. 550, 2020.

He Z., O'Connor F., Thijssen J.: Is gold a Sometime Safe Haven or an Always Hedge for equity investors? A Markov-Switching CAPM approach for US and UK stock indices. International Review of Financial Analysis, vol. 60, pp. 30-37, 2018.

Henrique B., Sobreiro V., Kimura H.: Stock price prediction using support vector regression on daily and up to the minute prices. The Journal of Finance and Data Science, vol. 4, pp. 183-201, 2018.

Huang B., Ding Q., Sun G.: Stock Prediction based on Bayesian-LSTM. Proceedings of the 2018 10th International Conference on Machine Learning and Computing, pp. 128-133, 2018.

Huynh H., Dang L., Duong D.: New Model for Stock Price Movements Prediction Using Deep Neural Network. Proceedings of the Eighth International Symposium on Information and Communication Technology, pp. 57-62, 2017.

J. Christy Jackson, J. Prassanna, Md. Abdul Quadir, V. Sivakumar: Stock market analysis and prediction using time series analysis. Materials Today: Proceedings, 2021.

J. de Mello Assis, A. C. M. Pereira, R. C. e Silva: Designing Financial Strategies based on Artificial Neural Networks Ensembles for Stock Markets. International Joint Conference on Neural Networks (IJCNN), pp. 1-8, 2018.

Joosery B., Deepa G.: Comparative Analysis of Time-Series Forecasting Algorithms for Stock Price Prediction. Proceedings of the International Conference on Advanced Information Science and System, pp. 1-6, 2019.

Khattak A., Ullah H., Khalid H., Habib A., Asghar M., Kundi F.: Stock Market Trend Prediction using Supervised Learning. Proceedings of the Tenth International Symposium on Information and Communication Technology, pp. 85-91, 2019.

Kirikkaleli D.: The effect of domestic and foreign risks on an emerging stock market: A time series analysis. The North American Journal of Economics and Finance, vol. 51, 2020.

Kulaglic A., Üstündağ B.: Stock Price Forecast using Wavelet Transformations in Multiple Time Windows and Neural Networks. 3rd International Conference on Computer Science and Engineering (UBMK), pp. 518-521, 2018.

Liu H., Long Z.: An improved deep learning model for predicting stock market price time series. Digital Signal Processing, vol. 102, 2020.

Luo R.: Short-Term Stock Price Prediction Models Based on Economic Background. Proceedings of the 3rd International Conference on Information Technologies and Electrical Engineering, pp. 33-38, 2020.

Majumder M., Hossain M., Hasan M.: Indices prediction of Bangladeshi stock by using time series forecasting and performance analysis. International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-5, 2019.

McMillan D. G.: When and why do stock and bond markets predict US economic growth?. The Quarterly Review of Economics and Finance, vol. 80, pp. 331-343, 2021.

Ojo S., Owolawi P., Mphahlele M., Adisa J.: Stock Market Behaviour Prediction using Stacked LSTM Networks. International Multidisciplinary Information Technology and Engineering Conference (IMITEC), pp. 1-5, 2019.

Ouahilal M., Mohajir M., Chahhou M., Mohajir B.: Optimizing stock market price prediction using a hybrid approach based on HP filter and support vector regression. 4th IEEE International Colloquium on Information Science and Technology (CiSt), pp. 290-294, 2016.

Patil P., Wu C., Potika K., Orang M.: Stock Market Prediction Using Ensemble of Graph Theory, Machine Learning and Deep Learning Models. Proceedings of the 3rd International Conference on Software Engineering and Information Management, pp. 85-92, 2020.

Plastun A., Sibande X., Gupta R., Wohar M. E.: Historical evolution of monthly anomalies in international stock markets. Research in International Business and Finance, vol. 52, 2020.

Reyhani R., Moghadam A.: A heuristic method for forecasting chaotic time series based on economic variables. Sixth International Conference on Digital Information Management, pp. 300-304, 2011.

Ryota K., Tomoharu N.: Stock market prediction based on interrelated time series data. IEEE Symposium on Computers & Informatics (ISCI), pp. 17-21, 2012.

Salisu A., Vo X.: The behavior of exchange rate and stock returns in high and low interest rate environments. International Review of Economics & Finance, vol. 74, pp. 138-149, 2021.

Spierdijk L., Umar Z.: Stocks for the long run? Evidence from emerging markets. Journal of International Money and Finance, vol. 47, pp. 217-238, 2014.

Stona F., Morais I. A. C., Triches D.: Economic dynamics during periods of financial stress: Evidences from Brazil. International Review of Economics & Finance, vol. 55, pp. 130-144, 2018.

Tang J., Chen X.: Stock Market Prediction Based on Historic Prices and News Titles. Proceedings of the 2018 International Conference on Machine Learning Technologies, pp. 29-34, 2018.

Tang L. Pan H. Yao Y.: K-Nearest Neighbor Regression with Principal Component Analysis for Financial Time Series Prediction. Proceedings of the 2018 International Conference on Computing and Artificial Intelligence, pp. 127-131, 2018.

Toraman C., Başarir C.: The Long Run Relationship Between Stock Market Capitalization Rate and Interest Rate: Co-integration Approach. Procedia - Social and Behavioral Scienc-es, vol. 143, pp. 1070-1073, 2014.

Wang H.: Stock Price Prediction Based on Machine Learning Approaches. Proceedings of the 3rd International Conference on Data Science and Information Technology, pp. 1-5, 2020.

Wang Z., Ho S., Lin Z.: Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment. IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1375-1380, 2018.

Xia Y., Liu Y., Chen Z.: Support Vector Regression for prediction of stock trend. 6th International Conference on Information Management, Innovation Management and Industrial Engineering, pp. 123-126, 2013.

Zhang C. Y., Jacobsen B.: The Halloween indicator, “Sell in May and Go Away”: Every-where and all the time. Journal of International Money and Finance, vol. 110, 2021.

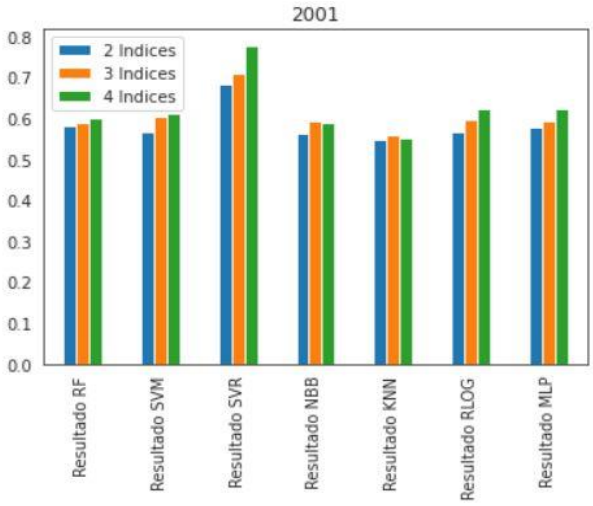
Zhang L., Aggarwal C., Qi G.: Stock Price Prediction via Discovering Multi-Frequency Trading Patterns. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2141-2149, 2017.

Zhao L., Wang L.: Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm. IEEE Fifth International Conference on Big Data and Cloud Computing, pp. 93-98, 2015.

APÊNDICE A – RESULTADOS POR ANO

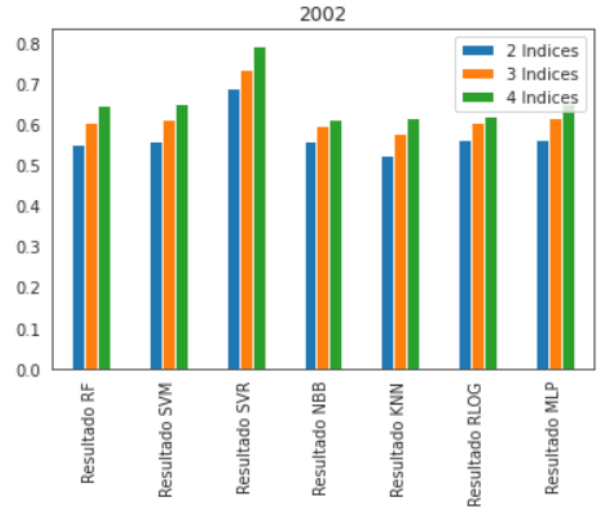
Em complemento ao Capítulo 5, seção 5.2.2, aqui são apresentados os desempenhos por ano de cada algoritmo de acordo com o conjunto de índices.

Figura A.1 – Desempenho dos conjuntos para o ano de 2001.



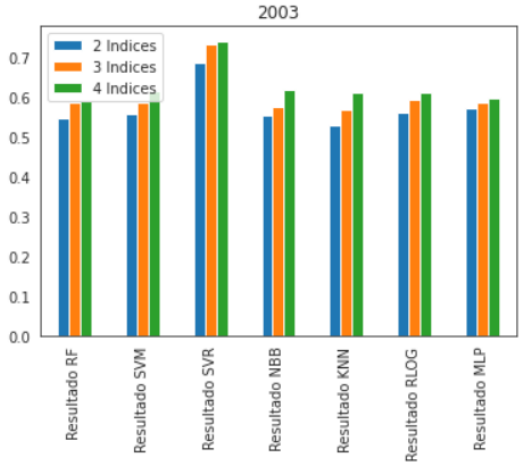
Fonte: Autores.

Figura A.2 – Desempenho dos conjuntos para o ano de 2002.



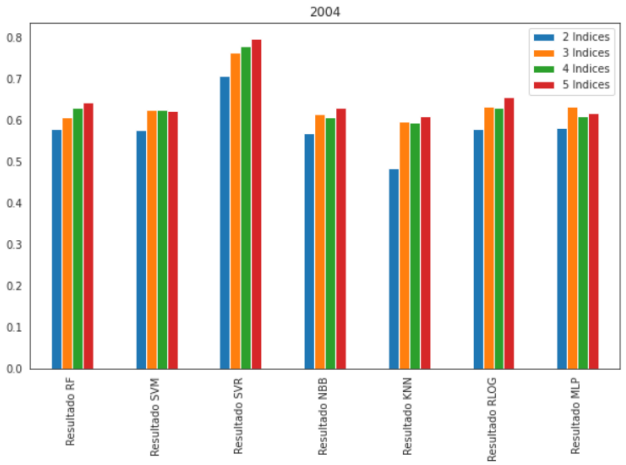
Fonte: Autores.

Figura A.3 – Desempenho dos conjuntos para o ano de 2003.



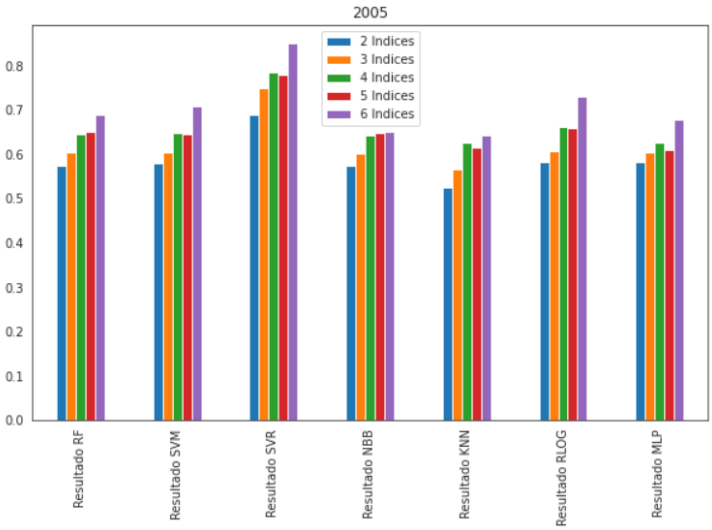
Fonte: Autores.

Figura A.4 – Desempenho dos conjuntos para o ano de 2004.



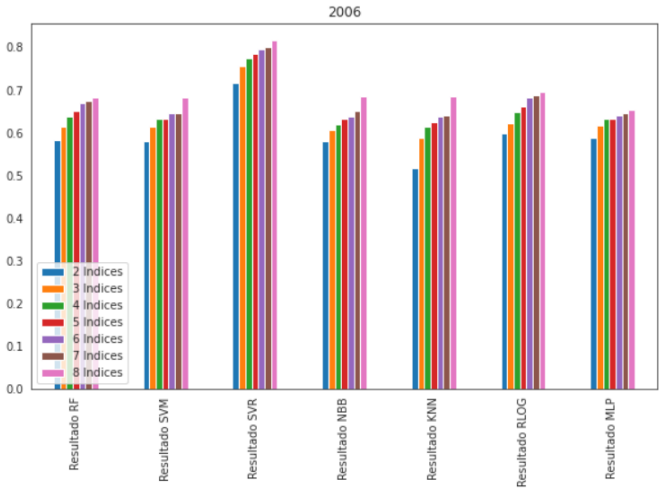
Fonte: Autores.

Figura A.5 – Desempenho dos conjuntos para o ano de 2005.



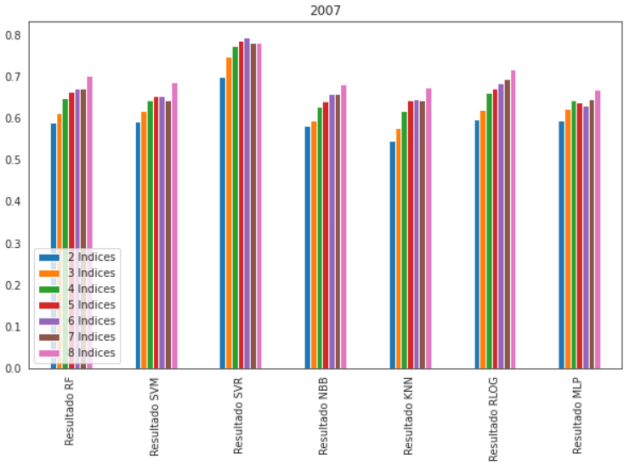
Fonte: Autores.

Figura A.6 – Desempenho dos conjuntos para o ano de 2006.



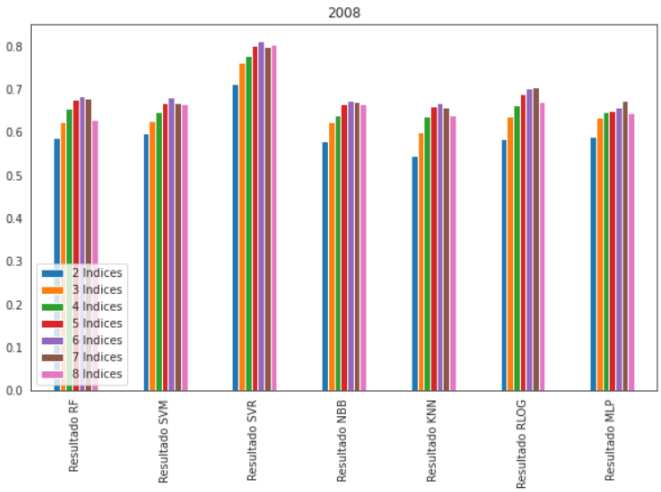
Fonte: Autores.

Figura A.7 – Desempenho dos conjuntos para o ano de 2007.



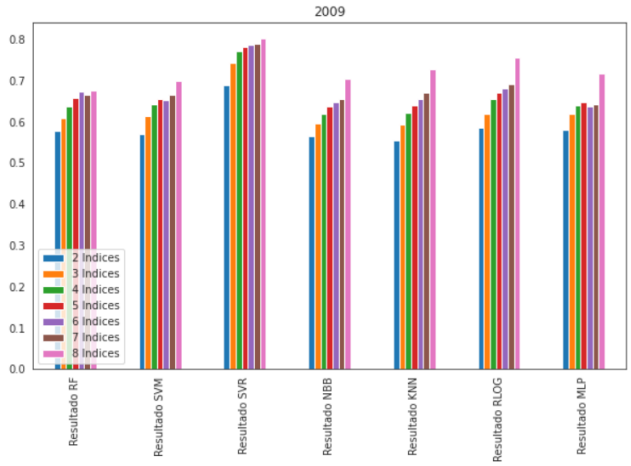
Fonte: Autores.

Figura A.8 – Desempenho dos conjuntos para o ano de 2008.



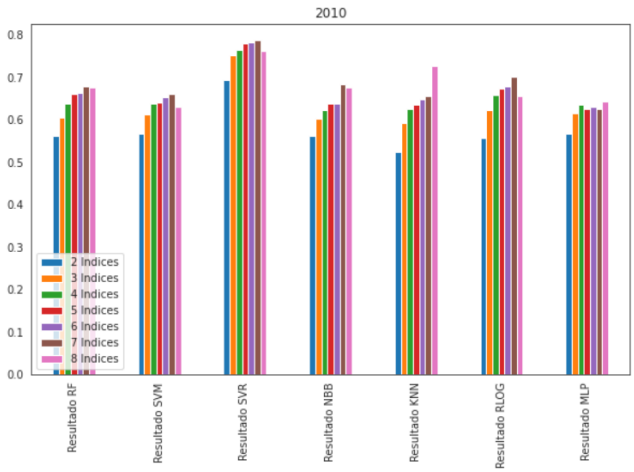
Fonte: Autores.

Figura A.9 – Desempenho dos conjuntos para o ano de 2009.



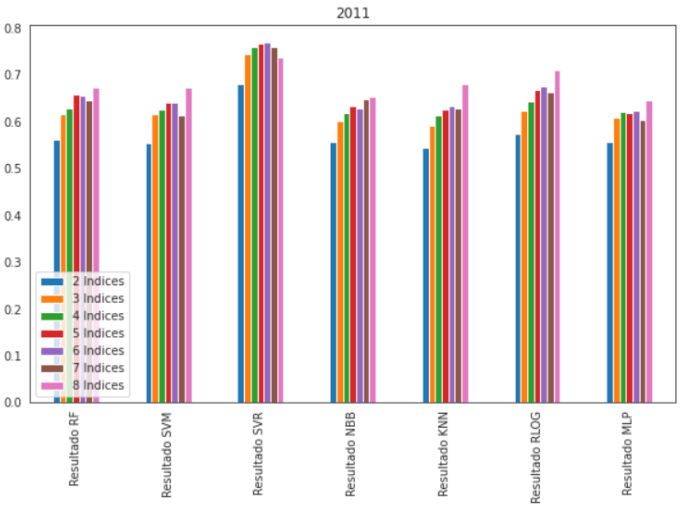
Fonte: Autores.

Figura A.10 – Desempenho dos conjuntos para o ano de 2010.



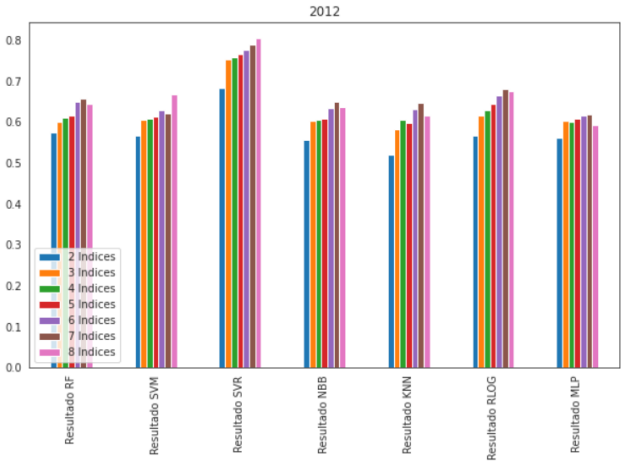
Fonte: Autores.

Figura A.11 – Desempenho dos conjuntos para o ano de 2011.



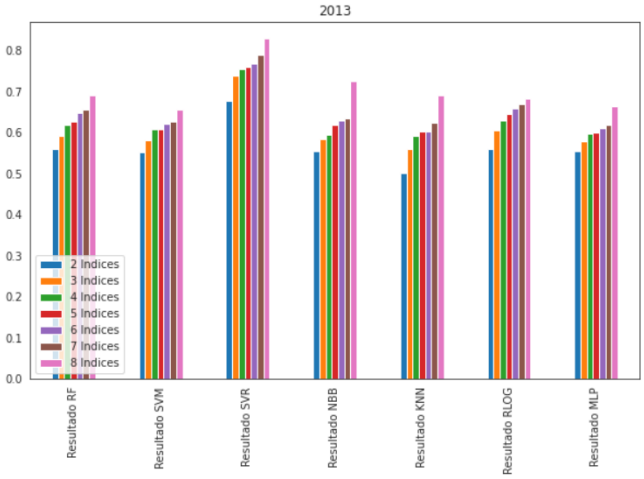
Fonte: Autores.

Figura A.12 – Desempenho dos conjuntos para o ano de 2012.



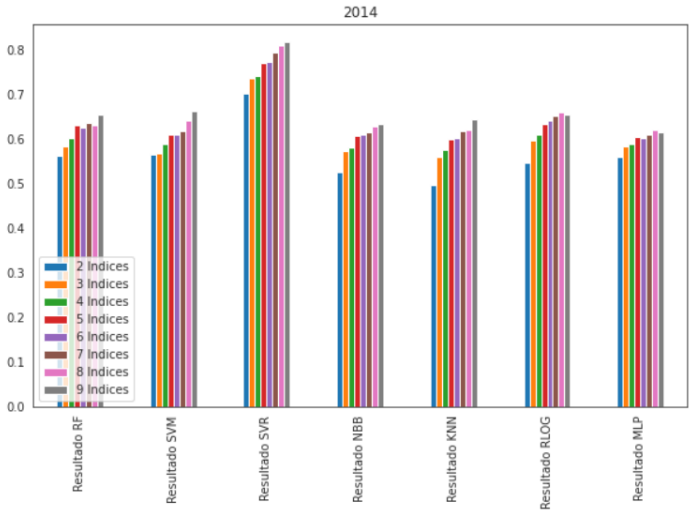
Fonte: Autores.

Figura A.13 – Desempenho dos conjuntos para o ano de 2013.



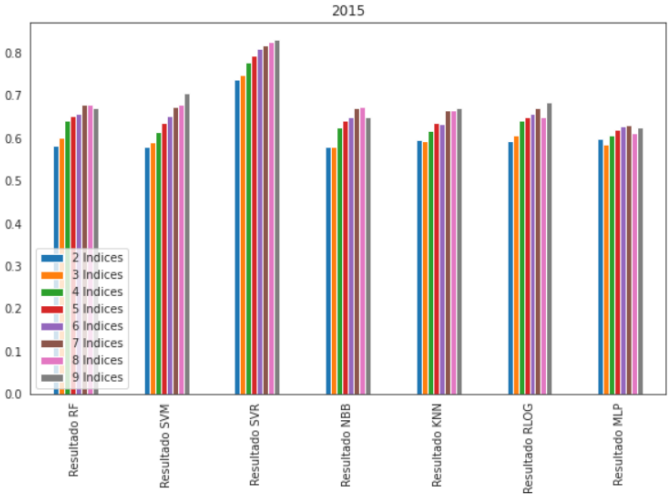
Fonte: Autores.

Figura A.14 – Desempenho dos conjuntos para o ano de 2014.



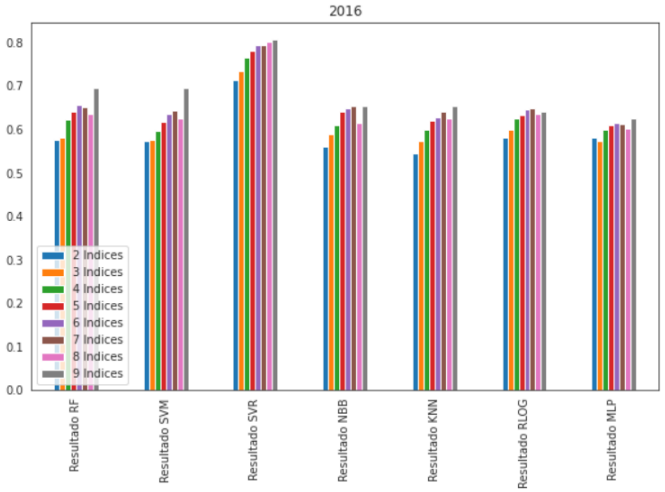
Fonte: Autores.

Figura A.15 – Desempenho dos conjuntos para o ano de 2015.



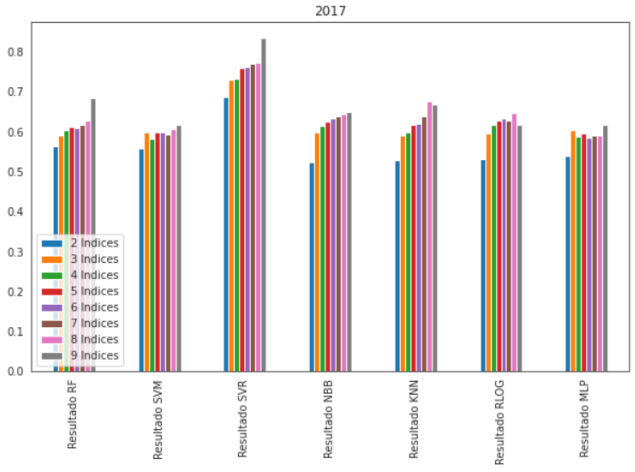
Fonte: Autores.

Figura A.16 – Desempenho dos conjuntos para o ano de 2016.



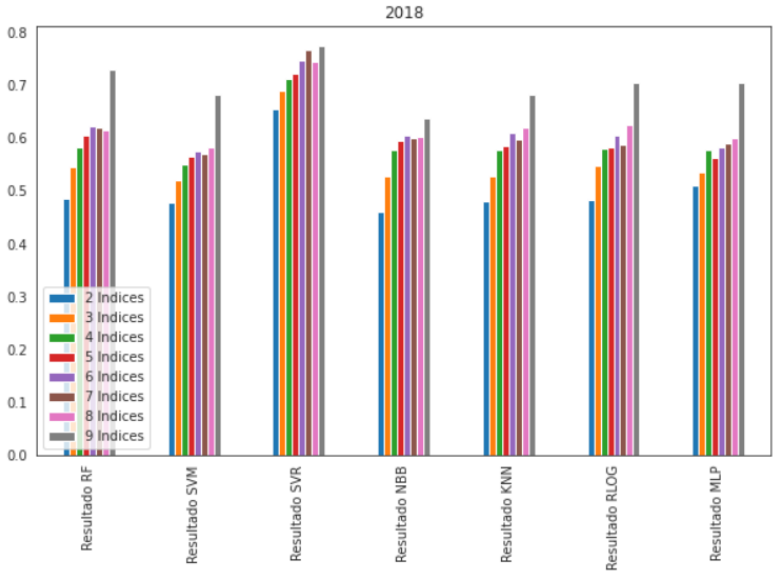
Fonte: Autores.

Figura A.17 – Desempenho dos conjuntos para o ano de 2017.



Fonte: Autores.

Figura A.18 – Desempenho dos conjuntos para o ano de 2018.



Fonte: Autores.