



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

*Índice NASDAQ 100: O impacto de diferentes tipos de indicadores
na classificação da tendência*

Diogo António Alexandre Cerqueira

Mestrado em Ciência de Dados

Orientadora:

Dra. Anabela Ribeiro Dias da Costa, Professora Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Co-Orientadora:

Dra. Diana Elisabeta Aldea Mendes, Professora Associada,
ISCTE - Instituto Universitário de Lisboa

Dezembro, 2024

*Índice NASDAQ 100: O impacto de diferentes tipos de indicadores
na classificação da tendência*

Diogo António Alexandre Cerqueira

Mestrado em Ciência de Dados

Departamento de Métodos Quantitativos para Gestão e Economia
e Departamento de Ciências e Tecnologias da Informação

Orientadora:

Dra. Anabela Ribeiro Dias da Costa, Professora Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Co-Orientadora:

Dra. Diana Elisabeta Aldea Mendes, Professora Associada,
ISCTE - Instituto Universitário de Lisboa

Dezembro, 2024

Agradecimento

Primeiramente, gostaria de agradecer à minha família por me ter proporcionado sempre todas as condições para estudar, por estarem ao meu lado nos bons e maus momentos e por apoiarem incondicionalmente todas as minhas decisões. Em segundo lugar, um agradecimento especial às minhas orientadoras pelo apoio e ajuda incansável ao longo desta caminhada. Gostaria ainda de agradecer aos meus amigos, sem eles teria sido muito mais difícil. Por último, mas não menos importante, um agradecimento à minha namorada, Filipa, por todo o apoio e ajuda nas horas mais difíceis.

Resumo

A previsão dos mercados financeiros é uma tarefa complexa, contudo, tentadora, pois uma previsão certa pode-se traduzir em retornos significativos. A maioria dos investidores de retalho perde dinheiro no mercado de ações, visto que estes são influenciados por vários fatores externos como a oferta e a procura, o sentimento dos investidores, o estado atual da economia, medidas governamentais, entres outros. Assim sendo, o principal objetivo desta dissertação é perceber o impacto da junção de diferentes tipos de indicadores, indicadores técnicos, fundamentais e de sentimento, na classificação da tendência do índice *NASDAQ 100*. Para tal, foi desenvolvido um modelo de *machine learning*, utilizando o algoritmo *eXtreme Gradient Boosting (XGBoost)*, que provou ser um algoritmo eficiente com as melhores métricas (*accuracy*, *precision*, *recall* e *f1-score*), a rondarem os 92%, aquando da junção das três categorias de indicadores, tendo-se recorrido ao método de *randomized search* para alavancar o processo de otimização dos hiper-parâmetros. Apesar da dificuldade de conciliar os indicadores fundamentais, com os indicadores técnicos e de sentimento, devido à sua periodicidade ser trimestral e os outros diária, estes contribuíram para uma ligeira melhoria na *performance* do modelo conjugados com as outras duas categorias de indicadores.

Abstract

Forecasting financial markets is a complex task, as accurate predictions can lead to significant returns. Most retail investors lose money in the stock market, because markets are influenced by various external factors such as supply and demand, investor sentiment, the current state of the economy, and government measures, among others. Therefore, the primary objective of this dissertation is to understand the impact of combining different types of indicators—technical, fundamental, and sentiment indicators—on the classification of the NASDAQ 100 index trend. To achieve this, a machine learning model was developed using the eXtreme Gradient Boosting (XGBoost) algorithm, which proved to be an efficient algorithm with the best metrics (accuracy, precision, recall and f1-score), reaching around 92% when combining the three categories of indicators. The optimization of hyperparameters was leveraged using the randomized search method. Despite the challenge of reconciling fundamental indicators with technical and sentiment indicators, due to their quarterly and daily periodicity respectively, the performance of the model slightly improved when these indicators were combined with the other two categories.

Índice

Agradecimento	iii
Resumo	v
Abstract	vii
 Capítulo 1. Introdução	 1
 Capítulo 2. Revisão da Literatura	 3
2.1. Introdução	3
2.2. Indicadores	4
2.3. Modelos de <i>Machine Learning</i>	6
2.4. Metodologia	7
2.5. Resultados	8
2.5.1. Informação Bibliográfica	8
2.5.2. Indicadores Utilizados	11
2.5.3. Modelos de <i>Machine Learning</i> Utilizados	14
2.6. Conclusões Finais	15
 Capítulo 3. Metodologia	 17
3.1. Contextualização <i>NASDAQ 100</i>	18
3.2. Variáveis recolhidas	23
3.3. Seleção de variáveis	27
3.4. Modelação e Resultados	29
3.4.1. Iteração 1: <i>Baseline</i>	31
3.4.2. Iteração 2: Indicadores Técnicos	31
3.4.3. Iteração 3: Indicadores fundamentais	32
3.4.4. Iteração 4: Indicadores de sentimento	32
3.4.5. Iteração 5: Indicadores técnicos e fundamentais	33
3.4.6. Iteração 6: Indicadores técnicos e de sentimento	34
3.4.7. Iteração 7: Indicadores fundamentais e de sentimento	34
3.4.8. Iteração 8: Indicadores técnicos, fundamentais e de sentimento	35

Capítulo 4. Conclusões	39
Referências Bibliográficas	43
Anexo A	47
Anexo B	49
Anexo C	50
Anexo D	52
Anexo E	54
Anexo F	57

Índice de Figuras

Figura 1 - Número de artigos selecionados em cada fase por base de dados.	8
Figura 2 - Número de artigos por ano de publicação.	9
Figura 3 - Número de artigos por ano de publicação.	9
Figura 4 - Frequência de países por mercado.	10
Figura 5 - Frequência de artigos por índices.	10
Figura 6 - Tipos de indicadores utilizados.	11
Figura 7 - Indicadores técnicos utilizados.	12
Figura 8 - Frequência dos indicadores fundamentais utilizados.	12
Figura 9 - Frequência dos indicadores de sentimento utilizados.	13
Figura 10 - Frequência dos modelos preditivos utilizados.	14
Figura 11 - Frequência dos modelos utilizados.	14
Figura 12 - Metodologia adotada para o desenvolvimento da tese.	17
Figura 13 - Evolução do preço de fecho do índice <i>NASDAQ 100</i> .	20
Figura 14 - Retorno anual do índice <i>NASDAQ 100</i> .	20
Figura 15 - Peso dos setores presentes no índice <i>NASDAQ 100</i> .	21
Figura 16 - Top 13 empresas com maior peso no índice <i>NASDAQ 100</i> .	22
Figura 17 - Visualização do <i>Relative Strength Index</i> (RSI).	23
Figura 18 - Visualização do efeito do RSI no preço de fecho do índice <i>NASDAQ 100</i> .	25
Figura 19 - Visualização do <i>Economic Policy Uncertainty Index for United States</i> .	27
Figura 20 - Variáveis mais explicativas para as previsões do modelo (<i>shap values</i>).	38

Índice de Tabelas

Tabela 1 - Filtros para a seleção dos artigos.	7
Tabela 2 - Peso dos países representados no índice <i>NASDAQ 100</i> .	21
Tabela 3 - Variáveis selecionadas.	28
Tabela 4 - Definição do espaço dos hiper-parâmetros.	30
Tabela 5 - Iteração 1: <i>Baseline</i> .	31
Tabela 6 - Iteração 2: Indicadores técnicos.	31
Tabela 7 - Iteração 3: Indicadores fundamentais.	32
Tabela 8 - Iteração 4: Indicadores de sentimento.	32
Tabela 9 - Iteração 5: Indicadores técnicos e fundamentais.	33
Tabela 10 – Iteração 6: Indicadores técnicos e de sentimento.	34
Tabela 11 - Iteração 7: Indicadores fundamentais e de sentimento.	34
Tabela 12 - Iteração 8: Indicadores técnicos, fundamentais e de sentimento.	35
Tabela 13 - Resultados.	36

Introdução

No decorrer dos últimos anos, o campo da previsão de mercados financeiros tem sido alvo de uma evolução contínua, impulsionada em grande parte pelos avanços da Inteligência Artificial, mais concretamente pelo desenvolvimento de diversas técnicas de *machine learning* e pela disponibilidade de grandes volumes de dados financeiros. Entre os maiores mercados financeiros do mundo encontra-se, o índice *NASDAQ 100*, que representa 100 das maiores empresas não financeiras listada na bolsa de valores norte americana. É composto por uma ampla gama de empresas de diferentes setores, como o tecnológico, o da saúde, industrial, das telecomunicações, entre outros. Devido à sua concentração em tecnologia e inovação, o *NASDAQ 100* é frequentemente visto como um barômetro do sentimento e do desempenho do setor tecnológico, uma vez que empresas como a *Apple*, *Microsoft*, *Amazon*, *Alphabet (Google)* fazem parte deste índice.

Devido ao tamanho e à importância das empresas que o compõem, mudanças no índice podem afetar outros mercados, tanto nos Estados Unidos como noutros mercados internacionais. Assim sendo, para os investidores, o *NASDAQ 100* serve como um *benchmark* crucial para avaliar o desempenho de portfólios concentrados em tecnologia e inovação. Para investidores que procuram uma diversificação do seu portfólio, o *NASDAQ 100* oferece exposição a uma ampla gama de setores de alto crescimento. Embora seja demasiado ponderado em relação a empresas tecnológicas, o índice também inclui empresas líderes em outros setores, permitindo aos investidores beneficiar de diferentes dinâmicas do mercado.

A previsão do mercado de ações é uma área de grande interesse tanto a nível académico quanto a nível profissional, dada a sua complexidade e o impacto significativo que pode ter na tomada de decisões de investimento. Desta forma, vários métodos têm sido desenvolvidos e aplicados na tentativa de prever os movimentos do mercado de ações, cada um com as suas vantagens e desvantagens. Os mercados financeiros são influenciados por múltiplos fatores, como eventos económicos, políticos e desastres naturais, muitos dos quais são imprevisíveis. Essa incerteza inerente torna a previsão do mercado de ações particularmente desafiadora.

Uma boa previsão dos mercados financeiros pode trazer uma série de vantagens significativas para os investidores, tanto ao nível da tomada de decisão como na gestão de riscos e retornos, contudo, a complexidade e a natureza dinâmica dos mercados financeiros exigem uma abordagem cautelosa e bem informada para a utilização de previsões na estratégia de investimento.

A presente dissertação visa explorar a eficácia da combinação simultânea de três tipos de indicadores (técnicos, fundamentais e de sentimento) na previsão do índice *NASDAQ 100*. Esta abordagem multidimensional é fundamentada na hipótese de que a integração de diferentes categorias de indicadores pode proporcionar uma visão mais holística e precisa dos movimentos futuros do mercado, superando as limitações das abordagens que se concentram num único tipo de indicador.

Assim sendo, o primeiro objetivo é avaliar a eficácia individual de cada tipo de indicador na previsão do índice *NASDAQ 100*. Isso envolve a análise de como cada tipo de indicador, isoladamente, pode influenciar ou prever os movimentos do índice. Esta etapa é crucial para estabelecer uma *baseline* para compreender o valor agregado da combinação desses indicadores. O segundo objetivo é explorar como a combinação desses três tipos de indicadores pode proporcionar uma previsão mais precisa do índice *NASDAQ 100* do que o uso de qualquer tipo de indicador isoladamente. Isso implica investigar a sinergia entre os indicadores técnicos, que se baseiam em padrões de preços e volumes históricos; os indicadores fundamentais, que analisam a situação económica e financeira das empresas e fatores macroeconómicos; e os indicadores de sentimento, que capturam as emoções e percepções dos agentes do mercado.

Neste contexto, esta dissertação propõe a utilização do algoritmo *XGBoost*, uma técnica de *machine learning* que tem demonstrado uma eficácia notável em várias tarefas de previsão, para modelar e prever a direção do índice *NASDAQ 100*.

O *XGBoost* (*eXtreme Gradient Boosting*), tem sido amplamente reconhecido pela sua capacidade de lidar com diferentes tipos de dados estruturados e a sua robustez em capturar interações complexas entre variáveis. Neste contexto, o *XGBoost* será utilizado para integrar e analisar três categorias distintas de indicadores: técnicos, fundamentais e de sentimento.

A presente dissertação encontra-se estruturada em quatro capítulos, sendo este a introdução. O seguinte capítulo, é constituído pela revisão de literatura, abrangendo os diferentes tipos de indicadores que foram utilizados perante os artigos selecionados, os algoritmos mais utilizados para a classificação da tendência e previsão dos preços de diferentes mercados de ações. O terceiro capítulo está relacionado fundamentalmente sobre a metodologia adotada, CRISP-DM. Para finalizar esta dissertação, encontram-se as discussões dos resultados, as principais conclusões obtidas neste estudo e a definição de eventuais próximos passos.

Revisão da Literatura

2.1. Introdução

Os mercados financeiros são um tema central no campo da economia, onde são movidos milhares de milhões de dólares diariamente (*Hoseinzade & Haratizadeh, 2019*). Estes têm, um papel importante para os investidores (*Chen et. al., 2023*), logo a capacidade de antecipar movimentos do mercado não apenas oferece vantagens económicas significativas (*Wen et al., 2019*), mas também contribui para a compreensão mais profunda dos mecanismos que regem as dinâmicas financeiras globais, diminuindo o risco dos investidores. A previsão dos mercados financeiros é uma tarefa desafiante (*Zhao et al., 2023*), devido à complexidade dos mesmos e à sua alta volatilidade (*An et. al., 2024*). Historicamente, a previsão dos mercados financeiros tem oscilado entre a procura por padrões nos dados históricos e a aplicação de teorias económicas e financeiras. Com o advento da tecnologia da informação e o aumento da disponibilidade de grandes conjuntos de dados, novas abordagens baseadas em métodos quantitativos e algoritmos de *machine learning* têm ganho destaque. Estas técnicas modernas oferecem novas perspetivas e possibilidades, desafiando as abordagens tradicionais e contribuindo para um campo em constante evolução. A capacidade de antecipar movimentos do mercado pode ser crucial para a tomada de decisões de investimento e para a estratégia de gestão do portfólio. Neste contexto, os indicadores desempenham um papel fundamental, ajudando a decifrar padrões complexos e a prever tendências futuras.

A escolha entre indicadores técnicos, fundamentais ou de sentimento muitas vezes reflete a filosofia de investimento e os objetivos do analista. Indicadores técnicos providenciam informações incrementais, dando mais valor às análises de mercados financeiros (*Mamaysky et al., 2000*), outros inclinam-se para a análise qualitativa e orientada para o valor dos indicadores fundamentais (*Barak et al., 2017*). Segundo *Murphy (1999)* a abordagem fundamental tem como objetivo perceber os fatores externos ao preço das ações, de modo a entenderem o valor intrínseco do mercado, e quais são as causas das oscilações dos mesmos. Paralelamente a estas abordagens, os indicadores de sentimento emergem como uma ferramenta vital (*Wang et al., 2019*), oferecendo uma perspetiva única que transcende os dados puramente quantitativos ou fundamentais. Estes indicadores tentam capturar o sentimento, as expectativas e as atitudes dos investidores em relação ao mercado, refletindo como estes sentimentos coletivos podem influenciar o comportamento do mesmo. Contrariamente aos indicadores técnicos e fundamentais, que se concentram em dados históricos e indicadores

económicos, os indicadores de sentimento são mais dinâmicos e podem fornecer informações em tempo real sobre as tendências do mercado.

2.2. Indicadores

Neste contexto, a Hipótese do Mercado Eficiente (HME) emerge como um conceito fundamental, desafiando a premissa de que a análise de mercado, seja técnica ou fundamental, pode gerar retornos acima da média de forma consistente. Segundo *Fama (1970)*, a HME postula que os preços das ações refletem todas as informações disponíveis e, portanto, negociam no seu valor justo, tornando difícil para os investidores superarem consistentemente o mercado. A relevância da HME na previsão dos mercados financeiros é um tópico de debate contínuo, especialmente considerando a natureza volátil e inovadora das empresas que compõem o índice *NASDAQ*, o que pode, por vezes, desafiar as premissas tradicionais da eficiência do mercado.

Na literatura, são frequentemente utilizados indicadores técnicos na análise e previsão dos mercados financeiros. Indicadores técnicos são ferramentas analíticas utilizadas para prever as tendências futuras do mercado com base em dados históricos de preços e volumes de transações. Um dos indicadores técnicos mais utilizados na literatura é o *Relative Strength Index (RSI)*, que é um oscilador do *momentum* que mede a velocidade e a mudança dos movimentos dos preços. O RSI é frequentemente usado para identificar condições de sobrecompra (quando o RSI está acima de 70), sugerindo que o preço da ação está sobrevalorizado, podendo ocorrer uma correção (queda) do mesmo; ou sobrevenda, quando o preço da ação está subvalorizado (valor do RSI abaixo de 30), indicando que a ação está prestes a subir de preço. Deste modo, o RSI fornece sinais de potenciais reversões de tendência dos mercados (*González et al., 2011*). As médias móveis também costumam ser amplamente utilizadas na literatura como indicadores técnicos, destaque para a *Moving Average Convergence/Divergence (MACD)* que é um indicador que demonstra a relação entre duas médias móveis, funcionando como um indicador de *momentum*, sendo utilizada para indicar se o investidor deve ou não comprar uma ação num determinado período temporal (*Nabipour et al., 2020*).

Enquanto os indicadores técnicos concentram-se em padrões de preços e volumes de negociação, os indicadores fundamentais mergulham profundamente nos aspetos económicos e financeiros subjacentes a cada empresa e do mercado em geral, estando relacionados com as demonstrações financeiras das empresas e indicadores económicos/macroeconómicos (*Beniwal et al., 2024*). A análise fundamental é mais utilizada para horizontes temporais médios ou longos, uma vez que as empresas só costumam apresentar este tipo de informações com alguma distância temporal, normalmente todos os trimestres (*Chou, S.J. et al., 2024*). Um dos indicadores fundamentais mais conhecidos, está

relacionado com as demonstrações financeiras das empresas que é o *Return Of Equity* (ROE), que mede o desempenho da empresa quando esta gera lucros através dos capitais próprios. Esta métrica é utilizada para selecionar as empresas que mais maximizam o investimento feito nelas próprias, isto é, por cada dólar investido, as empresas são capazes de gerar um lucro líquido maior em termos de percentagem do capital investido (Silva *et al.*, 2015). Por outro lado, o mercado de ações é muito sensível a fatores externos, segundo Niu *et al.* (2023), que demonstraram como os riscos geopolíticos influenciam a volatilidade dos mercados. Vários estudos, foram e continuam a ser realizados para perceber os impactos inerentes aos mercados financeiros. Ariel (1990) realizou um estudo sobre como os dias que antecederem os feriados afetam os retornos do índice *Dow Jones Industrial*. Este estudo revelou que o retorno médio das ações no dia anterior aos feriados é de 9 a 14 vezes superior em comparação com os dias não pré-feriados. Este padrão sugere que o comportamento dos investidores pode ser influenciado por fatores psicológicos e sociais, além dos puramente financeiros.

Indicadores de sentimento são ferramentas analíticas usadas para avaliar as atitudes e expectativas dos investidores em relação ao mercado, sendo cruciais para entender o comportamento do mercado. Este tipo de indicadores ajuda a identificar sentimentos positivos ou negativos, que podem sinalizar reversões de tendência ou continuação do movimento do mercado, que não totalmente justificados pelas notícias fundamentais (Shleifer & Summers, 1990). Na literatura existem diversas abordagens ao tema do sentimento, nos mercados financeiros. Alguns autores tentam-se focar em sentimentos das pessoas de uma forma mais analítica, isto é, através da análise de sentimento em plataformas como o Twitter ou em notícias *online* (Nti *et al.*, 2021). Outros autores, seguem uma abordagem mais tradicional ao abordar este conceito de sentimento, utilizando indicadores como por exemplo o CBOE Volatility Index (VIX), ou o CBOE NASDAQ Volatility Index (VXN), que são indicadores comumente utilizados na literatura com o objetivo de identificar o sentimento em torno dos mercados financeiros (Campisi *et al.*, 2023). Hoseinzade & Haratizadeh (2019) utilizaram outros índices, commodities, pares de moedas e taxas de juro dos títulos do tesouro, de modo a conseguirem sinalizar o sentimento geral externo aos mercados de ações. Segundo Zhang *et al.* (2023), o sentimento dos investidores pode melhorar a previsão do retorno das ações. Contudo, Gunnarsson *et al.* (2024) fizeram um estudo onde combinaram dados relacionados com o sentimento dos investidores, variáveis macroeconómicas e dados financeiros de modo a melhorar a performance do modelo e concluíram que só havia melhorias na performance para espaços temporais curtos. Por outro lado, Li *et al.* (2024), melhoraram a performance dos seus modelos quando a previsão era feita num espaço temporal mais longo (1 semana ou 1 mês), sendo sinal de que este tema dos indicadores de sentimento ainda não é algo consensual dentro da literatura, o que também pode ter resultados diferentes tendo em conta a abordagem seguida, se de uma forma mais tradicional utilizando, por

exemplo, indicadores como o VIX, VXN ou outros índices, ou uma abordagem mais analítica passando pela análise de sentimento de tweets, notícias, entre outros.

2.3. Modelos de *Machine Learning*

Nos últimos anos, a análise preditiva tem experimentado avanços significativos, impulsionados em grande parte pelo desenvolvimento de algoritmos de *machine learning* e *deep learning* sofisticados e pela disponibilidade crescente de grandes volumes de dados.

Patel et al. (2015) utilizaram quatro modelos preditivos para a previsão de dois índices, o *NIFTY50* e o *S&P Bombay Stock Exchange* ambos do mercado de ações indiano, que consistiu no desenvolvimento de algoritmos de *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), *Random Forest* e *Naive Bayes*. O modelo com pior performance foi o que utilizou a ANN. Sendo que os restantes três algoritmos foram muito semelhantes com métricas de *accuracy* e *F1-Score* a rondar os 89%.

Ghosh et al. (2022), desenvolveram um modelo de previsão intradiária do índice *S&P 500*, onde o *output* do modelo era a probabilidade de as ações subirem. Com base nestas probabilidades, compravam as 10 ações com maior probabilidade de subirem e *short* nas 10 ações com a menor probabilidade de subirem. Utilizaram *Long Short-Term Memory* (LSTM) e *Random Forests*, onde a abordagem LSTM foi ligeiramente melhor que a alternativa, dando retornos diários médios de 0.64% e 0.54%, respetivamente.

Naik e Mohan (2021), por outro lado tentaram prever crises no mercado de ações mercado, utilizando dois modelos preditivos diferentes, *XGBoost* e *Deep Neural Networks* (DNN), tendo obtido os melhores resultados com o algoritmo *XGBoost*.

Deng et al. (2021) utilizaram o algoritmo *XGBoost*, *SVM* e *Random Forest* para a previsão das ações da *Apple*, tendo obtido os melhores resultados com modelos *XGBoost*, rondando os 10% de retorno anual. Também *Han et al.* (2023), através do algoritmo *XGBoost*, geraram sinais de *trading*, isto é, se o momento é de compra ou venda, para a previsão de 92 ações listadas no índice *NASDAQ*.

Zhou et al. (2019) recorreram ao uso de redes neuronais para a previsão da tendência dos índices *Shanghai Stock Exchange Composite*, *NASDAQ* e *S&P 500*. Estes conseguiram os melhores resultados para a previsão do preço do *S&P 500*, com valores de *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE) e *Mean Absolute Percentage Error* (MAPE), de cerca de 15, 20 e 0.012, respetivamente.

Um dos algoritmos mais utilizados e com melhores resultados destes artigos mencionados foi o *XGBoost*. Destes estudos, alguns retratavam classificação de tendência e outros a previsão dos preços dos mercados.

2.4. Metodologia

Para o desenvolvimento da presente revisão bibliográfica foram definidas um conjunto de regras iniciais que os artigos têm de seguir, de modo a serem considerados válidos para a sua utilização. O primeiro passo (Tabela 1) passou por selecionar artigos relacionados com o tema desta dissertação e para isso foi definida uma *query* que contivessem algumas palavras-chave que estivessem relacionadas com o tema. Após essa pesquisa inicial, os artigos foram filtrados pelo seu ano, o objetivo passava por selecionar apenas os artigos mais recentes e relevantes, em alguns casos os artigos utilizados eram mais antigos, pois eram artigos que eram encontrados nos artigos selecionados, que se mostravam relevantes para o tema. O tipo de documento tinha de ser uma revista em língua inglesa ou portuguesa, sendo que o artigo tinha de estar já finalizado. Para esta pesquisa, foram utilizadas três das bases de dados científicas mais relevantes na academia como é o caso *Scopus*, *IEEE Xplore* e *Science Direct*.

Tabela 1 - Filtros para a seleção dos artigos

Filtros	Especificações
Ano de Publicação	2000 – 2024
Tipo de Documento	Revista
Língua	Inglês ou português
Estado de Publicação	Final
Bases de Dados	<i>Scopus</i> , <i>IEEE Xplore</i> e <i>Science Direct</i>
Palavras-chave	((“Stock Market Prediction” OR “Stock Market Forecasting”) AND (“AI” OR “Machine Learning”))

Após esta primeira seleção de artigos, recorreu-se à utilização de duas plataformas, o *Journal Citation Reports (JCR)* e o *Scimago Journal Rank (SJR)*, de modo a conseguir selecionar as revistas mais relevantes. Para esta seleção, contribuíram a utilização de três métricas chave neste processo, na plataforma *JCR*, a métrica utilizada foi o *Journal Impact Factor (JIF)*, que mede o número médio de citações de artigos de uma determinada revista sobre um período específico. Revistas que obtenham um fator de impacto maior, tendem a ser mais relevantes e consequentemente terem mais prestígio entre a academia. A outra métrica é o *H-Index*, que é calculada com base nas citações, avaliando o prestígio do investigador. A última métrica e que serviu como base para a seleção das revistas mais prestigiadas, foi o quartil das revistas, sendo que, uma revista no quartil 1, está melhor classificada que revistas no quartil 2,3 e 4. De referir que, só foram selecionadas revistas que pertencessem ao primeiro quartil.

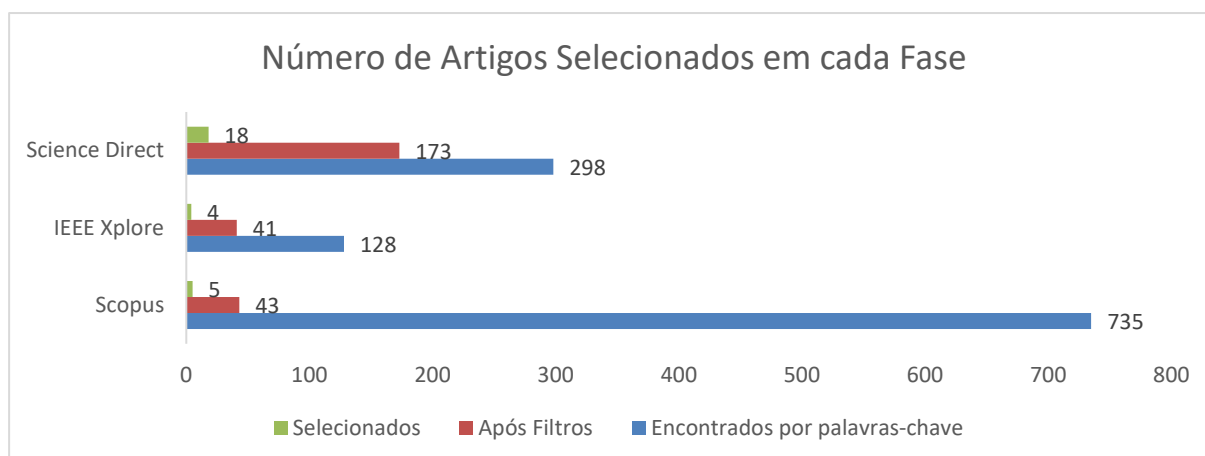


Figura 1 - Número de artigos selecionados em cada fase por base de dados

Assim sendo, como se pode ver na Figura 1 estes são os artigos que foram selecionados em cada uma destas fases. De salientar, que também foram selecionados artigos que estivessem presentes em artigos selecionados, desde que cumprissem estes requisitos de seleção das revistas.

Numa primeira fase, através da *query* referida na Tabela 1, existiam muitos artigos relacionados com a previsão de mercados financeiros, cerca de mil, nas diferentes bases de dados. Após a aplicação dos filtros este número desceu consideravelmente (cerca de 250 artigos), sendo que acabaram por ser selecionados 27 artigos (Anexo A), os mais relevantes para o tema, 18 na base dados da *Science Direct*, 5 na base de dados *Scopus* e 4 na base de dados *IEEE Xplore*.

Os objetivos desta revisão passaram por recolher informações sobre três pontos chave:

1. Informação geral da literatura selecionada;
2. Indicadores utilizados, entre cada categoria de indicadores (técnicos, fundamentais e sentimento);
3. Técnicas de *machine learning*/AI utilizadas;

2.5. Resultados

Nesta secção, são apresentados os resultados da presente revisão dos artigos selecionados, de forma a responder às três questões enumeradas anteriormente. Foram selecionados 27 artigos de 12 revistas diferentes.

2.5.1. Informação bibliográfica

Ao analisar-se a Figura 2, é possível observar a distribuição das publicações dos artigos por ano.

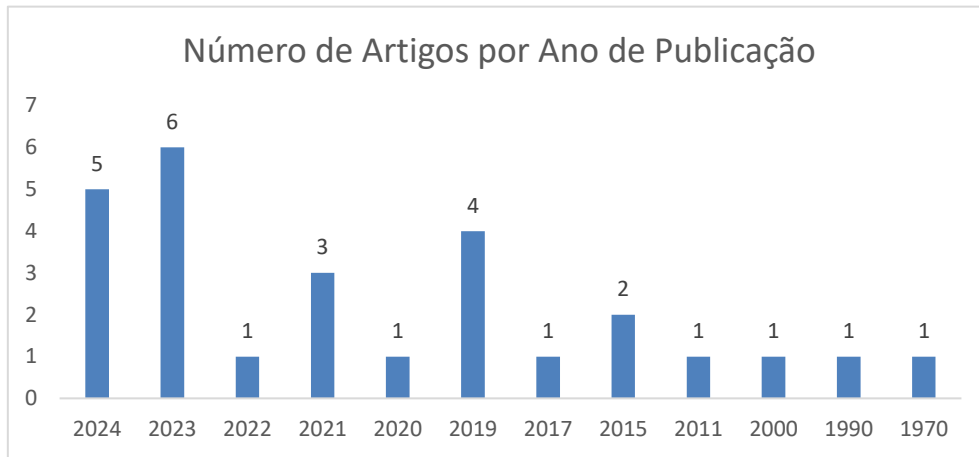


Figura 2 - Número de artigos por ano de publicação

Mais de metade dos artigos selecionados foram publicados depois do ano de 2020. Destaque para os artigos de Ariel (1990) e Fama (1970), que não cumprindo o critério dos anos de publicação, foram selecionados devido à sua importância neste campo de pesquisa.

A Figura 3, ilustra o número de artigos selecionados tendo em conta as revistas.



Figura 3 - Número de artigos selecionados por revista

Dos 27 artigos selecionados, estes foram publicados por 12 revistas diferentes, com claro destaque para a revista *Expert Systems With Applications*, que representam cerca de 30% dos artigos selecionados. Destaque também para a revista *Information Fusion* que tem um *impact factor* de 18,6 e a revista *Journal of Finance* com um *H-Index* de 352.

Por outro lado, é também importante perceber quais os mercados mais frequentes entre os artigos selecionados. Alguns estudos focaram-se na previsão de índices, como por exemplo *S&P 500*, outros focaram-se na previsão de ações individuais, nestes casos, por exemplo se o estudo tentou prever as ações da *Tesla, Inc.* estas irão ser associadas ao mercado americano e ao índice *NASDAQ*, pois é onde esta ação está representada. Assim sendo, verifica-se através da Figura 4, que os Estados Unidos da América (EUA) e a China são os países mais representativos nos estudos selecionados. A soma das frequências não irá dar 27 (igual ao número de artigos selecionados), pois, por vezes, em cada artigo os autores tentavam prever a tendência de mais do que um mercado, do mesmo país ou de países diferentes.

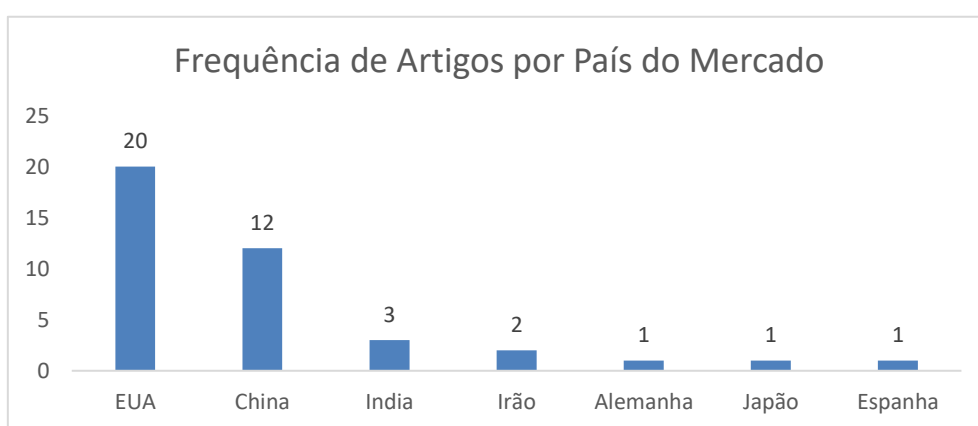


Figura 4 – Frequência países por mercado

Os índices mais frequentes nos artigos selecionados, foram o *S&P 500*, empatado com o índice *NASDAQ*, seguidos logo pelo *SSE Index*, como pode ser observado na Figura 5. A descrição completa dos índices pode ser consultada no Anexo B.

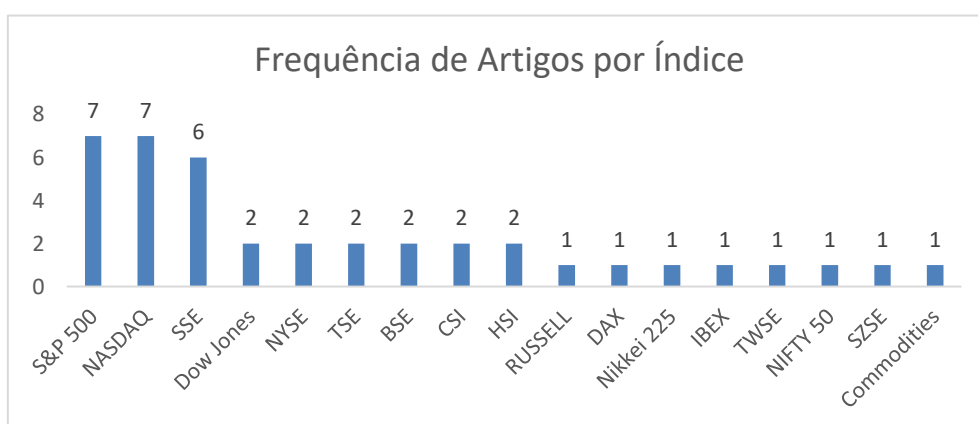


Figura 5 - Frequência de artigos por índices

2.5.2. Indicadores Utilizados

Nesta fase, serão enunciadas as abordagens que os diferentes autores abordaram para a previsão dos mercados financeiros (Figura 6), de modo a conhecer melhor quais foram os tipos de indicadores utilizados.

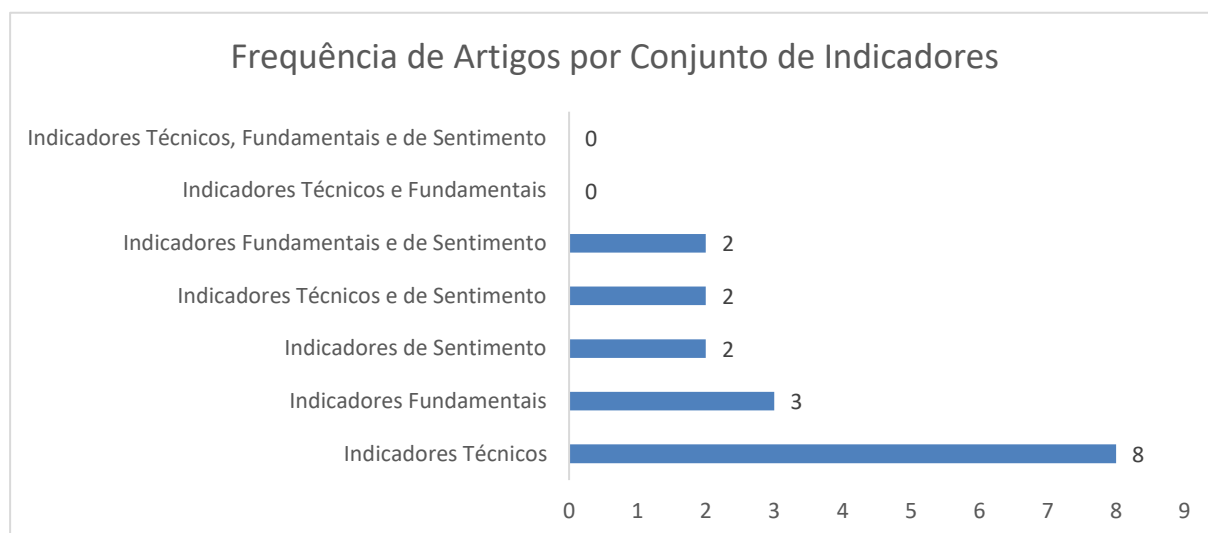


Figura 6 - Tipos de indicadores utilizados

Destaca-se a utilização de indicadores técnicos, dos 17 estudos que utilizam indicadores para previsão dos mercados, 8 estudos utilizam este tipo de indicadores, sendo o tipo de variáveis mais utilizadas, desta amostra. Indicadores fundamentais e de sentimento, também são empregues, mas de uma forma menos frequente. De salientar, o facto de 4 estudos terem adotado abordagens complementares, isto é, a junção de dois tipos de indicadores diferentes. Nota-se também que nenhum destes estudos, juntou os três tipos de indicadores.

Entre os diversos estudos que utilizam indicadores técnicos, foram encontradas 19 variáveis diferentes como pode ser observado na Figura 7 (página 12).

O *Relative Strenght Index*, mais conhecido como RSI, é o indicador mais frequente entre os artigos. Dos 10 artigos que utilizaram pelo menos indicadores técnicos, 7 utilizaram o RSI, mostrando bem como é comum à sua utilização. Outros indicadores técnicos, também foram utilizados por mais do que um autor, sendo estes uma amostra dos indicadores técnicos presentes na literatura. Indicadores como *Head and Shoulders*, *Inverted Head and Shoulders*, *Rectangle tops*, entre outros foram aplicados apenas num estudo no ano de 2000.

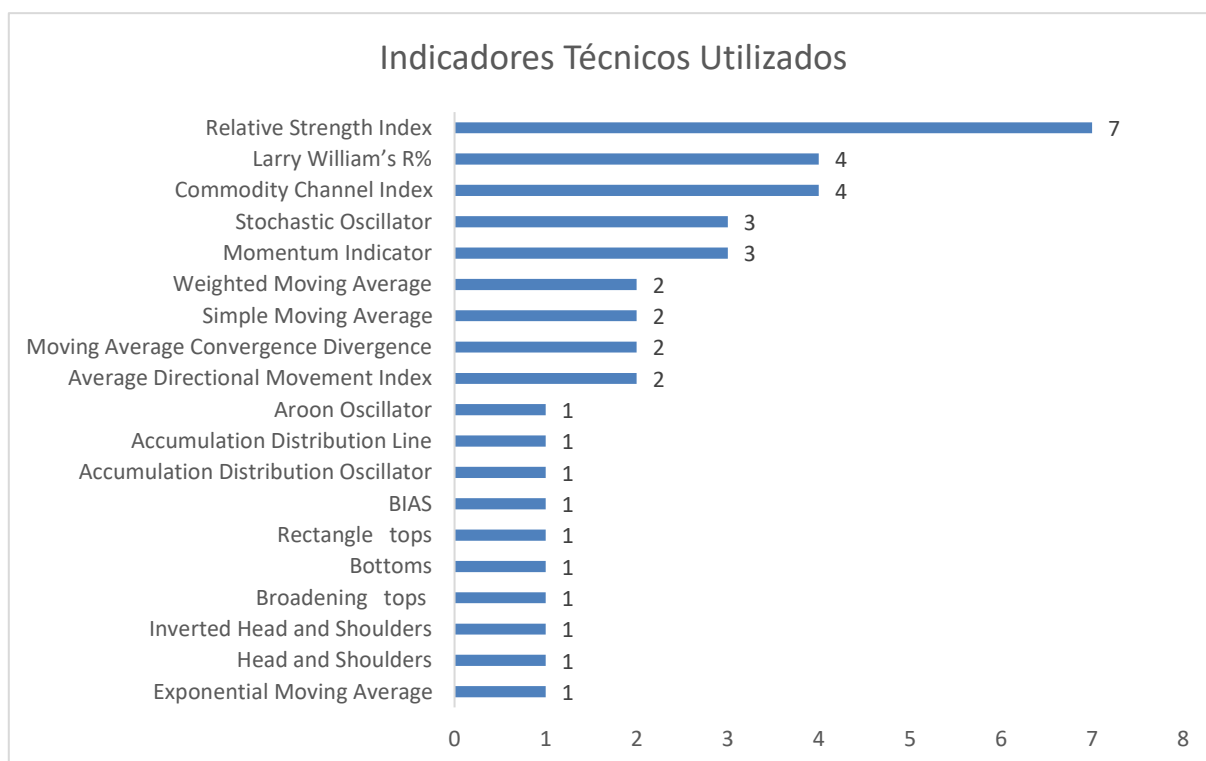


Figura 7 - Frequência dos indicadores técnicos utilizados

Quanto aos indicadores fundamentais, foram utilizados em menos artigos comparados com os indicadores técnicos. Estes baseiam-se na sua generalidade em variáveis presentes nas demonstrações financeiras das empresas, como pode ser observado na Figura 8.

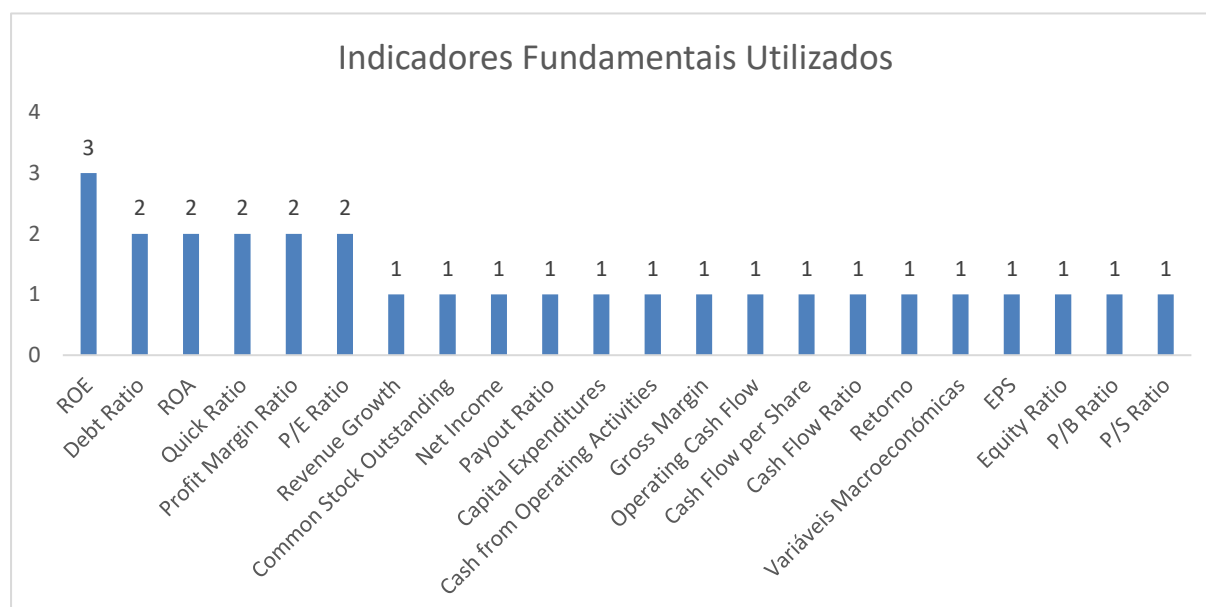


Figura 8 - Frequência dos indicadores fundamentais utilizados

Relativamente aos indicadores de sentimento, surge a necessidade de utilização do *VIX* e *VXN*, conhecidos como índices de volatilidade, ou pelos índices do medo, relativamente ao *S&P 500* e *NASDAQ*, respetivamente. Também o sentimento dos investidores nas redes sociais, ou as notícias *online* são aplicados como indicadores de sentimento. Destacam-se também outros indicadores de sentimento, como a utilização da variação de outros índices, *commodities* e taxas de títulos do tesouro. Estes últimos podem não estar diretamente relacionados com o sentimento dos investidores, como a análise de publicações em redes sociais, mas conseguem medir o sentimento de uma forma mais generalizada relativamente aos mercados.

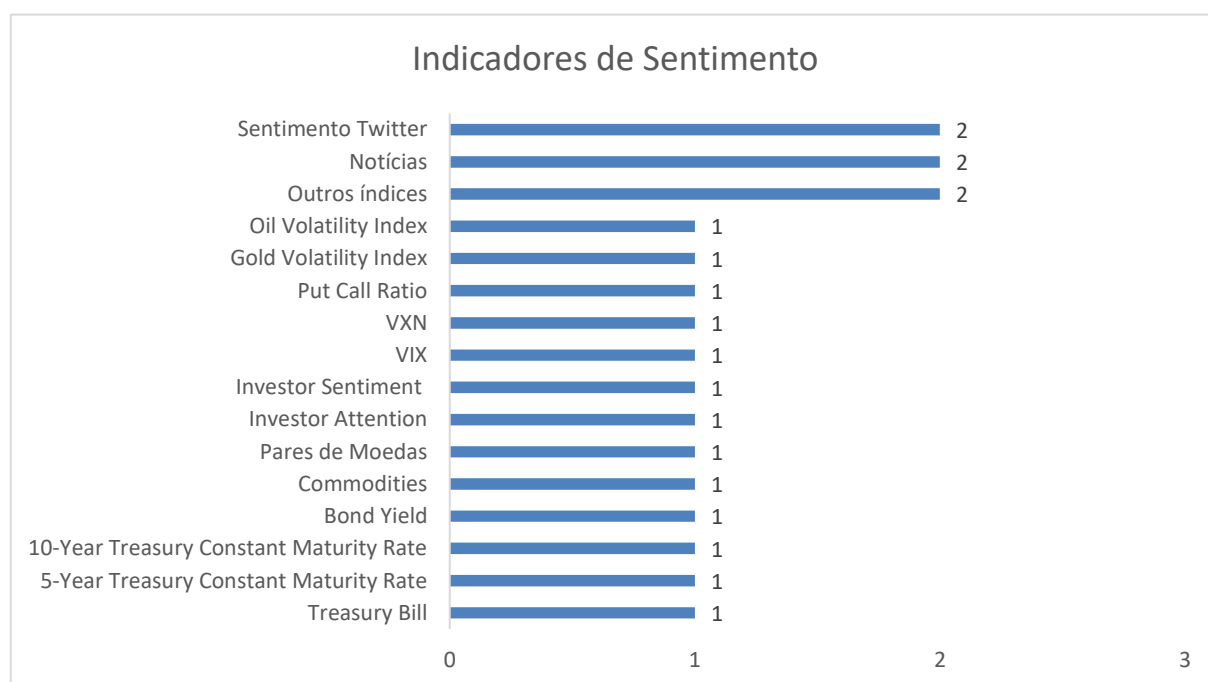


Figura 9 – Frequência dos indicadores de sentimento utilizados

2.5.3. Modelos de *Machine Learning* Utilizados

Dos 27 artigos selecionados, 18 desenvolveram modelos preditivos. Destes 18, 10 artigos desenvolveram modelos de regressão e 8 desenvolveram modelos de classificação, como pode ser observado na Figura 10.

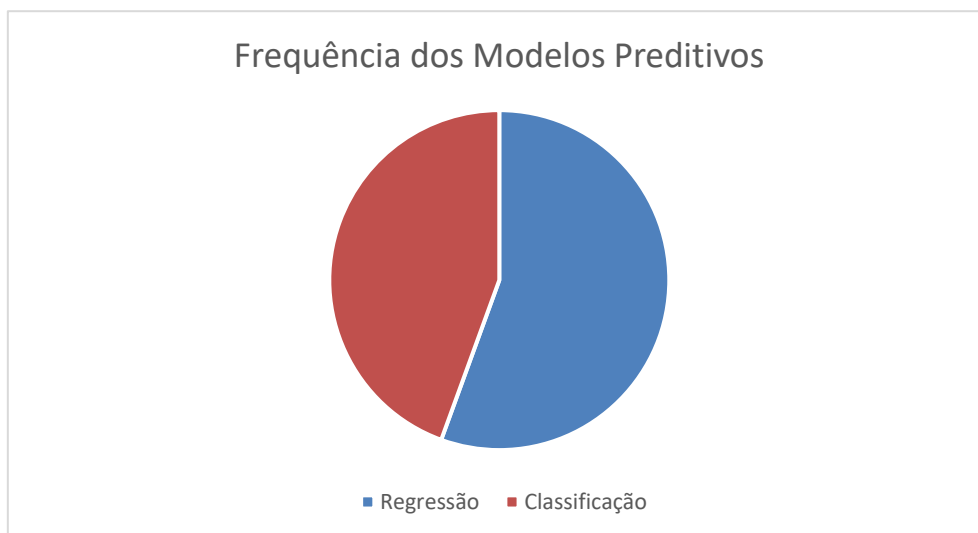


Figura 10 – Frequência dos modelos preditivos utilizados

Na Figura 11, pode-se observar os modelos mais frequentes nos estudos selecionados.

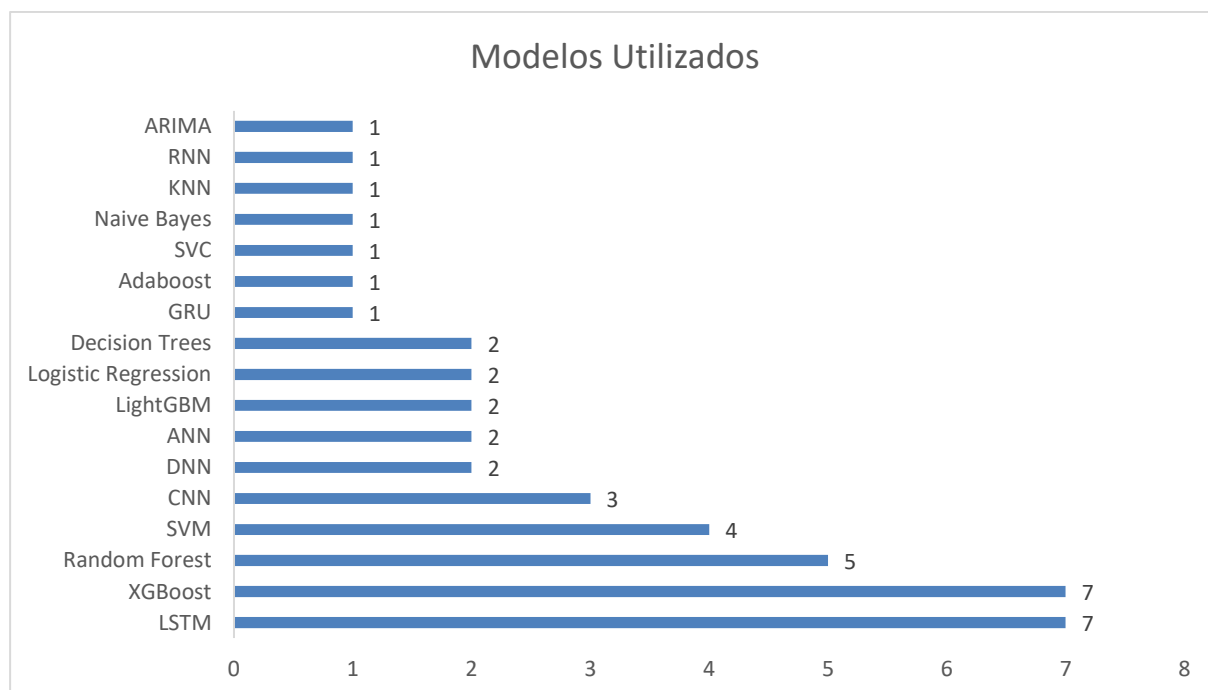


Figura 11 - Frequência dos modelos utilizados

Entre os modelos mais frequentes, destacam-se o *XGBoost*, *LSTM* e *Random Forest*, sendo que modelos como o *XGBoost*, *Random Forest*, *SVM*, entre outros foram sendo mais utilizados para

problemas de classificação, enquanto *LSTM* e outros tipos de redes neurais para a resolução de problemas de regressão.

2.6. Conclusões Finais

Ao analisar os resultados dos artigos selecionados, pode-se concluir que grande parte dos artigos selecionados são recentes, publicados nas revistas mais prestigiadas na área, assegurando maior integridade do conteúdo dos artigos. Quando se trata de prever mercados financeiros existem dois mercados que emergem dos demais, os mercados financeiros norte-americanos e chineses, sendo os índices *S&P 500*, *NASDAQ* e *SSE* os mais requisitados pelos autores.

Entre a literatura selecionada, não é muito comum a utilização conjunta de diferentes tipos de indicadores, sendo que dos artigos considerados nenhum utiliza a junção dos três tipos de indicadores (técnicos, fundamentais e de sentimento). Os indicadores mais utilizados na previsão dos mercados financeiros foram os indicadores técnicos, onde o *RSI* destaca-se dos demais, contudo foram utilizados 19 indicadores diferentes, o que mostra a diversidade existente entre os indicadores técnicos. Os indicadores fundamentais utilizados, são indicadores presentes nas demonstrações de resultados das empresas, tentando medir a sua saúde económica e financeira, sendo um bom complemento com outros tipos de indicadores, como ilustraram *Nti et al.* (2021). É difícil medir o que são indicadores de sentimento, contudo foram utilizadas várias abordagens diferentes pelos autores. *Whang et al.* (2019), analisam o sentimento das publicações em redes sociais para conseguir medir o sentimento dos investidores, enquanto *Hoseinzade et al.* (2019), utilizam a variação de outros índices de modo a conseguir perceber qual o sentimento à volta dos mercados. Por outro lado, temos *Campisi et al.* (2023), que de uma forma mais pragmática utilizam o *VIX* e o *VXN* como indicadores de sentimento. Assim sendo, podemos observar que existe uma lacuna na literatura selecionada, relacionada com a pouca existência de abordagens complementares aquando da seleção dos tipos de indicadores a utilizar na previsão dos mercados financeiros.

A literatura divide-se no tipo de abordagem à previsão dos mercados, há autores que preferem prever o preço das ações (regressão), enquanto outros preferem prever a tendência (classificação). Em termos de previsão da tendência o modelo *XGBoost*, *Random Forest* e *SVM* destacam-se dos demais, sendo que o algoritmo que tende a ter melhor performance é o *XGBoost*, seguindo-se as *Random Forests*. Quanto à previsão dos preços das ações, sobressai o desenvolvimento de modelos *deep learning*, como são os casos das redes neurais.

CAPÍTULO 3

Metodologia

Para o desenvolvimento desta dissertação, optou-se pela adoção da metodologia *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Este modelo metodológico é amplamente reconhecido e adotado devido à sua estrutura flexível e centrada no projeto (Figura 12), que facilita a compreensão e a gestão das complexidades inerentes aos projetos de ciência de dados.

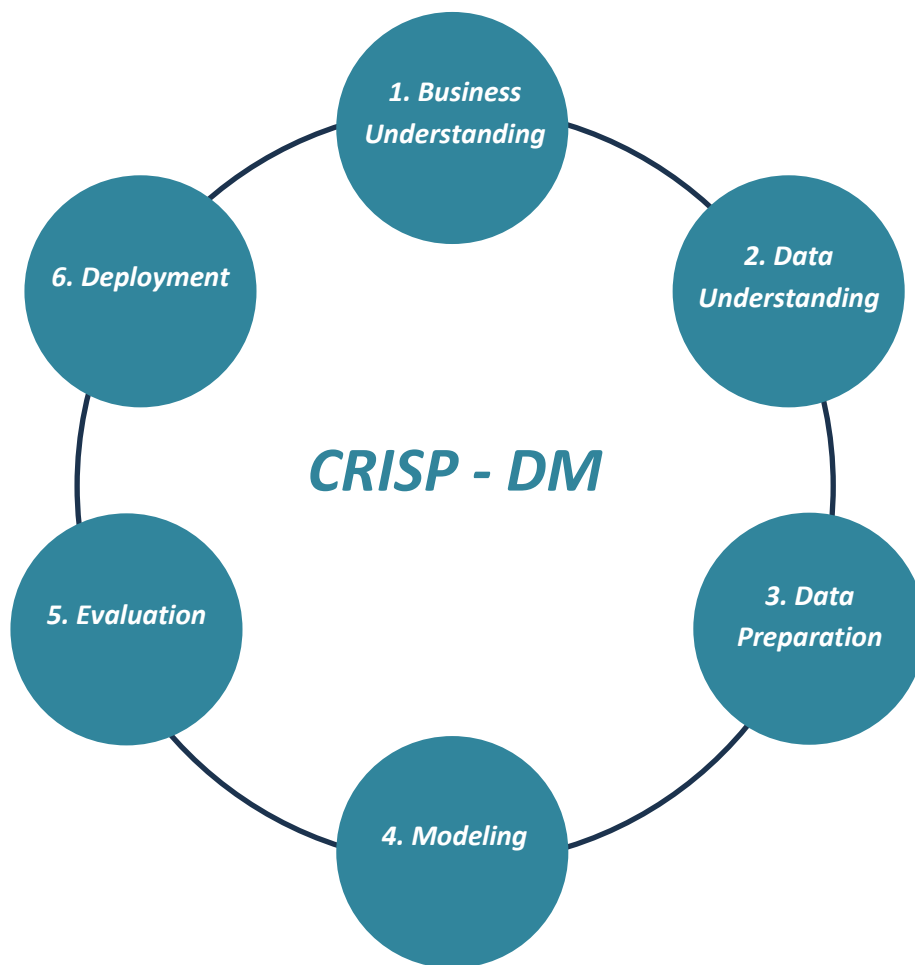


Figura 82 - Metodologia adotada para o desenvolvimento da tese

A primeira etapa, *business understanding*, é uma etapa crucial, pois deve atender aos objetivos definidos inicialmente, desenvolvendo um plano de ação para alcançá-los. Neste caso, passa por compreender como a junção de vários tipos de indicadores melhora ou não, a classificação da tendência do índice *NASDAQ 100*. A segunda fase, *data understanding*, consiste na recolha de dados e posterior análise exploratória e descritiva das variáveis recolhidas. É também, de suma importância

avaliar a qualidade dos dados, tentando identificar possíveis problemas com os mesmos. A terceira etapa, *data preparation*, é uma etapa intensiva que envolve a transformação e a escolha das variáveis mais relevantes tendo em conta o problema em análise. É nesta fase que se tratam os problemas detetados nos dados na fase anterior. Por vezes também é necessário moldar os dados de modo a atenderem os requisitos dos métodos de modelação selecionados. Na quarta etapa, *modeling*, é onde são aplicados os algoritmos aos dados já tratados e na próxima fase, *evaluation*, é onde se avalia a qualidade dos modelos e se verifica se os resultados obtidos atendem aos objetivos definidos na primeira etapa. Na última etapa, *deployment*, os resultados gerados pelo modelo são aplicados num ambiente real de negócio.

Deste modo, pretende-se assumir um compromisso com o rigor científico, relevância prática e a inovação metodológica, enfrentando eficazmente os desafios apresentados ao longo desta dissertação.

3.1. Contextualização do NASDAQ 100

Os mercados financeiros constituem a espinha dorsal da economia global, facilitando a alocação de recursos e a transferência de risco. Estes mercados abrangem uma variedade de instrumentos, incluindo ações, obrigações, pares de moedas, derivados, entre outros, operando em plataformas que vão desde bolsas de valores tradicionais até mercados de balcão. Um estudo realizado pela *International Finance Corporation* durante 1976 e 1993, revela que o PIB tem tendência de crescimento mais acentuada em economias com mercados financeiros líquidos, caracterizados por um alto volume de transações, do que em economias com mercados financeiros mais ilíquidos, mercados onde o nível de transações é menor. Pensar-se que os mercados financeiros são uma atividade simples, é enganador, talvez por isso é que cerca de 80% dos investidores de retalho perdem dinheiro no mercado de ações. Há vários fatores que afetam o mercado, por exemplo fatores externos, os quais o investidor individual não tem perceção do momento, dado que estes episódios tendem a acontecer recorrentemente ao longo do tempo.

Primeiramente, o crash de 1929 e a Grande Depressão tiveram implicações duradouras nos mercados financeiros e na regulamentação económica. Uma das consequências imediatas deste acontecimento foi a perda de confiança no sistema financeiro. O público e os investidores tornaram-se céticos em relação à estabilidade dos mercados financeiros o que levou, consequentemente, a uma redução significativa do investimento.

Anos mais tarde, mais conhecido como, *Black Monday*, entrou para a história como um dos dias mais catastróficos dos mercados financeiros, onde o índice *Dow Jones*, caiu cerca de 22,6%. Ainda hoje as suas causas são objeto de estudo e análise entre os economistas, no entanto, um dos principais motivos está relacionado com o pânico sentido pelos investidores pela chegada de uma possível recessão.

No início deste milénio, presenciamos a explosão da *Dotcom Bubble*. Os anos 90 ficaram marcados pelo avanço rápido da tecnologia em diversas áreas, sendo liderada pelo surgimento da Internet, e pelas empresas *dotcom* que foram supervalorizadas ao longo dessa década. Em novembro de 2000 o índice *NASDAQ 100*, fortemente associado ao setor tecnológico, tinha triplicado o seu preço quando comparado com o seu surgimento e no mês seguinte a bolha do *dotcom* rebentou nos mercados financeiros tendo levado a uma queda de cerca de 10% no índice.

Em 2008, devido à crise do *subprime* e à falência de um dos maiores bancos do mundo, o *Lehman Brothers*, o índice *Dow Jones* caiu mais de 50%.

Mais recentemente, a pandemia do COVID-19, o início da guerra na Ucrânia e os elevados níveis de inflação, têm provocado consequências visíveis nos mercados financeiros.

Como referido anteriormente, o índice *NASDAQ 100*, é um dos maiores índices dos mercados globais e, portanto, à semelhança dos mercados também foi afetado por alguns destes eventos (e outros) mencionados anteriormente. Ao analisar a evolução do preço de fecho do índice *NASDAQ* entre o início dos anos 90 e o final de 2023 (ver Figura 13), é possível perceber que este foi afetado em quatro momentos, com destaque para dois deles. De 1990 houve um crescimento exponencial, provocado essencialmente pelo surgimento das empresas *dotcom*, que acaba por rebentar e afetar significativamente o índice, tendo só recuperado em 2017. A crise de 2008 afetou o índice, mas de uma forma menos acentuada, assim como a COVID-19. No final de 2022, início de 2023, este volta a ter uma grande queda muito devido à instabilidade vivida com os elevados níveis de inflação. De referir que, nos dias de hoje o índice já recuperou os valores anteriores, não tendo sido muito afetado ao longo do tempo ao contrário da bolha do *dotcom*, que levou alguns anos a recuperar.

Assim sendo, destaca-se a explosão da bolha do *dotcom* e a instabilidade devido à inflação como os dois momentos que mais marcaram negativamente o preço de fecho do índice *NASDAQ 100*.

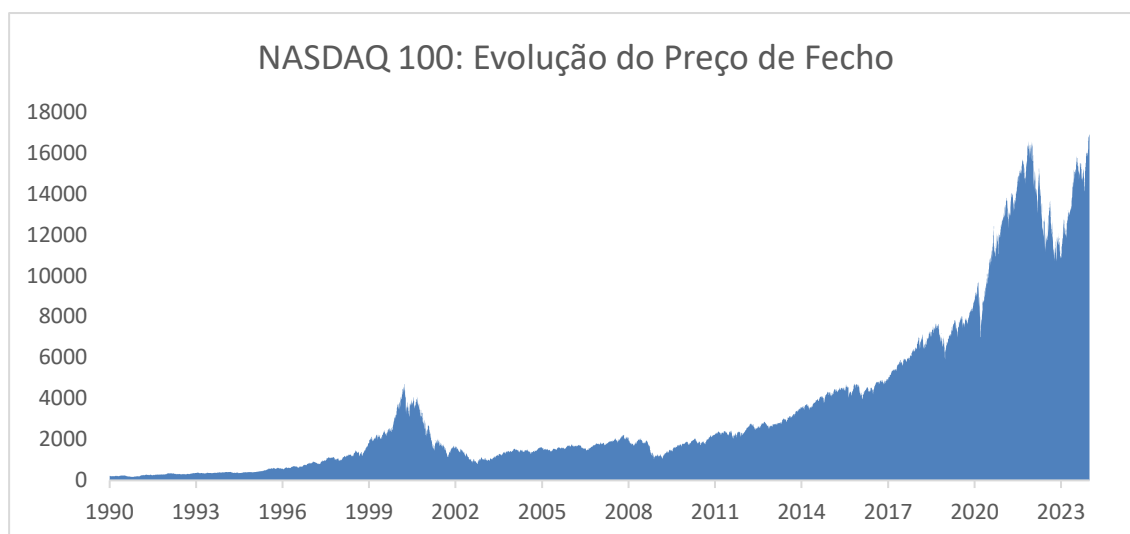


Figura 93 - Evolução do preço de fechamento do índice NASDAQ 100

Explicada a dificuldade em obter sucesso com os investimentos realizados nos mercados de ações, uma forma mais cautelosa do ponto de vista do risco do investidor passa pelo investimento em índices que, tendencialmente, ao longo do tempo têm retornos consistentes.

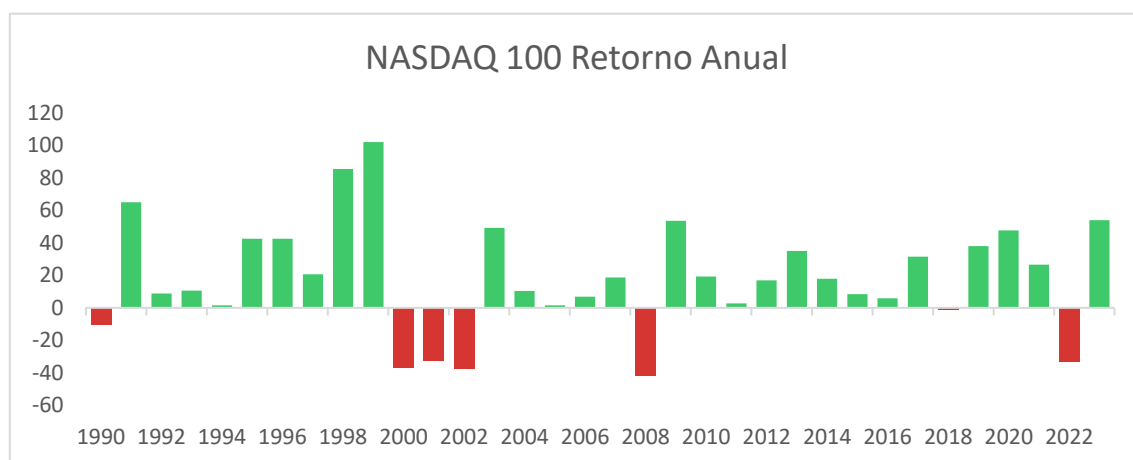


Figura 104 - Retorno anual do índice NASDAQ 100

Ao analisar o retorno anual do índice *NASDAQ 100* desde 1990 (ver Figura 14, página 20), é possível destacar dois acontecimentos: o primeiro é que ao longo do tempo este índice mostra-se bastante consistente no que respeita aos retornos anuais, nos últimos 34 anos, em 7 anos os retornos foram negativos e em 26 anos foram positivos; o segundo, está relacionado com o que foi mencionado anteriormente acerca dos fatores externos que afetam significativamente os mercados, 6 dos 7 anos em que o índice obteve retornos negativos coincidiu com períodos em que toda a economia global, incluindo os mercados financeiros, foram afetados por fatores externos (explosão da bolha do *dotcom*, crise financeira do *subprime*, COVID-19 e inflação). Assim sendo, podemos confirmar que durante este

período o índice *NASDAQ 100* é um índice regular, uma vez que obtém consistentemente retornos positivos. De mencionar, que o índice é muito sensível a crises tecnológicas (como foi a crise do *dotcom*) e crises globais. Contudo, ao contrário de outros índices durante a COVID-19 o índice obteve retornos positivos.

Ao contrário do índice *S&P 500* e *Dow Jones*, o índice *NASDAQ 100* é constituído por algumas empresas sediadas fora dos Estados Unidos da América¹, como é o caso da *Lululemon Athletica* (Canadá), *MercadoLibre* (Argentina), *Coca-Cola* (Países Baixos), entre outras como pode ser observado na Tabela 2.

Tabela 2 - Peso dos países representados no índice *NASDAQ 100*

País	Peso
Estados Unidos da América	96,54%
Países Baixos	0,94%
Ilhas Caimão	0,81%
Argentina	0,63%
Reino Unido	0,57%
Canadá	0,49%

Regra geral o índice *NASDAQ 100*, está associado ao setor tecnológico, contudo existem mais nove setores que são representados no índice, embora com uma representatividade bem menor, como se pode observar na Figura 15.

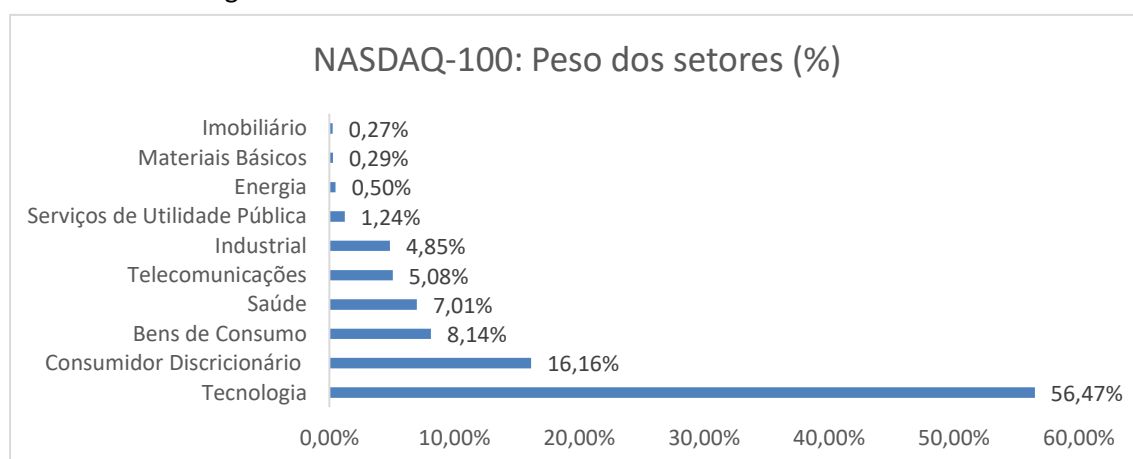


Figura 115 - Peso dos setores presentes no índice *NASDAQ 100*

¹ <https://etf.dws.com/en-gb/produkte/nasdaq-100-ucits-etf-1c/> - informações consultadas a dia 6 de janeiro de 2024.

O índice *NASDAQ 100* é um índice tecnológico, visto que é o setor que predomina na sua composição, o que pode levar a que este tenha um impacto mais reduzido aquando de choques externos (por exemplo, COVID-19). Contudo, a diversificação de vários setores, como os setores de bens de consumo e consumidor discricionário pode ajudar a explicar o impacto da inflação no índice, ao albergar um vasto leque de empresas de diferentes setores.

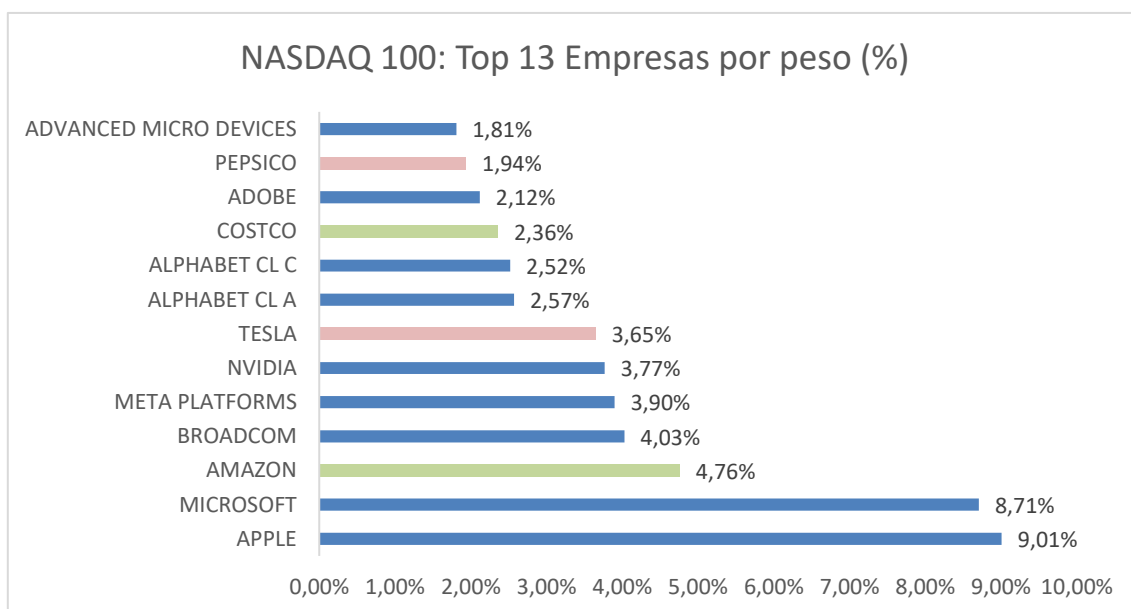


Figura 126 - Top 13 empresas com maior peso no índice NASDAQ 100

O índice *NASDAQ 100* é constituído por 101 empresas, sendo os respetivos pesos no índice muito variáveis e, consequentemente, a influência das empresas é bastante diferenciada. As treze empresas com maior peso no índice representam cerca de 51% do peso do índice, pelo que é perceptível a influência das mesmas. Na Figura 16, são apresentadas as treze empresas e os respetivos pesos, como se pode observar, nove empresas pertencem ao setor tecnológico, duas ao setor bens de consumo (*Tesla* e *Pepsico*) e duas ao setor consumidor discricionário (*Amazon* e *Costco*).

3.2. Variáveis Recolhidas

As variáveis foram recolhidas (Anexo C) através de fontes públicas como é o caso do *Yahoo Finance*², *FRED*³, *Macrotrends*⁴, entre outros. Esta fase divide-se em dois passos: validar a qualidade e consistência dos dados; quando não era possível garantir a qualidade dos mesmos, recorreu-se a técnicas de forma a alcançar a mesma.

O primeiro passo consiste em recolher o máximo de informação relacionada com o índice *NASDAQ*. Deste modo foram recolhidos os preços de abertura e fecho, os valores máximo e mínimo diário e o volume transacionado. A variável *target* foi calculada através da diferença relativa entre o preço de fecho n e o preço de fecho de $n-1$.

Os indicadores técnicos são ferramentas matemáticas e estatísticas aplicadas a séries temporais financeiras para auxiliar a previsão de futuras oscilações dos mercados. Estes são fundamentais na análise técnica, uma abordagem que consiste na previsão dos preços através do estudo de dados passados, principalmente o preço e o volume transacionado. A premissa subjacente é que os padrões históricos do mercado e as tendências de preços tendem a repetir-se ao longo do tempo, e esses padrões podem ser identificados e explorados com a ajuda de indicadores técnicos.

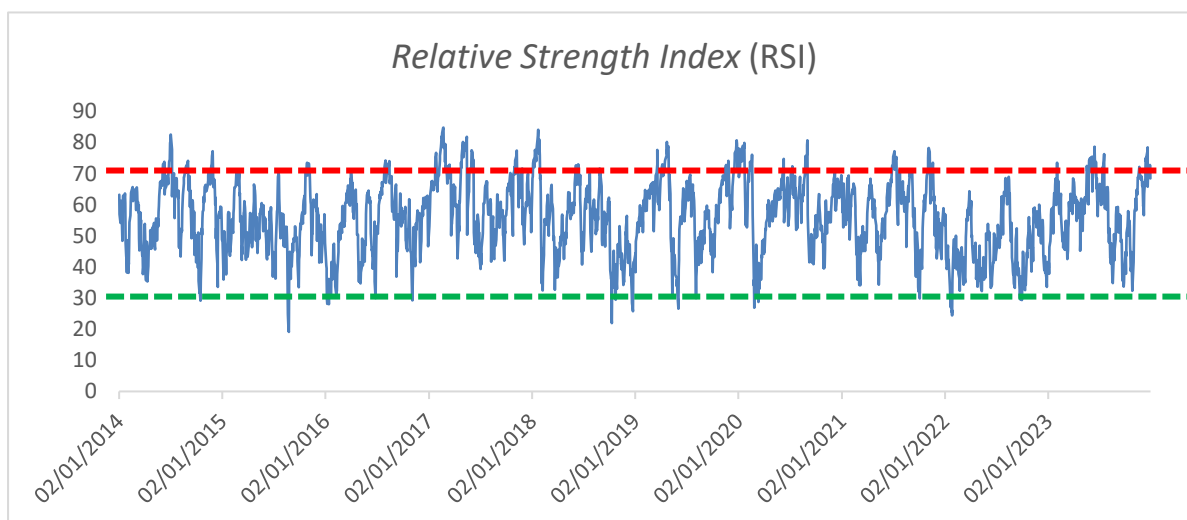


Figura 137 - Visualização do Relative Strength Index (RSI)

² <https://finance.yahoo.com/quote/%5ENDX/history/>

³ <https://fred.stlouisfed.org/>

⁴ <https://www.macrotrends.net/>

Devido à definição de indicadores técnicos, o processo de criação dos mesmos decorreu sem infortúnios, uma vez que são calculados através das variáveis recolhidas inicialmente relacionadas com o índice *NASDAQ* (preços abertura/fecho, volume, etc..). O cálculo dos mesmos foi efetuado tendo em conta as suas fórmulas que se encontram disponíveis em *sites* como a *investopedia*⁵, *corporate finance institute*⁶ ou *trading view*⁷, que têm uma grande reputação dentro da área dos mercados financeiros. Em alguns casos, a fórmula dos indicadores técnicos também pode ser consultada entre a literatura selecionada.

Em relação às médias móveis foram calculadas oito tipos de médias móveis a 14 e 50 dias, que posteriormente levaram à criação de oito indicadores, com o objetivo de sinalizar o movimento do índice. De mencionar que, quando a média móvel a 14 dias fosse superior à média móvel a 30 dias assumia valor 1 e o inverso valor 0.

O RSI é um dos indicadores técnicos mais conhecidos dentro da análise técnica, sendo uma ferramenta utilizada para identificar a força de uma ação (ou mercado). Valores acima de 70 tendem a indicar que o mercado está em sobrecompra e é esperada uma reversão da tendência a qualquer momento, enquanto valores abaixo de 30, sugere que o mercado está em sobrevenda, pelo que o ativo pode estar subvalorizado (Figura 17). Contudo, o RSI, à semelhança dos outros indicadores, não pode ser usado de forma isolada, pois as tendências do mesmo, não correspondem sempre às tendências dos ativos.

Foram escolhidos 3 períodos, onde o RSI ultrapassou os 30 e os 70 pontos, setas verdes e vermelhas, respetivamente (Figura 18). Desta forma, é possível verificar que a longo prazo as duas primeiras setas verdes funcionaram bem, pois o índice *NASDAQ* continuou a sua tendência de subida após esses dois momentos, contudo as duas primeiras setas vermelhas (valores acima de 70) falharam completamente. As últimas setas verdes e vermelhas também não confirmaram a hipótese do RSI, visto que tiveram tendências contrárias aos valores do RSI nesses períodos. Portanto, pode-se confirmar que o RSI, é um dos indicadores mais utilizados na análise técnica, mas que usado de forma isolada, não tem um grande valor acrescentado para ajudar na previsão dos ativos financeiros.

⁵ <https://www.investopedia.com/>

⁶ <https://corporatefinanceinstitute.com/>

⁷ <https://www.tradingview.com/support/folders/43000587405-built-in-indicators/>



Figura 148 - Visualização do efeito do RSI no preço de fecho do índice NASDAQ 100

Os indicadores fundamentais são métricas económicas e financeiras utilizadas para avaliar a saúde financeira e o potencial de crescimento de uma empresa ou economia. São cruciais na análise fundamental, uma abordagem que se concentra na determinação do valor intrínseco de um ativo com base em fatores económicos, financeiros e outros qualitativos e quantitativos relevantes. Diferentemente da análise técnica, que tem em conta o preço e o volume transacionado para prever movimentos futuros, a análise fundamental examina rendimentos, dividendos, novos produtos, taxas de juros, e outros dados para avaliar o valor e o potencial de investimento de um ativo.

O primeiro indicador criado sinaliza os dois dias antes de um feriado. Segundo, Ariel (1990), existe uma correlação positiva entre os dias anteriores a feriados e a subida dos mercados financeiros. Dado que o índice *NASDAQ* não é uma empresa isolada, mas sim um conjunto de empresas procedeu-se à recolha de informações macroeconómicas das 23 empresas com maior peso no índice, criando-se de seguida uma média ponderada para cada variável de modo a termos uma aproximação dos indicadores fundamentais do índice *NASDAQ*. Deste processo foram recolhidas 16 variáveis, em que não houve nenhum constrangimento devido há falta ou inconsistências dos dados. De salientar que, devido à natureza dos indicadores fundamentais, a menor periodicidade em que estes se encontram disponíveis é trimestralmente, pelo que foi aplicado o método de imputação *forward fill*, de modo a preencher os valores anunciados no fim do trimestre para os restantes dias até ao próximo lançamento dos resultados.

Os indicadores de sentimento são ferramentas analíticas que medem o sentimento dos investidores em relação aos mercados financeiros, sendo uma componente vital na análise do comportamento do mercado. Outros índices e *commodities* atuam, frequentemente, como

indicadores de sentimento, refletindo as expectativas e percepções coletivas dos investidores sobre a saúde económica global. Por exemplo, o aumento no preço do ouro pode indicar um sentimento de aversão ao risco, uma vez que os investidores tendem a procurar refúgio em ativos considerados seguros em tempos de incerteza económica ou política. A premissa subjacente é que o sentimento do mercado pode influenciar ou prever tendências dos preços, antes que estes se materializem em movimentos concretos.

Assim sendo, foram recolhidos indicadores de sentimento que demonstram o grau de confiança ou receio dos investidores, mas também o índice *S&P 500*, *Dow Jones* e algumas *commodities* como, por exemplo os preços do ouro ou do petróleo.

Uma das variáveis recolhidas, o *Economic Policy Uncertainty Index* para os Estados Unidos (EPU) é uma medida quantitativa projetada para refletir o nível de incerteza política e económica do país. Desenvolvido por *Scott Baker*, *Nicholas Bloom* e *Steven Davis*⁸, o índice visa capturar a extensão em que a incerteza política afeta as decisões económicas de indivíduos, empresas e governos. A incerteza económica pode ter implicações significativas para o investimento, o emprego e o crescimento, tornando este índice uma ferramenta valiosa para analistas, investidores e formuladores de políticas.

O EPU é amplamente utilizado por economistas e analistas para entender como a incerteza política e económica pode afetar a economia. Altos níveis de incerteza podem levar a uma redução no investimento e no consumo, que, por sua vez, pode desacelerar o crescimento económico.

Ao analisar graficamente o EPU (Figura 19), verifica-se que, regra geral, este oscila entre 0 e 200, sendo que depois existem picos relacionados com a incerteza predominante no período em questão. Acontecimentos globais como o resultado do referendo sobre a permanência ou saída da União Europeia por parte do Reino Unido são momentos que trazem algum tipo de incerteza para a economia global. O *Brexit*, em finais de junho de 2016, foi o primeiro acontecimento que fez o EPU ultrapassar a marca dos 300 pontos desde o início de 2014. O segundo acontecimento foram as eleições presidenciais dos Estados Unidos da América, dia 8 de novembro de 2016, que colocaram frente a frente os candidatos *Hillary Clinton* e *Donald Trump*, tendo sido o último o vencedor, contrariando as sondagens. Até ao início de 2020 o EPU voltou a registar algumas oscilações, sendo que, no final de março de 2020 este indicador chegou a máximos históricos, devido à pandemia global da COVID-19.

⁸ <https://fred.stlouisfed.org/series/USEPUINDXD>

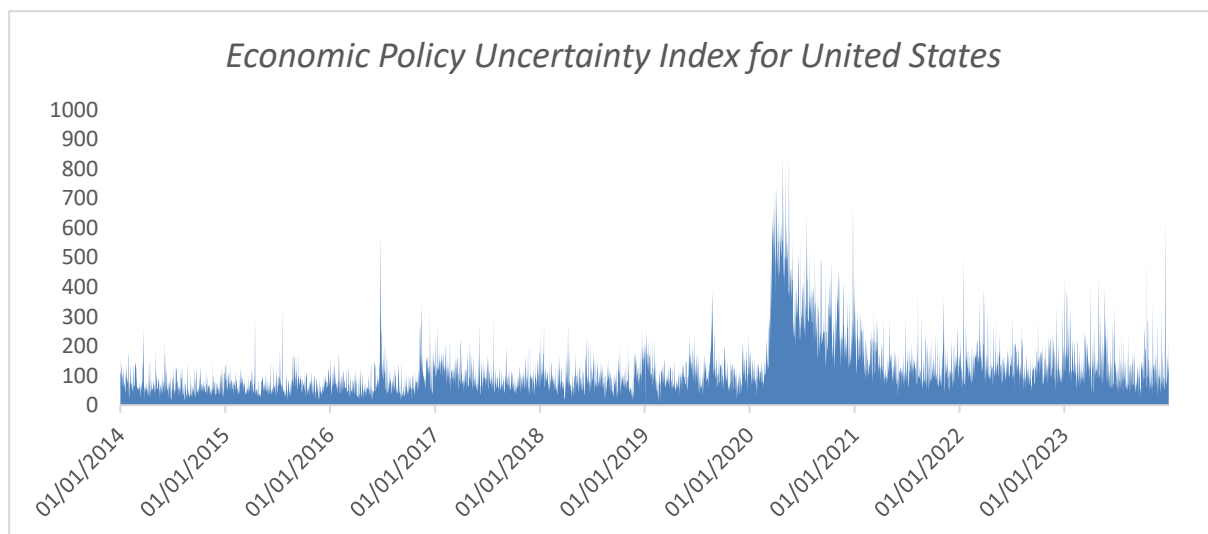


Figura 159 - Visualização do Economic Policy Uncertainty Index for United States

Durante cerca de 2 anos, o EPU alcançou níveis muito elevados devido à incerteza causada pela pandemia que levou a um aumento assinalável da mortalidade, tendo sido necessário a adoção de medidas restritivas pelos governos como o fecho das fronteiras. Recentemente, os conflitos na Europa e Médio Oriente, têm tido consequências fatais nos países onde as mesmas decorrem, contribuindo significativamente para a subida generalizada dos preços. Como resposta à inflação instalada foram adotadas medidas, intensificando a incerteza do período que vivemos.

Num âmbito geral foram recolhidas 68 variáveis (Anexo C), onde a consistência dos dados foi garantida na maioria das variáveis, uma vez que se tratavam de dados profissionais previamente tratados, com exceção para o *ICE Bofa US High Yield Index Effective Yield*, que mede a performance dos títulos de alto rendimento e *10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity*, que avalia o *spread* (a diferença) dos títulos do tesouro do mercado norte-americano com maturidade de 10 e 2 anos. Para estas duas variáveis verificou-se a presença de valores nulos, 35 e 107 casos, respetivamente, que foram tratados fazendo a imputação através da média do valor anterior e o valor posterior ao valor nulo.

3.3. Seleção de variáveis

A seleção das variáveis representa um pilar fundamental para o desenvolvimento de modelos preditivos eficazes e eficientes. A escolha criteriosa das variáveis influencia diretamente a capacidade preditiva do modelo, tentando reduzir variáveis ruidosas para o mesmo. Neste contexto, a abordagem adotada para a seleção de variáveis, foi realizada individualmente para cada categoria dos indicadores. Isto, porque o objetivo é testar como se comporta a previsão do índice *NASDAQ* utilizando os

diferentes tipos de indicadores, numa primeira parte de forma isolada e, posteriormente, combinando as diferentes variáveis de diferentes categorias de indicadores. Assim sendo, para realizar uma redução de dimensionalidade, os métodos utilizados tiveram por base o algoritmo LASSO⁹ (Anexo D) e a análise da matriz de correlações (Anexo E), visando uma seleção eficiente e robusta das variáveis dentro de cada tipo de indicador. A combinação destas duas técnicas metodológicas permite uma seleção de variáveis eficaz e fundamentada, adequada às especificidades de cada tipo de indicador considerado nesta pesquisa. Neste sentido, as variáveis a incluir na fase da modelação podem ser observadas na Tabela 3.

Tabela 3 - Variáveis selecionadas

Categoria	LASSO	Matriz de Correlações
Indicadores Técnicos	<i>Exponential Moving Average</i>	<i>Relative Strength Index (RSI)</i>
	<i>Smoothed Moving Average</i>	<i>Stochastic Oscillator</i>
	<i>Triangular Moving Average</i>	<i>Commodity Channel Index</i>
	<i>Price Percentage Oscillator (PPO)</i>	<i>Average Directional Index</i>
Indicadores Fundamentais	<i>Earnings per Share (EPS)</i>	<i>Pré Feriados</i>
	<i>Gross Margin</i>	<i>Cash on Hand</i>
	<i>Operating Margin</i>	<i>EBITDA Margin</i>
	<i>Pre Tax Profit Margin</i>	<i>Price to Earnings Ratio</i>
	<i>Debt Equity Ratio</i>	<i>Current Ratio</i>
Indicadores de Sentimento	<i>Dow Variation</i>	<i>US Economic Policy Uncertainty Index</i>
	<i>SP 500 Variation</i>	<i>Oil Variation</i>
	<i>Open Gas</i>	<i>Fear and Greed Index</i>
	<i>Gold Variation</i>	
	<i>10-Year Treasury Constant Maturity Minus 2- Year Treasury Constant Maturity</i>	
	<i>High Volatility Nasdaq Index</i>	
	<i>Sentimento do Consumidor</i>	

⁹ https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html

3.4. Modelação e Resultados

Nesta secção será apresentado o algoritmo utilizado para o desenvolvimento dos modelos e os respetivos resultados obtidos durante as diferentes experiências. Após concluídas as três primeiras etapas da metodologia CRISP-DM, onde assentou, por exemplo, a recolha de variáveis, o seu tratamento e respetiva seleção, obteve-se uma base de dados com 20 indicadores. Dos quais, 8 indicadores técnicos, 10 são indicadores fundamentais e 10 indicadores de sentimento (Tabela 3). Fazem ainda parte da base de dados, variáveis relacionadas com o índice, tais como, os preços de abertura e fecho e o volume transacionado do índice como o preço de abertura e fecho e o volume transacionado do índice. Desta forma, a base de dados resultante é constituída por 23 variáveis, relativas ao período compreendido entre janeiro de 2014 e dezembro de 2023, o que se traduz em 2516 observações. No Anexo F, podem ser consultadas algumas das variáveis incorporadas na fase da modelação.

O algoritmo selecionado, tendo em conta as boas *performances* obtidas nos artigos selecionados na revisão de literatura, foi o *XGBoost* através da utilização da biblioteca *xgboost* do *python*¹⁰. Este é um algoritmo de *machine learning* baseado em árvores de decisão, que pertence à categoria de *ensemble learning*, tendo um *framework* de *boosting* denominada de *gradient boosting*. Este método combina múltiplos modelos mais fracos, geralmente árvores de decisão, onde cada novo modelo corrige os erros cometidos pelos anteriores através da adição de aprendizagens mais fracas para o *ensemble*. Uma das características distintivas deste algoritmo é a capacidade que este tem na prevenção de *overfitting*, através da regularização do *alfa* (L1) e *lambda* (L2), ajudando também a reduzir a complexidade do modelo.

Recorreu-se à otimização dos valores dos hiper-parâmetros através da técnica de *Randomized Search*¹¹, a qual é diferente da *Grid Search*¹².

- *N_estimators*, define o número de árvores no modelo; caso seja definido um número elevado de árvores e, se a *learning rate* for alta, existe a possibilidade de *overfitting*.
- *Max_depth*, controla a profundidade máxima das árvores. Árvores mais profundas podem modelar relações mais complexas, mas também podem levar ao *overfitting*.
- A *learning rate*, é a taxa de aprendizagem, onde valores menores reduzem o *overfitting*, mas exigem mais árvores para um bom desempenho.
- *Min child weight*, define o peso mínimo de um nó folha da árvore. Um valor alto força o modelo a fazer *splits* apenas quando resultam numa redução significativa de perda, o que ajuda a evitar *overfitting*.

¹⁰ https://xgboost.readthedocs.io/en/stable/python/python_intro.html

¹¹ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

¹² https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

- *Subsample*, é a proporção de amostras que são utilizadas para treinar cada árvore. Ao usar-se frações dos dados em cada árvore reduz-se o risco de *overfitting*.
- *Colsample bytree*, é a proporção de variáveis que são utilizadas para treinar cada árvore. A utilização de um número reduzido de variáveis pode prevenir o *overfitting*.
- *Reg alpha* e *Reg lambda*, são os termos de regularização L1 e L2, respetivamente. Os quais, têm como objetivo penalizar os pesos para evitar *overfitting*.

A definição do espaço dos hiper-parâmetros a serem otimizados pode ser consultado na Tabela 4, página 30.

Tabela 4 - Definição do espaço dos hiper-parâmetros

Hiper-paramêtro	Intervalo
<i>N_estimators</i>	<i>[50, 1000]</i>
<i>Max_depth</i>	<i>[3, 10]</i>
<i>Learning_rate</i>	<i>[0.001, 0.3]</i>
<i>Min_child_weight</i>	<i>[1, 10]</i>
<i>Subsample</i>	<i>[0.5, 1]</i>
<i>Colsample_bytree</i>	<i>[0.5, 1]</i>
<i>Reg_alpha</i>	<i>[0, 1]</i>
<i>Reg_lambda</i>	<i>[0, 1]</i>

A divisão dos dados em conjunto de treino e teste foi feita de forma que os dados de 2014 a 2020 (inclusive) fossem do conjunto de treino e de 2021 a 2023 fosse do conjunto de teste. Esta divisão, traduz-se em 7 anos de dados para o conjunto de treino, que correspondem a 1763 observações e 3 anos de dados para o conjunto de teste, que correspondem a 753 observações. Conclui-se assim que a divisão treino-teste foi aproximadamente 70% e 30%, respetivamente.

De mencionar, que foram desenvolvidas 8 iterações. A primeira será a *baseline*, na qual serão consideradas as variáveis preço de abertura, o volume e um indicador de cada categoria de indicadores. Posteriormente serão realizadas 7 iterações nas quais serão incorporadas as três categorias de indicadores de forma isolada e combinada. As variáveis de cada categoria de indicadores serão as variáveis selecionadas na secção 3.3 (ver Tabela 3). De seguida, serão apresentados os resultados, onde serão detalhados os valores otimizados dos hiper-parâmetros, bem como os valores das métricas selecionadas, *accuracy*, *precision*, *recall* e *f1-score* para os dados dos conjuntos de treino e teste.

3.4.1. Iteração 1: *Baseline*

A primeira iteração tinha como propósito a definição de uma *baseline*, sendo este passo fundamental na fase da modelação, uma vez que permite uma análise comparativa com os modelos consequentes. Deste modo, foi usado o modelo *XGBoost*, tendo-se optado por utilizar o preço de abertura, o volume transacionado, o RSI (indicador técnico), EPS (indicador fundamental) e EPU (indicador de sentimento), uma vez que são das variáveis mais reconhecidas entre cada categoria de indicadores. Os valores otimizados dos hiper-parâmetros, bem como os valores das métricas são apresentados na Tabela 5 (ver página 31).

Tabela 5 - Iteração 1: *Baseline*

<i>Baseline</i>			
Hiper-paramêtro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	108	Accuracy: 67%	Accuracy: 60%
Max_depth	3		
<i>Learning_rate</i>	0.01	Precision: 66%	Precision: 64%
Min_child_weight	1		
<i>Subsample</i>	0.71	Recall: 86%	Recall: 57%
Colsample_bytree	0.84		
<i>Reg_alpha</i>	0.12	F1-Score: 74%	F1-Score: 60%
Reg_lambda	0.38		

De forma geral, com uma *accuracy* de 60%, a *baseline* é melhor que uma tomada de decisão aleatória (50%). No entanto, é importante salientar o valor alto do *recall* no conjunto de treino, com 86%, e a sua queda drástica no conjunto de teste para 57%, indicando que a *baseline* no conjunto de treino é bastante eficaz em detetar as subidas da tendência do índice *NASDAQ*, o que pode traduzir-se numa fraca generalização no conjunto de teste, nas próximas iterações, irá ser acompanhada a evolução desta métrica.

3.4.2. Iteração 2: Indicadores Técnicos

Após a criação da *baseline* procedeu-se à criação do modelo usando apenas indicadores técnicos. Os resultados obtidos são apresentados na Tabela 6.

Tabela 6 - Iteração 2: *Indicadores técnicos*

<i>Indicadores técnicos</i>			
Hiper-paramêtro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	96	Accuracy: 89%	Accuracy: 86%
Max_depth	4		
<i>Learning_rate</i>	0.07	Precision: 90%	Precision: 88%
Min_child_weight	2		
<i>Subsample</i>	0.63	Recall: 91%	Recall: 85%
Colsample_bytree	0.98		

<i>Reg_alpha</i>	0.1	F1-Score: 91%	F1-Score: 86%
<i>Reg_lambda</i>	0.9		

Comparando com a *baseline*, os indicadores técnicos aumentaram consideravelmente a performance em todas as métricas. A *accuracy*, a *precision* e o *f1-score* subiram mais de 20 pontos percentuais, enquanto o *recall* aumentou cerca de 30 pontos percentuais. A diferença do *recall* no conjunto de treino e teste não é tão significativa como verificado na iteração anterior.

3.4.3. Iteração 3: Indicadores Fundamentais

A terceira iteração é o desenvolvimento do modelo utilizando os indicadores fundamentais. Na Tabela 7, encontram-se os valores otimizados dos hiper-parâmetros, bem como os valores das métricas.

Tabela 7 - Iteração 3: Indicadores fundamentais

Indicadores fundamentais			
Hiper-parâmetro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	998	Accuracy: 63%	Accuracy: 53%
<i>Max_depth</i>	3		
<i>Learning_rate</i>	0.03	Precision: 63%	Precision: 55%
<i>Min_child_weight</i>	7		
<i>Subsample</i>	0.99	Recall: 84%	Recall: 73%
<i>Colsample_bytree</i>	0.54		
<i>Reg_alpha</i>	0.52	F1-Score: 72%	F1-Score: 63%
<i>Reg_lambda</i>	0.41		

A performance do modelo utilizando os indicadores fundamentais decresce acentuadamente registrando os piores resultados em termos de *accuracy* e *precision*. Contudo, todas as métricas apresentam valores superiores a 50%, o que se traduz em resultados ligeiramente melhores que escolhas arbitrárias.

3.4.4. Iteração 4: Indicadores de Sentimento

A quarta iteração consiste no desenvolvimento do modelo utilizando os indicadores de sentimento. Os resultados obtidos são apresentados na Tabela 8.

Tabela 8 - Iteração 4: Indicadores de sentimento

Indicadores de sentimento			
Hiper-parâmetro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	100	Accuracy: 90%	Accuracy: 90%
<i>Max_depth</i>	9		
<i>Learning_rate</i>	0.03	Precision: 92%	Precision: 91%

Min_child_weight	5		
Subsample	0.52	Recall: 91%	Recall: 91%
Colsample_bytree	0.90		
Reg_alpha	0.26	F1-Score: 91%	F1-Score: 91%
Reg_lambda	0.45		

A melhor performance é obtida utilizando os indicadores de sentimento, de forma isolada. Todas as métricas registam valores iguais ou superiores a 90%, valor não alcançado por nenhum outro modelo até então. Destaque para o facto das métricas do conjunto treino e teste serem bastante semelhantes.

3.4.5. Iteração 5: Indicadores Técnicos e Fundamentais

A quinta iteração consiste na criação de um modelo que combine indicadores técnicos e indicadores fundamentais.

Tabela 9 - Iteração 5: Indicadores técnicos e fundamentais

Indicadores técnicos e fundamentais			
Hiper-paramêtro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	762	Accuracy: 89%	Accuracy: 85%
Max_depth	4		
<i>Learning_rate</i>	0.007	Precision: 90%	Precision: 87%
Min_child_weight	6		
Subsample	0.76	Recall: 91%	Recall: 85%
Colsample_bytree	0.97		
Reg_alpha	0.028	F1-Score: 90%	F1-Score: 86%
Reg_lambda	0.8		

Os resultados da junção dos indicadores técnicos com os indicadores fundamentais são ligeiramente piores quando comparados com os resultados do modelo que utiliza apenas indicadores técnicos, como pode ser observado na Tabela 9.

3.4.6. Iteração 6: Indicadores Técnicos e Sentimento

Nesta iteração são combinados os indicadores técnicos e de sentimento, sendo os resultados obtidos apresentados na Tabela 10.

Tabela 10 - Indicadores técnicos e de sentimento

Indicadores técnicos e de sentimento			
Hiper-paramêtro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	171	Accuracy: 93%	Accuracy: 91%
Max_depth	5		
<i>Learning_rate</i>	0.007	Precision: 93%	Precision: 91%
Min_child_weight	1		
<i>Subsample</i>	0.69	Recall: 94%	Recall: 92%
Colsample_bytree	0.96		
<i>Reg_alpha</i>	0.45	F1-Score: 94%	F1-Score: 91%
Reg_lambda	0.1		

Até esta iteração, os melhores resultados tinham sido obtidos aquando da utilização isolada dos indicadores de sentimento. Após a junção destes dois tipos de indicadores, os resultados alcançados são os melhores, tendo a *accuracy* e o *recall* aumentado cerca de 1 ponto percentual, enquanto as outras métricas se mantiveram no nível dos 91%, no conjunto de teste.

3.4.7. Iteração 7: Indicadores Fundamentais e Sentimento

Nesta iteração são considerados os indicadores fundamentais e de sentimento. Os resultados obtidos são apresentados na Tabela 11.

Tabela 11 - Iteração 7: Indicadores fundamentais e de sentimento

Indicadores fundamentais e de sentimento			
Hiper-paramêtro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	126	Accuracy: 92%	Accuracy: 90%
Max_depth	3		
<i>Learning_rate</i>	0.03	Precision: 93%	Precision: 90%
Min_child_weight	6		
<i>Subsample</i>	0.98	Recall: 93%	Recall: 91%
Colsample_bytree	0.74		
<i>Reg_alpha</i>	0.96	F1-Score: 93%	F1-Score: 90%
Reg_lambda	0.01		

À semelhança do que sucedeu na quinta iteração, onde a inclusão dos indicadores fundamentais piorou ligeiramente a performance do modelo em comparação com os resultados obtidos quando foram apenas utilizados os indicadores técnicos, nesta iteração a junção dos indicadores

fundamentais com os indicadores de sentimento também piorou ligeiramente a performance do modelo em comparação com a quarta iteração, onde foram utilizados apenas os indicadores de sentimento. A *precision* e o *f1-score* desceram 1 ponto percentual, enquanto a *accuracy* e o *recall* se mantiveram nos 91%.

Até à sétima iteração, conclui-se que a junção dos indicadores técnicos com os indicadores de sentimento tende a melhorar ligeiramente a *performance* do modelo, enquanto a junção dos indicadores fundamentais, tanto com os indicadores técnicos como com os indicadores de sentimento tendem a piorar as *performances* dos modelos.

3.4.8. Iteração 8: Indicadores Técnicos, Fundamentais e Sentimento

A última iteração consiste na junção de todos os indicadores utilizados em cada categoria de indicador. Na Tabela 12, encontram-se os valores otimizados dos hiper-parâmetros, bem como os valores das métricas.

Tabela 12 - Iteração 8: Indicadores técnicos, fundamentais e de sentimento

Indicadores técnicos, fundamentais e de sentimento			
Hiper-parâmetro	Valor	Métricas treino	Métricas teste
<i>N_estimators</i>	258	Accuracy: 94%	Accuracy: 91%
<i>Max_depth</i>	8		
<i>Learning_rate</i>	0.008	Precision: 94%	Precision: 92%
<i>Min_child_weight</i>	4		
<i>Subsample</i>	0.69	Recall: 95%	Recall: 92%
<i>Colsample_bytree</i>	0.99		
<i>Reg_alpha</i>	0.19	F1-Score: 94%	F1-Score: 92%
<i>Reg_lambda</i>	0.6		

A última iteração foi a que proporcionou os melhores resultados, em todas as métricas, conseguindo superar de forma ligeira a melhor iteração, que tinha sido aquando da junção dos indicadores técnicos e de sentimento.

A *accuracy* manteve-se nos 91%, traduzindo-se numa ótima taxa de acerto das previsões do modelo relacionadas com a previsão da tendência do índice *NASDAQ 100*. A *precision* registou uma subida de 1 ponto percentual, indicando a boa capacidade do modelo na forma como evita os falsos positivos, ou seja, não tende a errar as previsões quando a tendência do índice é crescente, evitando a geração de previsões erradas aos investidores, pois se o modelo prevê que a tendência é de subida os investidores podem estar mais confiantes que esta previsão é certa e que poucas vezes irão incorrer em perdas. O *recall* manteve-se nos 92%, demonstrando a capacidade do modelo em evitar falsos negativos, ajudando os investidores a não perder oportunidades de entrar numa tendência que é crescente e que possivelmente irá contribuir com retornos. O *F1-Score*, sendo a métrica de

estabilização entre a *precision* e o *recall*, atribui o mesmo peso a estas duas métricas, salientando a boa capacidade do modelo tendo em consideração tanto os falsos positivos como os falsos negativos, teve um incremento de 1 ponto percentual, chegando aos 92%.

Assim sendo, é possível destacar que os modelos que incluem pelo menos indicadores de sentimento, obtêm resultados em todas as métricas superiores a 90%. Os indicadores técnicos utilizados de forma isolada originam bons resultados, mas a sua capacidade preditiva aumenta com a conjugação com os indicadores de sentimento. Por outro lado, os indicadores fundamentais, têm resultados fracos quando são utilizados de forma isolada, ao nível da *baseline* e piora os resultados quando são utilizados com os indicadores técnicos e indicadores de sentimento. Contudo, quando são juntos aos indicadores técnicos e de sentimento, a *performance* do modelo melhora em comparação com a utilização conjunta dos dois tipos de indicadores referidos.

Com o objetivo de facilitar a análise e comparação dos resultados obtidos, na Tabela 13 são apresentados os valores das métricas obtidos para cada um dos modelos desenvolvidos.

Tabela 13 - Resultados

Indicadores	Accuracy	Precision	Recall	F1-Score
Baseline	60%	64%	57%	60%
Indicadores Técnicos	86%	88%	85%	86%
Indicadores Fundamentais	53%	55%	73%	63%
Indicadores de Sentimento	90%	91%	91%	91%
Indicadores Técnicos e Fundamentais	85%	87%	85%	86%
Indicadores Técnicos e de Sentimento	91%	91%	92%	91%
Indicadores Fundamentais e de Sentimento	90%	90%	91%	90%
Indicadores Técnicos, Fundamentais e de Sentimento	91%	92%	92%	92%

Para uma melhor e mais fácil compreensão influência dos indicadores na previsão da *target*, recorreu-se à geração dos *shap values* das vinte variáveis que mais contribuem para o desempenho do modelo. Ao analisarmos a Figura 20, conseguimos perceber que a variação do índice *S&P 500* é o indicador que mais contribui para explicar as previsões geradas. Este acontecimento não é de todo surpreendente, uma vez que este índice e o índice *NASDAQ 100*, são dois dos maiores índices dos mercados financeiros globais, sendo indicadores do desempenho geral do mercado de ações norte-americano, sendo o índice *NASDAQ 100*, um índice mais relacionado com o setor tecnológico. De seguida, a variável que mais impacta o modelo é o indicador técnico *Stochastic Oscillator*, conhecido

pela sua utilização na análise técnica, especialmente em movimentos de curto prazo, sendo o %K a linha principal deste oscilador e o %D uma média móvel simples do %K, de modo a suavizar os sinais gerados pelo %K. A terceira variável que mais impacta o modelo é a variação do índice *Dow Jones Industrial Average* que é também outro grande índice presente no mercado norte-americano. Destaque para o indicador técnico *On Balance Volume*, que influencia apenas positivamente a previsão do modelo, este indicador é utilizado para ajudar a identificar a pressão de compra e venda relacionada com um determinado ativo. Também o *Relative Strength Index*, um dos indicadores técnicos mais consensuais na análise técnica tem um grande impacto no modelo, fechando o top 5 das variáveis que mais explicam as previsões do modelo.

Deste modo, das vinte variáveis que mais explicam as previsões do modelo dez são indicadores técnicos, nove são indicadores de sentimento, e claro destaque para a presença de um indicador fundamental, *Gross Margin*.

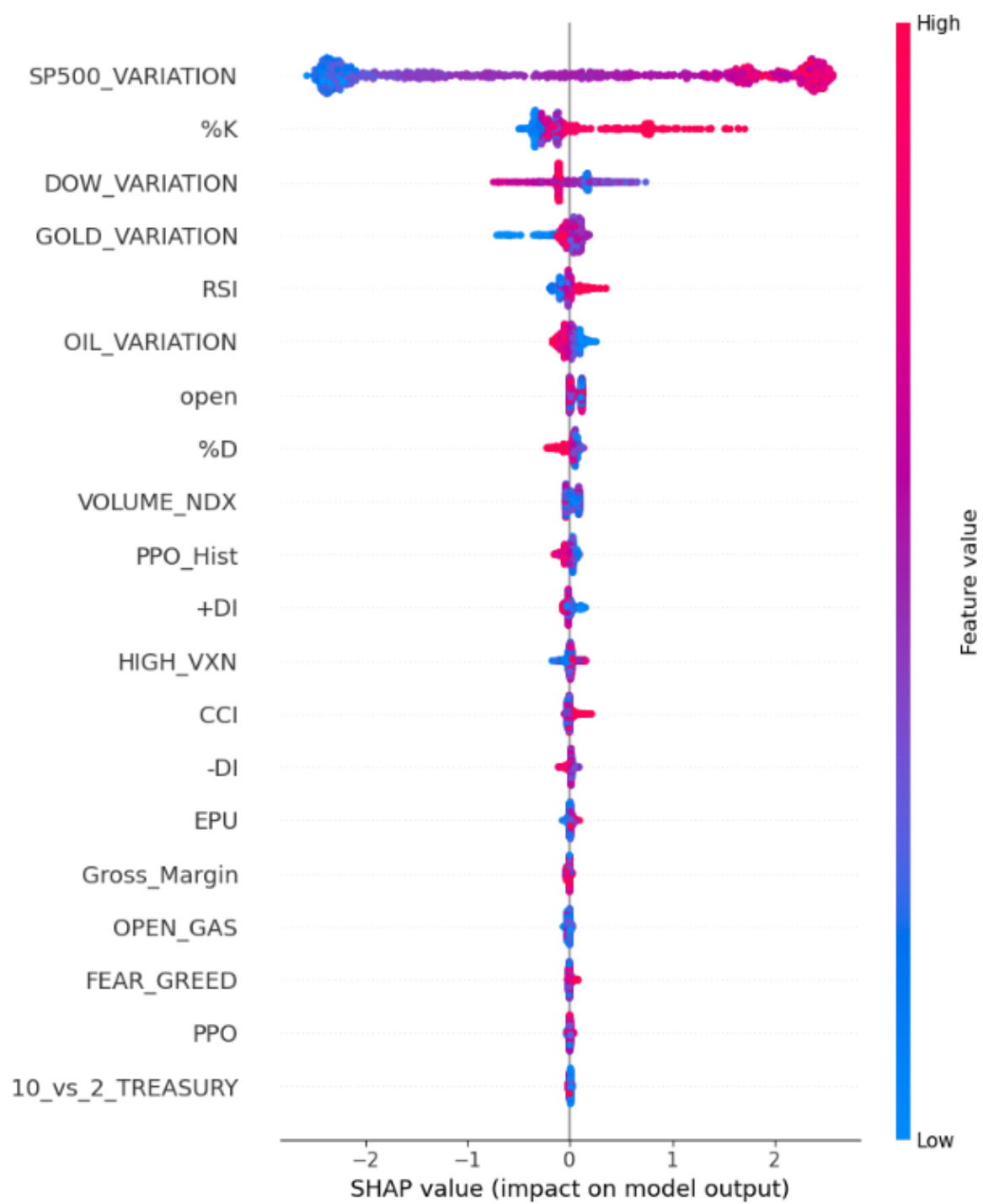


Figura 20 – Variáveis mais explicativas para as previsões do modelo (shap values)

CAPÍTULO 4

Conclusões

No decorrer desta dissertação, foi aprofundado o desafiante e dinâmico campo da previsão de mercados financeiros, com foco particular na previsão do índice *NASDAQ 100*. O principal objetivo era analisar o desempenho dos modelos utilizando três tipos de indicadores mais relevantes na literatura: técnicos, fundamentais e de sentimento.

Os dados utilizados no desenvolvimento deste estudo são de periodicidade diária, à exceção dos indicadores fundamentais em que a sua natureza é trimestral. Os dados foram recolhidos de fontes públicas, como o *Yahoo Finance*, *Macro trends*, *FRED (Federal Reserve Economic Data)*, *CBOE (Chicago Board Options Exchange)*, *Pyinvesting*, *Policy Uncertainty* e *Investing*. Os indicadores técnicos, por estarem relacionados com os preços/volume do índice *NASDAQ 100*, foram construídos recorrendo à biblioteca *Pandas*, onde algumas das fórmulas foram recolhidas no site *Investopedia*. No final deste processo foram obtidos 68 indicadores, onde 5 estavam relacionados com os preços de abertura, fecho, volume transacionado, valores mais alto e baixo; 19 indicadores técnicos; 22 indicadores fundamentais; e 22 indicadores de sentimento.

Após o processo de recolha de variáveis, procedeu-se ao processo de seleção de variáveis a incluir na fase da modelação. Esta seleção teve por base a aplicação de duas técnicas: algoritmo LASSO e matriz de correlações. As variáveis com um *score* mais alto tendo em conta a aplicação do algoritmo LASSO, foram selecionadas, assim como as variáveis que tivessem mais correlacionadas com a *target*. Esta etapa foi aplicada de forma isolada a cada tipo de indicador.

É também importante realçar a boa performance do algoritmo *XGBoost*, à semelhança dos estudos de *Naik e Mohan (2021)* e *Deng et al. (2021)*, este que foi um dos algoritmos mais utilizados entre os artigos selecionados, correspondeu às expectativas geradas aquando da revisão bibliográfica. A boa *performance* do melhor modelo, que surgiu da junção dos três tipos de indicadores, traduz-se em métricas na ordem dos 91% para todas as quatro métricas calculadas, *accuracy*, *precision*, *recall* e *F1-score*. Esta *performance* superou de forma ligeira os resultados obtidos quando apenas se juntou os indicadores técnicos e de sentimento, pelo que se pode concluir que os indicadores fundamentais tiveram um ligeiro impacto positivo. Estas métricas são ligeiramente inferiores comparadas com o estudo de *Nti et al. (2021)*, onde a *accuracy*, *precision*, *recall* e *f1-score* foram de cerca de 98%, 99%, 90% e 97%, respetivamente. Este estudo abordou a junção de indicadores fundamentais e de

sentimento para a previsão do *Ghana Stock Exchange* recorrendo a LSTM. Por outro lado, o estudo de Hoseinzade & Haratizadeh (2019), também recorreu a modelos de *deep learning* (CNN) para a previsão de índices como o *NASDAQ*, *S&P 500*, entre outros, recorrendo a indicadores técnicos e de sentimento obtendo métricas de *f1-score* a rondar os 55%.

Entre as vinte variáveis com mais impacto, destaca-se a presença das três categorias de indicadores, com claro destaque para os indicadores técnicos e de sentimento, com dez e nove indicadores, respetivamente e a presença de um indicador fundamental. Os indicadores técnicos mais relevantes foram o *Stochastic Oscillator*, o *Relative Strength Index*, o *Price Percentage Oscillator*, *Commodity Channel Index*, entre outros. Alguns destes indicadores, foram dos mais utilizados entre os artigos selecionados, como pode ser observado na Figura 7 (página 12). No estudo de Hoseinzade & Haratizadeh (2019), os indicadores de sentimento que mais impactaram o modelo foram outros índices. Neste estudo a variação dos índices *S&P 500* e *Dow Jones*, foram umas das variáveis mais importantes para o modelo, relevando alguma sintonia entre estes dois estudos. Também os títulos de alto rendimento (*junk bonds*) e o *Economic Policy Uncertainty for United States* fecharam o top 20 das variáveis mais relevantes.

A fraca representação de indicadores fundamentais nas variáveis mais relevantes, pode estar relacionada com o facto de os dados dos indicadores fundamentais serem trimestrais, enquanto os outros indicadores têm uma frequência diária. Por outro lado, quando os resultados financeiros das empresas são publicados, os indicadores mais sensíveis são comparados ao período homólogo ou ao período antecedente. Muitas vezes são anunciadas as estimativas das empresas e só depois é que o efeito se traduz nos mercados indicando se os indicadores na realidade cumpriram ou não com as expectativas geradas pelas estimativas. Neste estudo, os indicadores fundamentais foram utilizados considerando os valores reais. De salientar também a ausência da variável pré feriado entre as 20 mais relevantes. O estudo de Ariel (1990), concluía que os mercados em dias precedentes aos feriados, tendiam a crescer, neste estudo essa variável não se traduziu numa das variáveis que teve mais impacto, muito provavelmente, porque apenas existem 11 feriados importantes por ano nos Estados Unidos, e as outras variáveis eram mais relevantes na classificação da tendência nos dias antecedentes aos feriados.

Futuramente, à semelhança da literatura selecionada, a adoção de mais algoritmos por exemplo, redes neuronais, pode ser um passo a percorrer de modo a melhorar os resultados. Além disso, a melhoria da qualidade dos dados relacionados com os indicadores fundamentais deve ser uma preocupação, principalmente a recolha de indicadores esperados para o algoritmo ter a capacidade de comparar o resultado real e o resultado esperado tentando depois sinalizar de uma forma mais relevante os indicadores fundamentais, melhorando a *performance* dos modelos.

Diante destas descobertas, é evidente que a previsão de mercados financeiros continua a ser um desafio complexo e multifacetado, exigindo uma abordagem cuidadosa e aberta à inovação. Este estudo contribui para o avanço do campo, destacando a importância da integração de múltiplos tipos de indicadores e ressaltando a necessidade de adaptação às particularidades dos dados financeiros e dos mercados em constante evolução.

Referências Bibliográficas

- H. Akaike (1973), "Information Theory as an Extension of the Maximum Likelihood Principle", in B. N. Petrov, and F. Csaki, (Eds.), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp. 267-281.
- D.T. Anderson, J.C. Bezdek, M. Popescu, and J.M. Keller (2010), "Comparing Fuzzy, Probabilistic, and Possibilistic Partitions", *IEEE Transactions on Fuzzy Systems*, 18(5), 906-918.
- An, Z., Wu, Y., Hao, F., Chen, Y., He, X., A novel hierarchical feature selection with local shuffling and models reweighting for stock price forecasting. *Expert Systems with Applications, Volume 249, Part A*, 2024.
- Ariel, R. A. (1990). High Stock Returns before Holidays: Existence and Evidence on Possible Causes. *Journal of Finance*, 45(5), 1611–1626.
- Barak, S., Arjmand, A., Ortobelli, S., Fusion of multiple diverse predictors in stock market, *Information Fusion, Volume 36*, 2017, Pages 90-102.
- Beniwal, M., Singh, A. and Kumar, N., Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices. *Engineering Applications of Artificial Intelligence, Volume 129*, 2024.
- Campisi, G., Muzzioli, S., Baets, D. B., A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices, *International Journal of Forecasting*, 2023.
- Chen, J., Wen, Y., Nanekaran, Y., Suzaiddola, M., Chen, W. and Zhang, D., (2023), Machine learning techniques for stock price prediction and graphic signal recognition. *Engineering Applications of Artificial Intelligence, Volume 121*, 2023.
- Chou, S.J., Chen, E. K., Optimizing investment portfolios with a sequential ensemble of decision tree-based models and the FBI algorithm for efficient financial analysis, *Applied Soft Computing, Volume 158*, 2024.
- Deng, S., Huang, X., Qin, Z., Fu, Z., Yang, T., A novel hybrid method for direction forecasting and trading of Apple Futures, *Applied Soft Computing, Volume 110*, 2021.
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417.

- Ghosh, P., Neufeld, A., Sahoo, K. J., *Forecasting directional movements of stock prices for intraday trading using LSTM and random forests*, *Finance Research Letters*, Volume 46, Part A, 2022.
- González, A., & Garcia C. A., & Colomo-Palacios, r., & Guldrís-Iglesias, F., & Gómez B. J., (2011). *CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator*. *Expert Syst. Appl.*. 38. 11489-11500.
- Gunnarsson, S. E., Isern, R. H., Kaloudis, A., Risstad, M., Vigdel, B., Westgaard, S., *Prediction of realized volatility and implied volatility indices using AI and machine learning: A review*, *International Review of Financial Analysis*, Volume 93, 2024.
- Han, Y., Kim, J., Enke, D., *A machine learning trading system for the stock market based on N-period Min-Max labeling using XGBoost*, *Expert Systems with Applications*, Volume 211, 2023.
- Hoseinzade, E., and Haratizadeh, S. (2019), *CNNpred: CNN-based stock market prediction using a diverse set of variables*. *Expert Systems with Applications*, Volume 129, 2019, Pages 273-285.
- Lin, P., Ma, S., Fildes, R., *The extra value of online investor sentiment measures on forecasting stock return volatility: A large-scale longitudinal evaluation based on Chinese stock market*, *Expert Systems with Applications*, Volume 238, Part B, 2024.
- Lo, A.W., Mamaysky, H. and Wang, J. (2000), *Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation*. *Journal of Finance*, 55: 1705-1765.
- Murphy, J., *Technical analysis of the financial markets: A comprehensive guide to trading methods and Applications*. Prentice Hall Press 1999.
- Nabipour, M., Nayyeri, P., Jabani, H. S. S., and Mosavi, A., "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in *IEEE Access*, vol. 8, pp. 150199-150212, 2020.
- Naik, N., and Mohan, B., R., *Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market*, in *IEEE Access*, vol. 9, pp. 86230-86242, 2021.
- Niu, Z., Wang, C., Zhang, H., *Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models*, *International Review of Financial Analysis*, Volume 89, 2023.
- Nti, I.K., Adekoya, A.F. & Weyori, B.A. *A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction*. *J Big Data* **8**, 17 (2021).
- Patel J., Shah, S., Thakkar, P., Kotecha, K., *Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques*, *Expert Systems with Applications*, Volume 42, Issue 1, 2015.

- Shleifer, Andrei, and Lawrence H. Summers. 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives*, 4 (2): 19-33.
- Silva, A., Neves, R., Horta, N., A hybrid approach to portfolio composition based on fundamental and technical indicators, *Expert Systems with Applications*, Volume 42, Issue 4, 2015.
- Zhang, X., Li, G., Li, Y., Zou, G., Wu, G. J., Which is more important in stock market forecasting: Attention or sentiment?, *International Review of Financial Analysis*, Volume 89, 2023.
- Zhao, X., Liu, Y., and Zhao, Q., "Cost Harmonization LightGBM-Based Stock Market Prediction," in *IEEE Access*, vol. 11, pp. 105009-105026, 2023.
- Zhou, F., Zhou, M. H., Yang, Z., Yang, L., EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction, *Expert Systems with Applications*, Volume 115, 2019.
- Wang, H., Lu, S., Zhao, L., Aggregating multiple types of complex data in stock market prediction: A model-independent framework, *Knowledge-Based Systems*, Volume 164, 2019, Pages 193-204.
- Wen, M., Li, P., Zhang, L., and Chen, Y., "Stock Market Trend Prediction Using High-Order Information of Time Series" in *IEEE Access*, vol. 7, pp. 28299-28308, 2019.

ANEXOS

Anexo A: Estudos Selecionados

Título do Artigo	Revista	Impact Factor	H-Index
<i>A novel hybrid method for direction forecasting and trading of Apple Futures</i>	<i>Applied Soft Computing</i>	8.7	190
<i>Optimizing investment portfolios with a sequential ensemble of decision tree-based models and the FBI algorithm for efficient financial analysis</i>	<i>Applied Soft Computing</i>	8.7	190
<i>Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices</i>	<i>Engineering Applications of Artificial Intelligence</i>	8	137
<i>A hybrid approach to portfolio composition based on fundamental and technical indicators</i>	<i>Expert Systems With Applications</i>	8.5	271
<i>A machine learning trading system for the stock market based on N-period Min-Max labeling using XGBoost</i>	<i>Experted Systems with Applications</i>	8.5	171
<i>A novel hierarchical feature selection with local shuffling and models reweighting for stock price forecasting</i>	<i>Expert Systems With Applications</i>	8.5	271
<i>CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator</i>	<i>Expert Systems With Applications</i>	8.5	271
<i>EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction</i>	<i>Experted Systems with Applications</i>	8.5	171
<i>The extra value of online investor sentiment measures on forecasting stock return volatility: A large-scale longitudinal evaluation based on Chinese stock market</i>	<i>Expert Systems With Applications</i>	8.5	271
<i>CNNpred: CNN-based stock market prediction using a diverse set of variables</i>	<i>Experted Systems with Applications</i>	8.5	171
<i>Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques</i>	<i>Experted Systems with Applications</i>	8.5	171
<i>Forecasting directional movements of stock prices for intraday trading using LSTM and random forests</i>	<i>Finance Research Letters</i>	10.4	101
<i>Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market</i>	<i>IEEE Access</i>	3.9	242
<i>Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis</i>	<i>IEEE Access</i>	3.9	242
<i>Cost Harmonization LightGBM-Based Stock Market Prediction</i>	<i>IEEE Access</i>	3.9	242

<i>Stock Market Trend Prediction Using High-Order Information of Time Series</i>	<i>IEEE Access</i>	3.9	242
<i>Fusion of multiple diverse predictors in stock market</i>	<i>Information Fusion</i>	18.6	156
<i>A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices</i>	<i>International Journal of Forecasting</i>	7.9	119
<i>Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models</i>	<i>International Review of Financial Analysis</i>	8.2	91
<i>Prediction of realized volatility and implied volatility indices using AI and machine learning: A review</i>	<i>International Review of Financial Analysis</i>	8.2	91
<i>Which is more important in stock market forecasting: Attention or sentiment?</i>	<i>International Review of Financial Analysis</i>	8.2	91
<i>A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction</i>	<i>Journal of Big Data</i>	8.1	76
<i>The Noise Trader Approach to Finance</i>	<i>Journal of Economic Perspectives</i>	8.4	224
<i>Efficient Capital Markets: A Review of Theory and Empirical Work</i>	<i>Journal of Finance</i>	8	352
<i>Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation</i>	<i>Journal of Finance</i>	8	352
<i>High Stock Returns before Holidays: Existence and Evidence on Possible Causes</i>	<i>Journal of Finance</i>	8	352
<i>Aggregating multiple types of complex data in stock market prediction: A model-independent framework</i>	<i>Knowledge-Based Systems</i>	8.8	169

Anexo B: Descrição dos Índices

País	Abreviatura	Descrição índice
Estados Unidos da América	<i>S&P 500</i>	<i>Standard & Poor's 500 Index</i>
Estados Unidos da América	<i>NASDAQ</i>	<i>National Association of Securities Dealers Automated Quotations</i>
Estados Unidos da América	<i>Dow Jones</i>	<i>Dow Jones Industrial Average</i>
Estados Unidos da América	<i>RUSSEL</i>	<i>The Russell 2000 Index</i>
Estados Unidos da América	<i>NYSE</i>	<i>NYSE Arca Major Market Index</i>
China	<i>SSE</i>	<i>Shanghai Composite Index</i>
China	<i>CSI</i>	<i>China Securities 500 Index</i>
China	<i>HSI</i>	<i>Hang Seng Index</i>
China	<i>TWSE</i>	<i>Taiwan Stock Exchange Capitalization Weighted Stock Index</i>
China	<i>SZSE</i>	<i>Shenzhen Stock Exchange</i>
Irão	<i>TSE</i>	<i>Iran Tehran Stock Market Index</i>
Índia	<i>BSE</i>	<i>Bombay Stock Exchange</i>
Índia	<i>NIFTY 50</i>	<i>National Stock Exchange FIFTY</i>
Alemanha	<i>DAX</i>	<i>Deutscher Aktien Index</i>
Japão	<i>Nikkei 225</i>	<i>Nikkei Stock Average</i>
Espanha	<i>IBEX</i>	<i>Índice Bursátil Español</i>

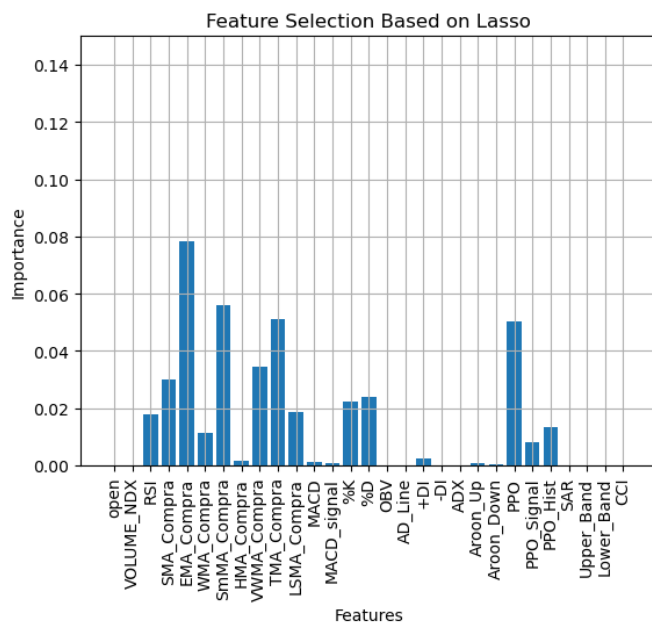
Anexo C: Variáveis Recolhidas

#	Abreviatura	Descrição	Tipo de Indicador	Fonte/Cálculo
1	<i>open</i>	Preço de abertura	Primária	<i>Yahoo Finance</i>
2	<i>high</i>	Preço mais elevado	Primária	<i>Yahoo Finance</i>
3	<i>low</i>	Preço mais baixo	Primária	<i>Yahoo Finance</i>
4	<i>close</i>	Preço de fecho	Primária	<i>Yahoo Finance</i>
5	<i>volume</i>	Volume transacionado	Primária	<i>Yahoo Finance</i>
6	<i>RSI</i>	<i>Relative Strength Index</i>	Técnico	<i>Investopedia</i>
7	<i>SMA</i>	<i>Simple Moving Average</i>	Técnico	<i>Investopedia</i>
8	<i>EMA</i>	<i>Exponential Moving Average</i>	Técnico	<i>Investopedia</i>
9	<i>WMA</i>	<i>Weighted Moving Average</i>	Técnico	<i>Investopedia</i>
10	<i>SmMA</i>	<i>Smoothed Moving Average</i>	Técnico	<i>Investopedia</i>
11	<i>HMA</i>	<i>Hull Moving Average</i>	Técnico	<i>Investopedia</i>
12	<i>VWMA</i>	<i>Volume Weighted Moving Average</i>	Técnico	<i>Investopedia</i>
13	<i>TMA</i>	<i>Triangular Moving Average</i>	Técnico	<i>Investopedia</i>
14	<i>LSMA</i>	<i>Least Squared Moving Average</i>	Técnico	<i>Investopedia</i>
15	<i>MACD</i>	<i>Moving Average Convergence Divergence</i>	Técnico	<i>Investopedia</i>
16	<i>STH</i>	<i>Stochastic Oscillator</i>	Técnico	<i>Investopedia</i>
17	<i>OBV</i>	<i>On Balance Volume</i>	Técnico	<i>Investopedia</i>
18	<i>AD</i>	<i>Accumulation Distribution Line</i>	Técnico	<i>Investopedia</i>
19	<i>ADX</i>	<i>Average Directional Index</i>	Técnico	<i>Investopedia</i>
20	<i>Aroon</i>	<i>Arron Indicator</i>	Técnico	<i>Investopedia</i>
21	<i>PPO</i>	<i>Price Percentage Oscillator</i>	Técnico	<i>Investopedia</i>
22	<i>SAR</i>	<i>Parabolic Stop and Reverse</i>	Técnico	<i>Investopedia</i>
23	<i>BB</i>	<i>Bollinger Bands</i>	Técnico	<i>Investopedia</i>
24	<i>CCI</i>	<i>Commodity Channel Index</i>	Técnico	<i>Investopedia</i>
25	<i>Pré Feriado</i>	Sinalizar dia antecedente a feriados	Fundamental	<i>Pandas</i>
26	<i>GM</i>	<i>Gross Margin</i>	Fundamental	<i>Macro Trends</i>
27	<i>OM</i>	<i>Operating Margin</i>	Fundamental	<i>Macro Trends</i>
28	<i>EBITDA_M</i>	<i>EBITDA Margin</i>	Fundamental	<i>Macro Trends</i>
29	<i>PTPM</i>	<i>Pre Tax Profit Margin</i>	Fundamental	<i>Macro Trends</i>
30	<i>NM</i>	<i>Net Margin</i>	Fundamental	<i>Macro Trends</i>
31	<i>PE</i>	<i>Price to Earnings Ratio</i>	Fundamental	<i>Macro Trends</i>
32	<i>PS</i>	<i>Price to Sales Ratio</i>	Fundamental	<i>Macro Trends</i>
33	<i>PB</i>	<i>Price to Book Ratio</i>	Fundamental	<i>Macro Trends</i>
34	<i>PFCF</i>	<i>Price to Free Cash Flow Ratio</i>	Fundamental	<i>Macro Trends</i>
35	<i>CR</i>	<i>Current Ratio</i>	Fundamental	<i>Macro Trends</i>
36	<i>QR</i>	<i>Quick Ratio</i>	Fundamental	<i>Macro Trends</i>
37	<i>DE</i>	<i>Debt Equity Ratio</i>	Fundamental	<i>Macro Trends</i>
38	<i>ROE</i>	<i>Return on Equity</i>	Fundamental	<i>Macro Trends</i>
39	<i>ROA</i>	<i>Retur non Assets</i>	Fundamental	<i>Macro Trends</i>
40	<i>ROI</i>	<i>Return on Investment</i>	Fundamental	<i>Macro Trends</i>
41	<i>RTE</i>	<i>Retur non Tanginle Equity</i>	Fundamental	<i>Macro Trends</i>
42	<i>EPS</i>	<i>Earnings per Share</i>	Fundamental	<i>Macro Trends</i>
43	<i>TA</i>	<i>Total Assets</i>	Fundamental	<i>Macro Trends</i>
44	<i>CoH</i>	<i>Cash on Hand</i>	Fundamental	<i>Macro Trends</i>

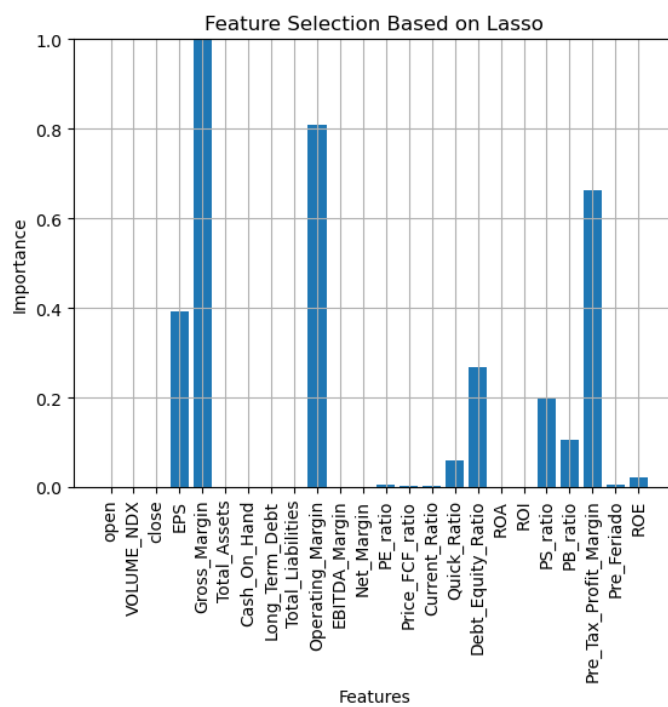
45	<i>LTD</i>	<i>Long Term Debt</i>	Fundamental	<i>Macro Trends</i>
46	<i>TL</i>	<i>Total Liabilities</i>	Fundamental	<i>Macro Trends</i>
47	<i>EPU</i>	Economic Policy Uncertainty Index for United States	Sentimento	<i>FRED</i>
48	<i>VXN</i>	NASDAQ-100 Volatility Index	Sentimento	<i>CBOE</i>
49	<i>DJIA_open</i>	Preço de abertura índice Dow Jones	Sentimento	<i>Yahoo Finance</i>
50	<i>DJIA_volume</i>	Volume transacionado índice Dow Jones	Sentimento	<i>Yahoo Finance</i>
51	<i>DJIA_variation</i>	Diferença do preço de abertura do índice Dow Jones no dia n menos no dia n-1	Sentimento	<i>Yahoo Finance</i>
52	<i>Gold_open</i>	Preço de abertura ouro	Sentimento	<i>Yahoo Finance</i>
53	<i>Gold_volume</i>	Volume transacionado ouro	Sentimento	<i>Yahoo Finance</i>
54	<i>Gold_variation</i>	Diferença do preço de abertura do ouro no dia n menos no dia n-1	Sentimento	<i>Yahoo Finance</i>
55	<i>Oil_open</i>	Preço abertura petróleo	Sentimento	<i>Yahoo Finance</i>
56	<i>Oil_volume</i>	Volume transacionado petróleo	Sentimento	<i>Yahoo Finance</i>
57	<i>Oil_variation</i>	Diferença do preço de abertura do petróleo no dia n menos no dia n-1	Sentimento	<i>Yahoo Finance</i>
58	<i>SP500_open</i>	Preço de abertura índice S&P 500	Sentimento	<i>Yahoo Finance</i>
59	<i>SP500_volume</i>	Volume transacionado índice S&P 500	Sentimento	<i>Yahoo Finance</i>
60	<i>SP500_variation</i>	Diferença do preço de abertura do índice S&P 500 no dia n menos no dia n-1	Sentimento	<i>Yahoo Finance</i>
61	<i>Junk Bonds</i>	ICE BofA US High Yield Index Effective Yield	Sentimento	<i>FRED</i>
62	<i>10 vs 2 Treasury</i>	10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity	Sentimento	<i>FRED</i>
63	<i>Gas_open</i>	Preço de abertura do gás	Sentimento	<i>Yahoo Finance</i>
64	<i>Gas_volume</i>	Volume transacionado gás	Sentimento	<i>Yahoo Finance</i>
65	<i>Gas_variation</i>	Diferença do preço de abertura do gás no dia n menos no dia n-1	Sentimento	<i>Yahoo Finance</i>
66	<i>FGI</i>	Fear and Greed Index	Sentimento	<i>Pyinvesting</i>
67	<i>DN</i>	US Daily News Index	Sentimento	<i>Policy Uncertainty</i>
68	<i>CCM</i>	Michigan Consumer Sentiment	Sentimento	<i>Investing</i>

Anexo D: Seleção das variáveis

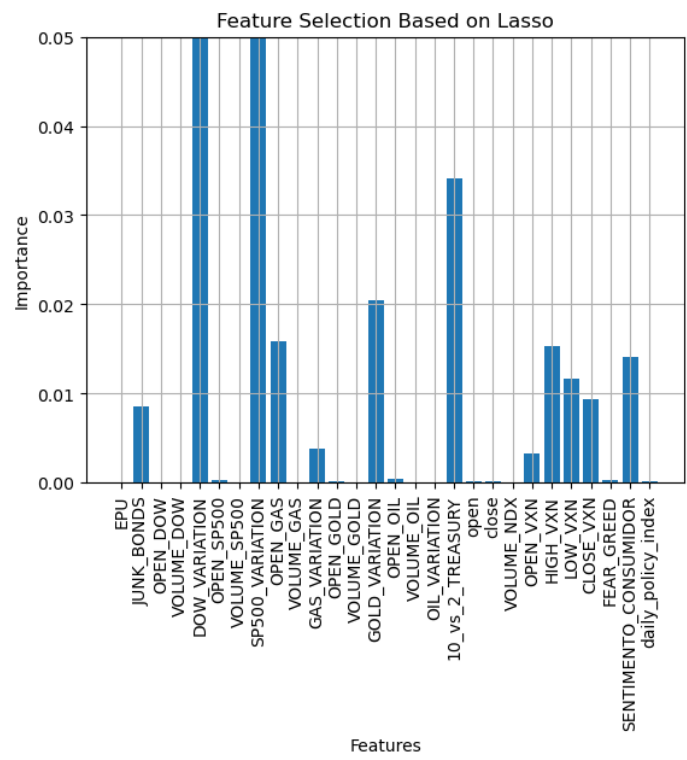
1. Indicadores Técnicos



2. Indicadores fundamentais

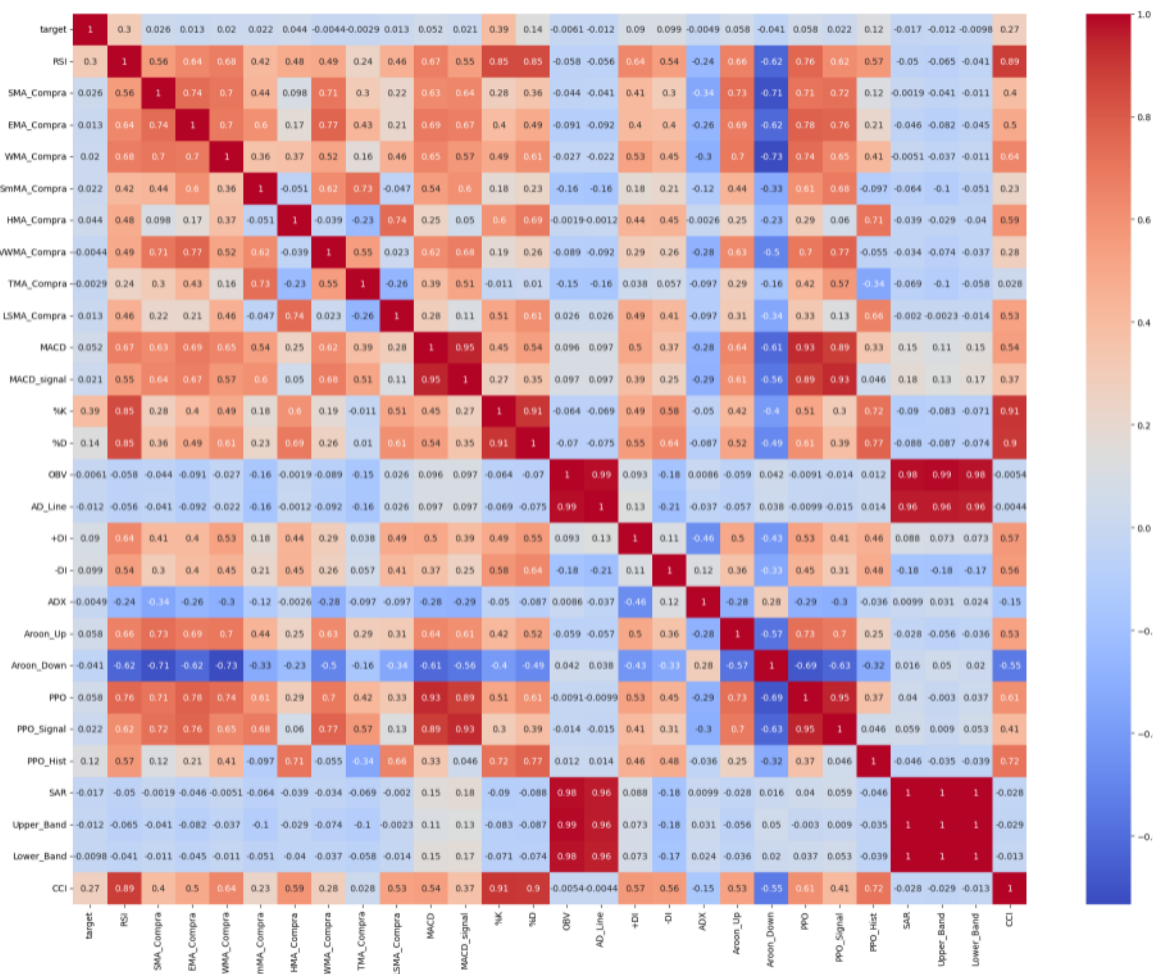


3. Indicadores de sentimiento

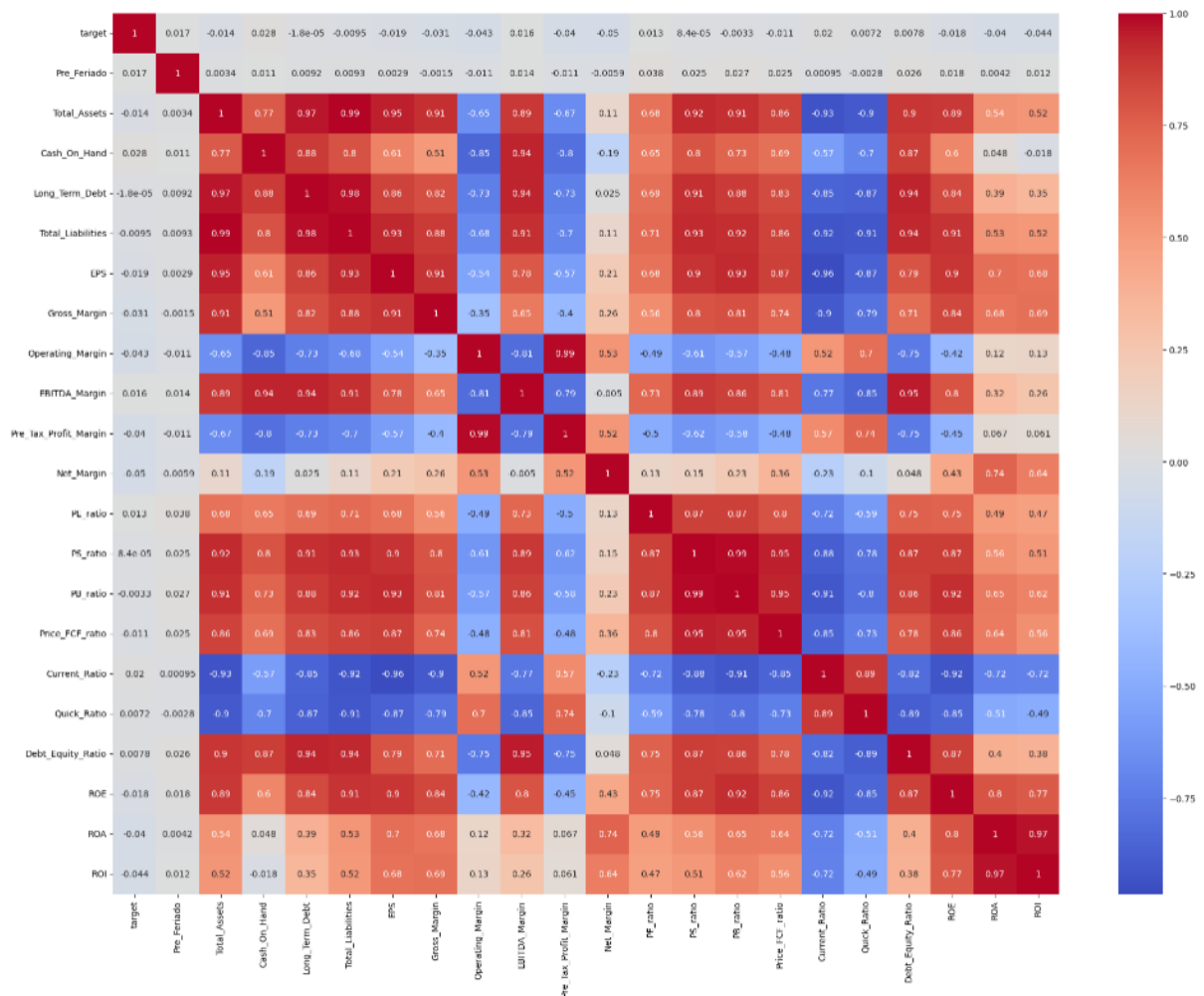


Anexo E: Matrizes de Correlações

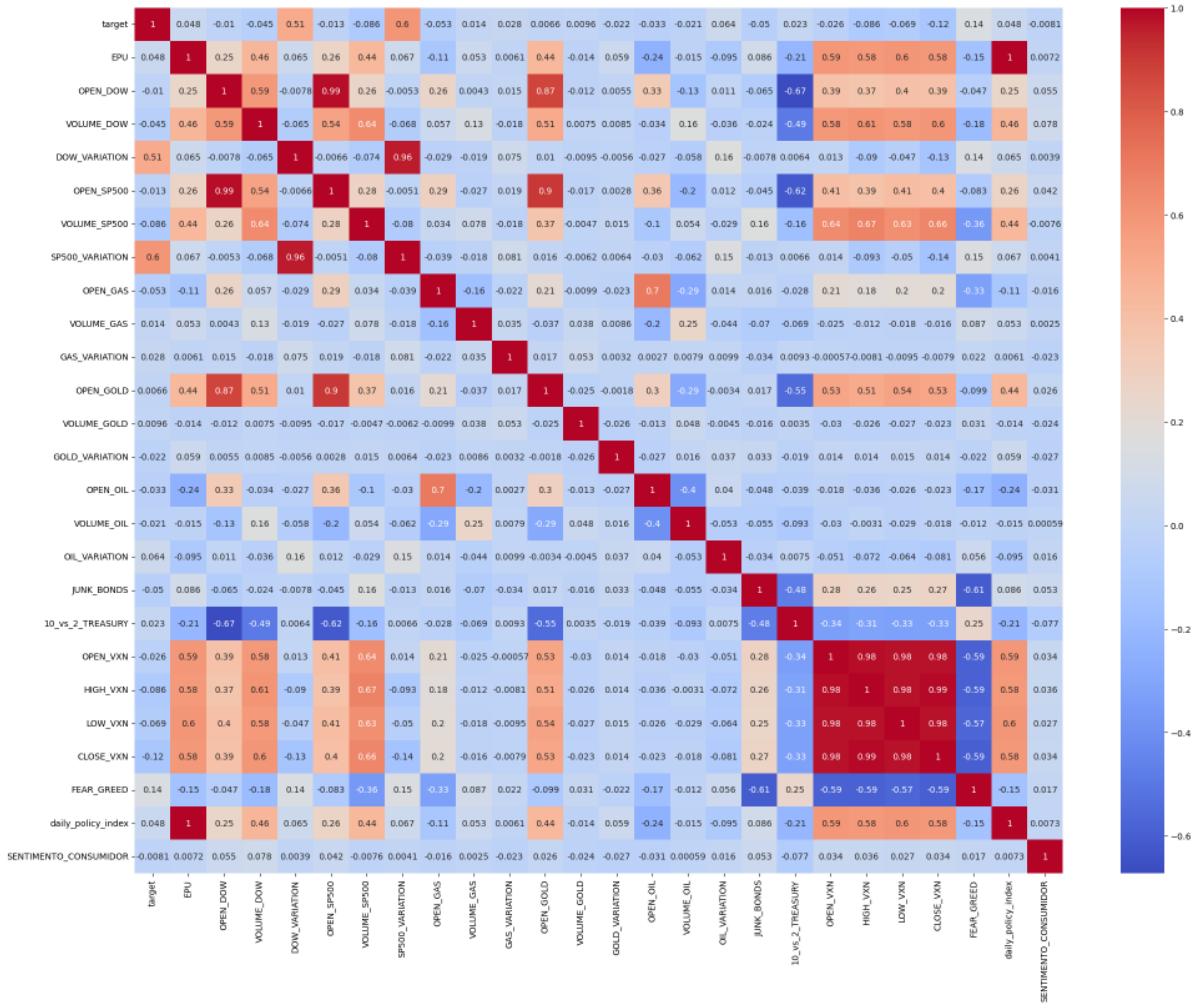
1. Indicadores técnicos



2. Indicadores fundamentais

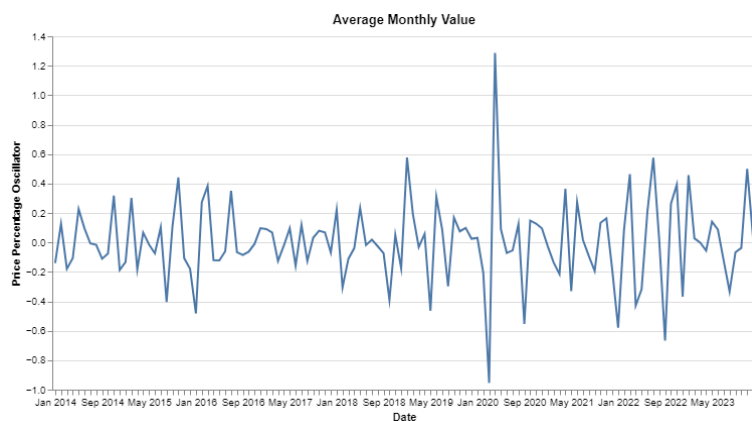


3. Indicadores de sentimiento

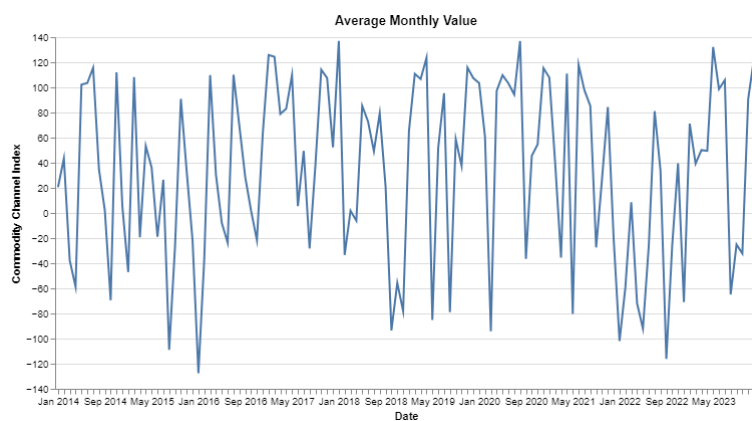


Anexo F: Variáveis utilizadas

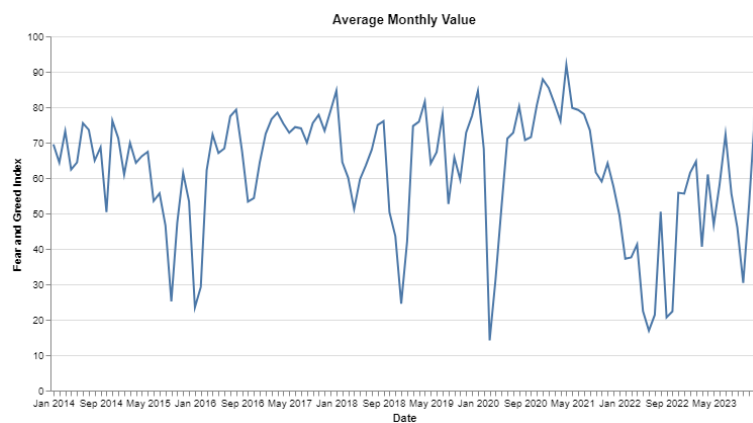
1. Price Percentage Oscillator



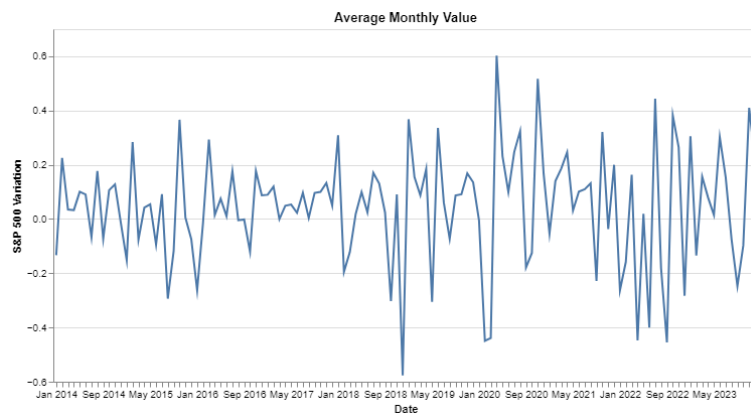
2. Commodity Channel Index



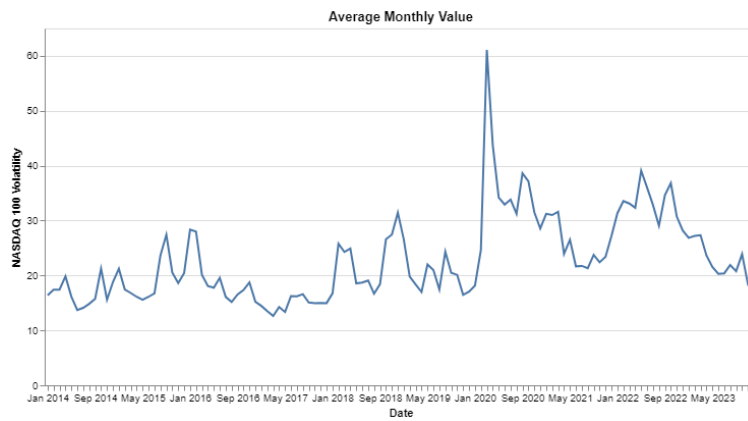
3. Fear and Greed Index



4. S&P 500 Variation



5. NASDAQ 100 Volatility



6. Junk Bonds

