

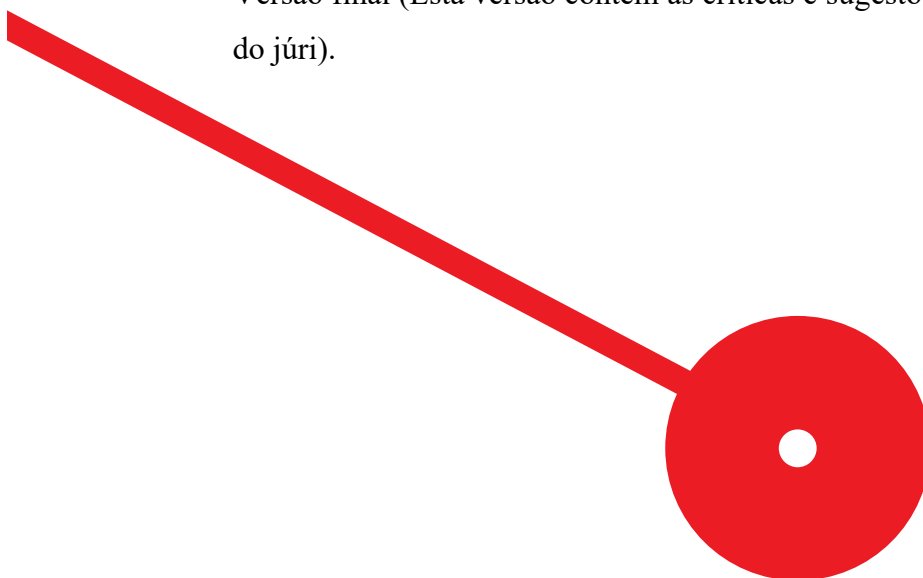


Análise de Sentimento no Contexto do Mercado Financeiro de Ações

Rui Alexandre de Granja Ribeiro

09/2025

Versão final (Esta versão contém as críticas e sugestões dos elementos do júri).

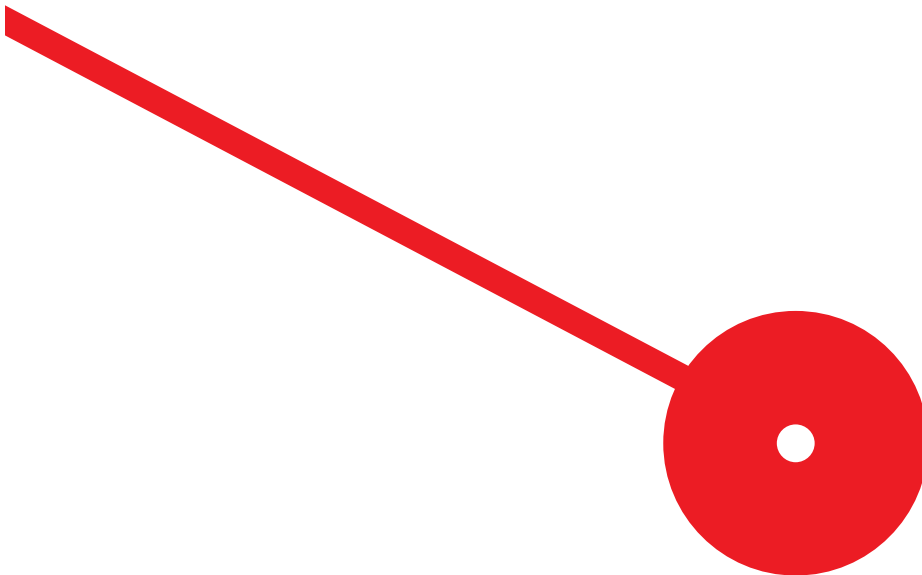




Análise de Sentimento no Contexto do Mercado Financeiro de Ações

Rui Alexandre de Granja Ribeiro

Dissertação de Mestrado apresentada ao Instituto Superior de Contabilidade e Administração do Porto para a obtenção do grau de Mestre em Inteligência e Análise para o Negócio, sob orientação de Professora Doutora Célia Talma Martins de Pinho Valente Oliveira Gonçalves e do Professor Doutor Henrique Daniel de Avelar Lopes Cardoso.



Dedicatória

Dedico este trabalho aos meus pais, Sofia e José, com a convicção de que, onde quer que estejam, se sentirão orgulhosos de mim.

Agradecimentos

A concretização deste trabalho só foi possível graças ao contributo de várias pessoas, a quem manifesto o meu profundo reconhecimento.

Em primeiro lugar, à minha esposa, Susana, companheira de vida, pelo apoio incondicional, pela motivação constante, por nunca me deixar desistir, mesmo nos momentos mais difíceis, e pela forma como enfrentou, ao meu lado, todas as dificuldades.

Ao meu filho Bruno, a minha maior inspiração, na esperança de que esta conquista, de alguma forma, possa também servir-lhe de exemplo para nunca desistir dos seus sonhos. Acredita sempre em ti!

Aos meus orientadores, Professora Doutora Célia Talma Gonçalves e Professor Doutor Henrique Cardoso, pela orientação, partilha de conhecimento e pela permanente disponibilidade.

Por fim, a todos os restantes docentes e colegas de curso, pelo espírito de partilha e amizade, que muito enriqueceram esta jornada.

A todos, o meu sincero agradecimento.

Bem hajam!

Resumo:

Esta dissertação tem como objetivo investigar a relação entre diferentes técnicas de Processamento de Linguagem Natural e o desempenho de diversos modelos de análise de sentimento no contexto do mercado financeiro de ações, utilizando comentários provenientes da rede social *StockTwits*. O estudo procura também estabelecer a relação entre o sentimento extraído a partir desses comentários e de notícias financeiras, disponibilizadas pelo *EODHD*, com a variação diária do valor de fecho das ações, obtido através do *Yahoo Finance*. Foram exploradas diferentes técnicas de pré-processamento e representações textuais, com e sem balanceamento de dados, comparando abordagens baseadas em léxicos, modelos de aprendizagem computacional tradicional, modelos baseados em redes neurais e modelos baseados em transformadores. A análise temporal incluiu a construção de índices de sentimento diário dos comentários e das notícias financeiras. Estes índices, bem como os valores de fecho das ações, foram organizados em séries temporais, que serviram de base para a detecção de desfasamentos através do algoritmo *Dynamic Time Warping* (Alinhamento Temporal Dinâmico) e para a verificação do alinhamento direcional com recurso a testes de significância estatística. Os resultados indicam que os modelos baseados em transformadores alcançaram o melhor desempenho, embora com maior exigência computacional, destacando-se o *RoBERTaStockTwits* como o mais eficaz. O impacto da técnica de balanceamento foi diferente entre as diferentes abordagens, sendo mais notório nos modelos baseados em transformadores e na aprendizagem computacional tradicional, e pouco significativo nos modelos baseados em redes neurais. A aplicação de técnicas de pré-processamento menos complexas produziu melhores resultados, com maior destaque nos modelos baseados em transformadores. As técnicas de representação textual baseadas em frequência de palavras mostraram melhor desempenho do que as representações densas nos modelos de aprendizagem computacional tradicional. A análise temporal revelou diferenças entre os comentários nas redes sociais e as notícias, com padrões mais evidentes e estatisticamente significativos nos comentários, sugerindo a possibilidade de ineficiências, tendo como referência a Teoria do Mercado Eficiente.

Palavras chave: Análise de Sentimento, Processamento de Linguagem Natural, Mercado Financeiro, Dynamic Time Warping

Abstract:

This dissertation aims to investigate the relationship between different Natural Language Processing techniques and the performance of various sentiment analysis models in the context of the stock market, using comments collected from the *StockTwits* social network. The study also seeks to establish the relationship between the sentiment extracted from these comments and from financial news, provided by *EODHD*, with the daily variation in stock closing prices, obtained through *Yahoo Finance*. Different preprocessing techniques and textual representations were explored, with and without data balancing, comparing approaches based on lexicons, traditional machine learning models, neural network-based models, and transformer-based models. The temporal analysis included the construction of daily sentiment indices from comments and financial news. These indices, together with stock closing prices, were organized into time series, which served as the basis for lag detecting through *Dynamic Time Warping* algorithm and for verifying directional alignment using statistical significance tests. The results indicate that transformer-based models achieved the best performance, albeit with higher computational requirements, with *RoBERTaStockTwits* standing out as the most effective. The impact of data balancing varied across the different approaches, being more pronounced in transformer-based and traditional machine learning models, and less significant in neural network models. The application of less complex preprocessing techniques led to better results, particularly in transformer-based models. Textual representation techniques based on word frequency outperformed dense representations in traditional machine learning models. The temporal analysis between sentiment indices and daily stock price variation revealed differences between social media comments and news, with more evident and statistically significant patterns in the comments, suggesting the possibility of inefficiencies, taking the Efficient Market Hypothesis as a reference.

Key words: Sentiment Analysis, Natural Language Processing, Financial Market, Dynamic Time Warping

Índice Geral

Capítulo I – Introdução.....	1
1.1 Objetivos e questões de investigação	3
1.2 Estrutura do documento.....	4
Capítulo II – Revisão da Literatura.....	6
2.1 Introdução à análise de sentimento.....	7
2.2 Métodos baseados em léxicos.....	10
2.3 Aprendizagem computacional tradicional.....	12
2.4 Aprendizagem profunda	14
2.5 Modelos baseados em transformadores.....	17
2.6 Técnicas de pré-processamento	19
2.7 Técnicas de balanceamento de dados	20
2.8 Técnicas de representação textual	20
2.9 Aplicação no contexto financeiro	21
2.10 Plataforma StockTwits	23
2.11 Teoria do Mercado Eficiente	25
2.12 Mercado NASDAQ	26
Capítulo III – Metodologia	29
3.1 Compreensão do negócio.....	30
3.2 Compreensão dos dados	30
3.2.1 Recolha inicial dos dados	31
3.2.2 Análise exploratória de dados.....	32
3.2.2.1 Comentários do <i>StockTwits</i>	32
3.2.2.2 Notícias financeiras do <i>EODHD</i>	34
3.2.2.3 Informação financeira das ações do <i>Yahoo Finance</i>	36
3.3 Preparação dos dados.....	36
3.3.1 Limpeza dos dados	36

3.3.2	Amostragem	37
3.3.3	Partição dos dados	38
3.3.4	Balanceamento dos dados.....	39
3.3.5	Técnicas de pré-processamento	40
3.4	Modelação	43
3.4.1	Analisadores baseados em léxicos.....	43
3.4.2	Modelos de aprendizagem computacional tradicional	44
3.4.3	Modelos baseados em redes neurais.....	48
3.4.4	Modelos baseados em transformadores.....	51
3.5	Avaliação	53
3.6	Implementação.....	55
3.6.1	Seleção do melhor modelo e classificação dos comentários.....	56
3.6.2	Construção das séries temporais de sentimento agregado.....	56
3.6.2.1	Agregação diária dos comentários.....	57
3.6.2.2	Índice de sentimento agregado dos comentários	57
3.6.2.3	Agregação diária das notícias	57
3.6.2.4	Índice de sentimento agregado das notícias	57
3.6.2.5	Construção da série de preços	58
3.6.3	Deteção do desfasamento ótimo	58
3.6.4	Alinhamento direcional em janelas deslizantes.....	59
3.6.5	Teste de significância estatística com permutações aleatórias	60
Capítulo IV – Análise e Discussão de Resultados		62
4.1	Comparação de desempenho dos classificadores	64
4.1.1	Analisadores baseados em léxicos.....	64
4.1.1.1	Análise de sensibilidade à técnica de pré-processamento	66
4.1.1.2	Melhores combinações	67
4.1.2	Modelos de aprendizagem computacional tradicional	67

4.1.2.1	Análise de sensibilidade à técnica de balanceamento	68
4.1.2.2	Análise de sensibilidade à técnica de pré-processamento	69
4.1.2.3	Análise de sensibilidade à técnica de representação textual.....	71
4.1.2.4	Melhores combinações	74
4.1.3	Modelos baseados em redes neuronais.....	74
4.1.3.1	Análise de sensibilidade à técnica de balanceamento	76
4.1.3.2	Análise de sensibilidade à técnica de pré-processamento	77
4.1.3.3	Análise de sensibilidade à técnica de representação textual.....	79
4.1.3.4	Melhores combinações	82
4.1.3.5	Análise ao desempenho computacional (utilização da GPU).....	82
4.1.4	Modelos baseados em transformadores.....	83
4.1.4.1	Análise de sensibilidade à técnica de balanceamento	84
4.1.4.2	Análise de sensibilidade à técnica de pré-processamento	85
4.1.4.3	Melhores combinações	87
4.1.4.4	Análise ao desempenho computacional (utilização da GPU).....	88
4.2	Análise de padrões temporais entre o sentimento e o comportamento do mercado	88
Capítulo V – Conclusões		92
5.1	Síntese dos resultados.....	93
5.2	Limitações do estudo	95
5.3	Trabalho futuro	95
Referências Bibliográficas		97

Índice de Figuras

Figura 1 - Visão geral do uso de redes sociais, baseada nos dados do DataReportal, referentes a outubro de 2024. Fonte: https://datareportal.com/global-digital-overview (acesso em 9 de janeiro de 2025).	8
Figura 2 - Evolução do interesse global pelo termo "Sentiment Analysis" de 2004 a 2024. Fonte: Google Trends (acesso em 9 de janeiro de 2025).	8
Figura 3 - Exemplo de mensagem. Fonte: StockTwits (acesso em 9 de janeiro de 2025).	23
Figura 4 - Adaptação do modelo CRISP-DM ao presente estudo. Fonte: Elaboração própria.....	30
Figura 5 - Distribuição do número de comentários por ação. Fonte: Elaboração própria a partir dos dados do StockTwits.	32
Figura 6 - Distribuição do número de comentários por ação e por ano. Fonte: Elaboração própria a partir dos dados do StockTwits.	32
Figura 7 - Distribuição do número de comentários por ação e por mês. Fonte: Elaboração própria a partir dos dados do StockTwits.	33
Figura 8 - Distribuição do sentimento nos comentários. Fonte: Elaboração própria a partir dos dados do StockTwits.	33
Figura 9 - Número de notícias por ação. Fonte: Elaboração própria a partir dos dados do EODHD.	34
Figura 10 - Número de notícias por ação por ano. Fonte: Elaboração própria a partir dos dados do EODHD.	34
Figura 11 - Número de notícias por ação por mês. Fonte: Elaboração própria a partir dos dados do EODHD.	35
Figura 12 - Distribuição do sentimento nas notícias. Fonte: Elaboração própria a partir dos dados do EODHD.	35
Figura 13 - Evolução do preço de fecho das ações. Fonte: Elaboração própria a partir dos dados do Yahoo Finance.	36
Figura 14 - Representação do fluxo de limpeza inicial dos dados. Fonte: Elaboração própria.....	37
Figura 15 - Representação do fluxo de preparação dos dados. Fonte: Elaboração própria.	43

Figura 16 - Representação do fluxo de processamento da categoria de analisadores baseados em léxicos. Fonte: Elaboração própria.....	44
Figura 17 - Representação do fluxo de processamento da categoria de modelos de aprendizagem computacional. Fonte: Elaboração própria.	48
Figura 18 - Representação do fluxo de processamento da categoria de modelos baseados em redes neuronais. Fonte: Elaboração própria.....	50
Figura 19 - Representação do fluxo de processamento da categoria de modelos baseados em transformadores. Fonte: Elaboração própria.....	53
Figura 20 - Exatidão máxima por analisador baseado em léxicos. Fonte: Elaboração própria.....	64
Figura 21 - Exatidão máxima por analisador baseado em léxicos e técnica de pré-processamento. Fonte: Elaboração própria.....	65
Figura 22 - Distribuição da exatidão dos analisadores baseados em léxicos, nas diferentes execuções. Fonte: Elaboração própria.	65
Figura 23 - Variação da exatidão dos analisadores baseados em léxicos por técnica de pré-processamento. Fonte: Elaboração própria.	66
Figura 24 - Exatidão máxima por modelo de aprendizagem computacional tradicional, com e sem aplicação de técnicas de balanceamento. Fonte: Elaboração própria.....	67
Figura 25 - Exatidão máxima por modelo de aprendizagem computacional tradicional e técnica de pré-processamento, com e sem balanceamento. Fonte: Elaboração própria.	68
Figura 26 - Distribuição da exatidão dos modelos de aprendizagem computacional tradicional, nas diferentes execuções, com e sem balanceamento. Fonte: Elaboração própria.....	68
Figura 27 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, com e sem balanceamento dos dados. Fonte: Elaboração própria.	69
Figura 28 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre combinações de técnicas de pré-processamento, sem balanceamento dos dados. Fonte: Elaboração própria.	70
Figura 29 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre combinações de técnicas de pré-processamento, com balanceamento dos dados. Fonte: Elaboração própria.....	71
Figura 30 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre diferentes representações textuais, sem balanceamento dos dados. Fonte: Elaboração própria.	72

Figura 31 - Exatidão máxima dos modelos de aprendizagem computacional tradicional para cada representação textual (sem balanceamento dos dados). Fonte: Elaboração própria.....	72
Figura 32 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre diferentes representações textuais, com balanceamento dos dados. Fonte: Elaboração própria.	73
Figura 33 - Exatidão máxima dos modelos de aprendizagem computacional tradicional para cada representação textual (com balanceamento dos dados). Fonte: Elaboração própria.....	73
Figura 34 - Exatidão máxima por modelo baseado em redes neuronais, com e sem aplicação de técnicas de balanceamento. Fonte: Elaboração própria.....	75
Figura 35 - Exatidão máxima por modelo baseado em redes neuronais e técnica de pré-processamento, com e sem balanceamento. Fonte: Elaboração própria.....	75
Figura 36 - Distribuição da exatidão dos modelos baseados em redes neuronais, nas diferentes execuções, com e sem balanceamento. Fonte: Elaboração própria.	76
Figura 37 - Variação da exatidão máxima dos modelos baseados em redes neuronais, com e sem balanceamento dos dados. Fonte: Elaboração própria.	77
Figura 38 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre combinações de técnicas de pré-processamento, sem balanceamento dos dados. Fonte: Elaboração própria.....	78
Figura 39 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre combinações de técnicas de pré-processamento, com balanceamento dos dados. Fonte: Elaboração própria.....	79
Figura 40 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre diferentes representações textuais, sem balanceamento dos dados. Fonte: Elaboração própria.....	80
Figura 41 - Exatidão máxima dos modelos baseados em redes neuronais para cada representação textual (sem balanceamento dos dados). Fonte: Elaboração própria.....	80
Figura 42 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre diferentes representações textuais, com balanceamento dos dados. Fonte: Elaboração própria.....	81
Figura 43 - Exatidão máxima dos modelos baseados em redes neuronais para cada representação textual (com balanceamento dos dados). Fonte: Elaboração própria.	81

Figura 44 - Exatidão máxima por modelo baseado em transformadores, com e sem aplicação de técnicas de balanceamento. Fonte: Elaboração própria.....	83
Figura 45 - Exatidão máxima por modelo baseado em transformadores e técnica de pré-processamento, com e sem balanceamento. Fonte: Elaboração própria.....	83
Figura 46 - Distribuição da exatidão dos modelos baseados em transformadores, nas diferentes execuções, com e sem balanceamento. Fonte: Elaboração própria.	84
Figura 47 - Variação da exatidão máxima dos modelos baseados em transformadores, com e sem balanceamento dos dados. Fonte: Elaboração própria.	85
Figura 48 - Variação da exatidão máxima dos modelos baseados em transformadores, entre combinações de técnicas de pré-processamento, sem balanceamento dos dados. Fonte: Elaboração própria.	86
Figura 49 - Variação da exatidão máxima dos modelos baseados em transformadores, entre combinações de técnicas de pré-processamento, com balanceamento dos dados. Fonte: Elaboração própria.	87

Índice de Tabelas

Tabela 1 - Níveis de análise de sentimentos. Fonte: Elaboração própria com base no estudo de (Shivaprasad & Shetty, 2017).	9
Tabela 2 - Categorias de polaridade. Fonte: Elaboração própria com base nos estudos de (Pang & Lee, 2005; Turney, 2002; Wilson et al., 2005).	9
Tabela 3 - Resumo de estudos com abordagem lexical na análise de sentimentos. Fonte: Elaboração própria.....	11
Tabela 4 - Resumo de estudos com aplicação de aprendizagem computacional tradicional na análise de sentimentos. Fonte: Elaboração própria.....	13
Tabela 5 - Resumo de estudos com aplicação de aprendizagem profunda na análise de sentimentos. Fonte: Elaboração própria.	16
Tabela 6 - Resumo de estudos comparativos sobre modelos baseados em transformadores aplicados à análise de sentimentos. Fonte: Elaboração própria.....	18
Tabela 7 - Estudos de abordagens específicas de análise de sentimentos adaptados ao domínio financeiro. Fonte: Elaboração própria.	22
Tabela 8 - Estudos que utilizam dados do StockTwits na análise de sentimentos e previsão do mercado financeiro. Fonte: Elaboração própria.	24
Tabela 9 - As 10 maiores bolsas de valores do mundo por capitalização bolsista, incluindo o número de empresas listada. Fonte: Elaboração própria. com base nos dados da World Federation of Exchanges de maio de 2025 (acesso em 24 de Agosto de 2025).....	27
Tabela 10 - Dados financeiros das ações das empresas FAANG mais Microsoft, incluindo capitalização de mercado, variação das ações (52 semanas) e P/E Ratio. Fonte: Elaboração própria com base nos dados do site Nasdaq.com, (acesso em 1 de janeiro de 2025).....	28
Tabela 11 - Comparação dos principais indicadores antes e depois da limpeza dos dados. Fonte: Elaboração própria.	37
Tabela 12 - Comparação dos dados após a limpeza inicial e após a amostragem aleatória. Fonte: Elaboração própria.	38
Tabela 13 - Estatísticas dos subconjuntos gerados após a partição dos dados. Fonte: Elaboração própria.....	39
Tabela 14 - Comparação após aplicação das técnicas de balanceamento no conjunto de treino. Fonte: Elaboração própria.	40

Tabela 15 - Exemplo do efeito da técnica de conversão para minúsculas (LOW). Fonte: Elaboração própria.....	41
Tabela 16 - Exemplo do efeito da técnica de eliminação de palavras sem significado semântico (RSW). Fonte: Elaboração própria.....	41
Tabela 17 - Exemplo do efeito da técnica de radicalização (STM). Fonte: Elaboração própria.....	41
Tabela 18 - Exemplo do efeito combinado das técnicas eliminação de palavras com pouco significado semântico (RSW) e radicalização (STM). Fonte: Elaboração própria.	42
Tabela 19 - Exemplo do efeito combinado das técnicas radicalização (STM) e remoção de palavras com pouco significado semântico (RSW). Fonte: Elaboração própria.	42
Tabela 20 - Exemplo do efeito combinado das técnicas de conversão para minúsculas (LOW) e remoção de palavras com pouco significado semântico (RSW) e radicalização (STM). Fonte: Elaboração própria.	42
Tabela 21 - Resumo das ferramentas de análise de sentimento baseadas em léxicos. Fonte: Elaboração própria.....	44
Tabela 22 - Resumo dos algoritmos de aprendizagem computacional tradicional. Fonte: Elaboração própria.....	45
Tabela 23 - Resumo das técnicas de representação textual. Fonte: Elaboração própria.	46
Tabela 24 - Parâmetros utilizados no treino dos modelos de representação densa. Fonte: Elaboração própria.....	47
Tabela 25 - Hiperparâmetros utilizados nos classificadores de aprendizagem computacional. Fonte: Elaboração própria.	47
Tabela 26 - Parâmetros utilizados nos vetorizadores textuais. Fonte: Elaboração própria.	48
Tabela 27 - Resumo das arquiteturas de redes neuronais utilizadas. Fonte: Elaboração própria.....	49
Tabela 28 - Hiperparâmetros dos modelos baseados em redes neuronais. Fonte: Elaboração própria.....	50
Tabela 29 - Resumo dos modelos baseados em transformadores utilizados. Fonte: Elaboração própria.....	51
Tabela 30 - Resumo dos parâmetros dos modelos baseados em transformadores. Fonte: Elaboração própria a partir dos dados da plataforma Hugging Face	52
Tabela 31 - Parâmetros de ajuste fino dos modelos baseados em transformadores. Fonte: Elaboração própria.....	53

Tabela 32 - Resumo das métricas de avaliação utilizadas. Fonte: Elaboração própria. .	54
Tabela 33 - Número de Execuções e Tempo Total de Execução para Analisadores Lexicais. Fonte: Elaboração própria.	63
Tabela 34 - Número de Execuções e Tempo Total de Execução para Modelos de Aprendizagem Computacional Tradicional. Fonte: Elaboração própria.	63
Tabela 35 - Número de Execuções e Tempo Total de Execução para Modelos Baseados em Redes Neurais. Fonte: Elaboração própria.	63
Tabela 36 - Número de Execuções e Tempo Total de Execução para Modelos Baseados em Transformadores. Fonte: Elaboração própria.	63
Tabela 37 - Variação da exatidão dos analisadores baseados em léxicos entre a melhor e a pior técnica de pré-processamento. Fonte: Elaboração própria.	66
Tabela 38 - Comparação das melhores combinações de pré-processamento e desempenho para cada analisador baseado em léxicos. Fonte: Elaboração própria.	67
Tabela 39 - Comparação da exatidão máxima dos modelos de aprendizagem computacional tradicional, com e sem balanceamento dos dados, incluindo a variação observada. Fonte: Elaboração própria.	69
Tabela 40 - Comparação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (sem balanceamento dos dados). Fonte: Elaboração própria.	70
Tabela 41 - Comparação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (com balanceamento dos dados). Fonte: Elaboração própria.	70
Tabela 42 - Exatidão dos modelos de aprendizagem computacional tradicional para cada representação textual, incluindo a melhor e a pior exatidão máxima registradas e a respectiva variação (sem balanceamento dos dados). Fonte: Elaboração própria.	72
Tabela 43 - Exatidão dos modelos de aprendizagem computacional tradicional para cada representação textual, incluindo a melhor e a pior exatidão registradas e a respectiva variação (com balanceamento de dados). Fonte: Elaboração própria.	73
Tabela 44 - Comparação das melhores combinações de pré-processamento, balanceamento, representação textual e desempenho para cada modelo de aprendizagem computacional. Fonte: Elaboração própria.	74

Tabela 45 - Comparação da exatidão máxima dos modelos baseados em redes neurais, com e sem balanceamento dos dados, incluindo a variação observada. Fonte: Elaboração própria.....	76
Tabela 46- Comparação da exatidão máxima dos modelos baseados em redes neurais, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (sem balanceamento dos dados). Fonte: Elaboração própria.....	77
Tabela 47 - Comparação da exatidão máxima dos modelos baseados em redes neurais, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (com balanceamento dos dados). Fonte: Elaboração própria.	78
Tabela 48 - Exatidão dos modelos baseados em redes neurais para cada representação textual, incluindo a melhor e a pior exatidão máxima registadas e a respetiva variação (sem balanceamento dos dados). Fonte: Elaboração própria.	79
Tabela 49 - Exatidão dos modelos baseados em redes neurais para cada representação textual, incluindo a melhor e a pior exatidão máxima registadas e a respetiva variação (com balanceamento dos dados). Fonte: Elaboração própria.	80
Tabela 50 - Comparação das melhores combinações de pré-processamento, balanceamento, representação textual e desempenho para cada modelo baseado em redes neurais. Fonte: Elaboração própria.	82
Tabela 51 - Utilização média de GPU (em MiB) durante o treino dos modelos baseados em redes neurais. Fonte: Elaboração própria.....	82
Tabela 52 - Comparação da exatidão máxima dos modelos baseados em transformadores, com e sem balanceamento dos dados, incluindo a variação observada. Fonte: Elaboração própria.....	84
Tabela 53 - Comparação da exatidão máxima dos modelos baseados em transformadores, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (sem balanceamento dos dados). Fonte: Elaboração própria.....	85
Tabela 54 - Comparação da exatidão máxima dos modelos baseados em transformadores, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (com balanceamento dos dados).	86
Tabela 55 - Comparação das melhores combinações de pré-processamento, balanceamento e desempenho para cada modelo baseado em transformadores. Fonte: Elaboração própria.....	87
Tabela 56 - Utilização média de GPU (em MiB) durante o treino dos modelos baseados em transformadores. Fonte: Elaboração própria.....	88

Tabela 57 - Distância DTW média para o sentimento dos comentários e respectivos desfasamentos ótimos por ação. Fonte: Elaboração própria.....	89
Tabela 58 - Distância DTW média para o sentimento das notícias e respectivos desfasamentos ótimos por ação. Fonte: Elaboração própria.....	89
Tabela 59 - Resultados do teste de alinhamento direcional com permutação para cada ação (comentários StockTwits). Fonte: Elaboração própria.....	90
Tabela 60 - Resultados do teste de alinhamento direcional com permutação para cada ação (notícias financeiras). Fonte: Elaboração própria.	91

Lista de Abreviaturas

API – *Application Programming Interface (Interface de Programação de Aplicações)*

BERT – *Bidirectional Encoder Representations from Transformers*

BoW – *Bag of Words*

CBOW – *Continuous Bag of Words*

CNN – *Convolutional Neural Network*

CRISP-DM – *Cross-Industry Standard Process for Data Mining*

EMH – *Efficient Market Hypothesis (Teoria do Mercado Eficiente)*

FNN – *Feedforward Neural Network*

GPU – *Graphics Processing Unit (Unidade de Processamento Gráfico)*

GRU – *Gated Recurrent Unit*

IoT – *Internet of Things (Internet das Coisas)*

KNN – *K-Nearest Neighbor*

MLP – *Multi Layer Perceptron*

NASDAQ – *National Association of Securities Dealers Automated Quotations*

NLP – *Natural Language Processing (Processamento de Linguagem Natural)*

LSTM – *Long Short-Term Memory*

RNN – *Recurrent Neural Network*

RoBERTa – *Robustly Optimized BERT Pretraining Approach*

SVM – *Support Vector Machines*

TF-IDF – *Term Frequency – Inverse Document Frequency*

VRAM – *Video Random Access Memory*

CAPÍTULO I – INTRODUÇÃO

A cada segundo é gerada uma enorme quantidade de dados, proveniente de diversas origens. Entre estas fontes estão as tecnologias associadas à Internet das Coisas (*IoT*), as operações bancárias e de bolsa, bem como a atividade registada nas redes sociais (Noori & Maktoubian, 2019). Existe um grande esforço para organizar melhor esta informação para os utilizadores. Os investigadores têm vindo a explorar ativamente o desafio da classificação automática de texto (Pang et al., 2002).

A Análise de Sentimento, uma subárea do Processamento de Linguagem Natural (*NLP*) (Birjali et al., 2021; Sharma & Goyal, 2023), dedica-se à identificação de opiniões, emoções, atitudes e sentimentos expressos pelas pessoas, em texto (Birjali et al., 2021; Sharma & Goyal, 2023). Estes podem referir-se a produtos, serviços, eventos, organizações e pessoas (Aguilar-Moreno et al., 2024; B. Liu, 2012).

O impacto que as emoções dos investidores têm no mercado de ações é uma área com interesse científico crescente na comunidade académica (Goodell et al., 2023), conforme demonstrado em diversos estudos (Agarwal & Tewari, 2025; Goodell et al., 2023; J. Lee et al., 2023; Vamossy, 2021, 2024). As plataformas de redes sociais têm uma grande importância para os investidores (C.-C. Lee et al., 2021). A informação extraída de grandes volumes de texto pode servir para prever retornos, estimar a volatilidade e apoiar estratégias de investimento e de gestão de risco (Deveikyte et al., 2022). Estes dados não estruturados são difíceis de processar manualmente, mas fornecem informações valiosas (Aguilar-Moreno et al., 2024).

A pesquisa realizada por Alshahrani et al. (2018) revelou que os sentimentos expressos em *tweets* estão diretamente relacionados com as flutuações nos preços de fecho e com as oscilações do mercado. Prever os preços das ações é crucial para a organização das atividades empresariais, e como consequência tem ocorrido uma procura crescente por soluções e sistemas que possibilitem antecipar, com o máximo de precisão possível, os acontecimentos do mercado de ações (Bozanta et al., 2021).

As redes sociais como *Twitter* e *StockTwits* permitem a recolha de dados que refletem as emoções manifestadas pelos utilizadores (Nousi & Tjortjis, 2021). Em comparação com o *Twitter*, o *StockTwits* é uma fonte mais informativa para dados financeiros, uma vez que é uma plataforma dedicada exclusivamente ao mercado de ações (Oliveira et al., 2013). O *StockTwits* serve como um meio privilegiado para investidores partilharem ideias e trocarem informações (Oliveira et al., 2013).

A Teoria do Mercado Eficiente (*EMH*), formulada por Fama (1970), defende que o valor dos ativos reflete, a cada instante, todas as informações disponíveis (Komariah et al., 2016). No entanto, estudos recentes desafiam esta teoria. O estudo de M. Mokhtari et al. (2023) identifica uma ligação entre o sentimento expresso no *Twitter* e os preços das ações e conclui que esses fatores podem influenciar o desempenho dos ativos. O estudo de Ranco et al. (2015) demonstra que poderão existir momentos em que se observa uma dependência entre o sentimento expresso no *Twitter* e o retorno das ações, principalmente durante picos de atividade. O estudo de Bollen et al. (2011) conclui que o humor coletivo medido no *Twitter* pode prever a variação diária do índice *Dow Jones*.

As primeiras abordagens à análise de sentimento financeiro baseavam-se em léxicos construídos a partir de dicionários de palavras ponderadas (associadas a valores numéricos que indicam intensidade e polaridade), utilizados para atribuir polaridade positiva ou negativa a palavras ou frases (Man et al., 2019). Com o tempo, foram sendo substituídos por modelos de aprendizagem computacional que exploravam a frequência de palavras para extrair características relevantes do texto. Mesmo sendo um avanço, revelaram limitações, pois falhavam em captar o contexto presente em determinadas sequências de palavras (Shen & Zhang, 2024).

Com os recentes avanços, as técnicas de aprendizagem profunda começaram a ocupar uma posição de destaque, tendo superado os modelos de aprendizagem computacional tradicional, embora impliquem um tempo de execução mais elevado (Ahmad & Umar, 2023).

1.1 Objetivos e questões de investigação

O objetivo principal desta dissertação é investigar a relação entre diferentes técnicas de processamento de linguagem natural e o desempenho de vários modelos de análise de sentimento no contexto do mercado financeiro de ações, recorrendo a comentários provenientes de redes sociais relacionados com este domínio. O estudo procura também estabelecer a relação entre o sentimento obtido a partir desses comentários e de notícias financeiras com a variação diária do valor das ações. Com isto pretende-se avaliar até que ponto poderão existir indícios de ineficiências, tendo em consideração a Teoria do Mercado Eficiente.

Além do objetivo principal, estabelecem-se os seguintes objetivos específicos:

- Avaliar o impacto da aplicação de diferentes técnicas de pré-processamento de dados no desempenho da análise de sentimento em dados financeiros;
- Avaliar o impacto do balanceamento de dados no desempenho da análise de sentimento em dados financeiros;
- Avaliar o impacto da aplicação de diferentes métodos de representação textual no desempenho da análise de sentimento;
- Avaliar o custo computacional das diferentes abordagens e modelos, tendo em conta a eficiência de cada técnica;
- Comparar os resultados das abordagens combinadas dos diferentes modelos para identificar as melhores combinações;
- Analisar a relação entre o sentimento gerado nos comentários e nas notícias com a variação diária do valor das ações, explorando padrões temporais.

Tendo em conta os objetivos definidos, as questões de investigação que orientam este trabalho podem ser formuladas da seguinte forma:

- Qual é a combinação mais eficaz de modelos e técnicas de processamento de linguagem natural para otimizar o desempenho da análise de sentimento no contexto do mercado financeiro de ações?
- De que forma os sentimentos gerados a partir dos comentários e das notícias se correlacionam com a variação diária do valor das ações?

1.2 Estrutura do documento

Em termos de estrutura, este trabalho encontra-se organizado em quatro capítulos adicionais, nomeadamente:

- **Capítulo II – Revisão de Literatura:** Apresenta a análise da literatura científica mais relevante, para identificar trabalhos semelhantes e lacunas, que permitam sustentar o enquadramento teórico e metodológico.
- **Capítulo III – Metodologia:** Descreve a abordagem adotada na realização do trabalho empírico, segundo os princípios da metodologia *CRISP-DM*, alinhada com os objetivos e questões de investigação.

- **Capítulo IV – Discussão de Resultados:** Neste capítulo são apresentados os resultados obtidos na modelação e implementação dos modelos, definidos no capítulo anterior.
- **Capítulo V – Conclusões:** Este capítulo apresenta as conclusões mais relevantes identificados no estudo. Também se abordam as principais limitações que condicionaram o trabalho, sendo igualmente apresentadas algumas possibilidades de implementação para estudos futuros.

CAPÍTULO II – REVISÃO DA LITERATURA

A revisão da literatura permite ao investigador obter uma visão do conhecimento existente para enquadrar o seu estudo e seguir um percurso fundamentado. Segundo Snyder (2019), este processo implica recolher e analisar trabalhos já publicados sobre um determinado tema, de modo a construir uma perspetiva completa. Para Tranfield et al. (2003), esta análise permite identificar lacunas e as áreas já exploradas. A partir desta pesquisa surgem novas questões, linhas de investigação e oportunidades para o avanço do tema.

Esta revisão da literatura foi efetuada em várias etapas. Foram utilizados quatro repositórios científicos: *Web of Science*, *IEEE Xplore*, *ScienceDirect* e *arXiv*. A escolha destas fontes teve em consideração a sua abrangência na cobertura das áreas para este estudo. Foram utilizados termos relacionados com a análise de sentimentos e o processamento de linguagem natural, incluindo a sua aplicação ao mercado financeiro.

Para aumentar a abrangência, foram utilizadas expressões associadas tanto a técnicas tradicionais como a métodos mais modernos, entre elas '*machine learning*', '*deep learning*' e '*transformers*'. Foram também utilizados termos específicos do domínio financeiro, como '*stock market*' e '*stocktwits*'.

A pesquisa foi efetuada em títulos, resumos e palavras-chave. Foram utilizados os operadores lógicos '*AND*' e '*OR*' para combinar termos com o objetivo de aumentar o máximo de resultados relevantes. Para assegurar atualidade das informações foi dada prioridade a publicações recentes.

Após a pesquisa inicial, selecionaram-se artigos de referência, recentes e com grande número de citações. A partir dessa seleção inicial, realizou-se uma análise mais detalhada aos resumos, metodologias utilizadas, conclusões e respetivas referências bibliográficas. Este processo permitiu identificar novas fontes de informação consideradas relevantes. Com estas novas fontes, a revisão ganhou mais profundidade, garantindo que a cobertura do tema permanecesse atualizada e abrangente para sustentar o enquadramento teórico.

2.1 Introdução à análise de sentimento

Segundo o estudo da *DataReportal*, apresentado na Figura 1, em outubro de 2024 mais de 5,22 biliões de pessoas (63,8% da população mundial) utilizam as redes sociais em todo o mundo, com um tempo médio diário superior a duas horas. A Internet, enquanto

sistema global, liga bilhões de dispositivos e indivíduos, garantindo esta expansão contínua de conteúdo digital (Sohangir, Wang, et al., 2018).

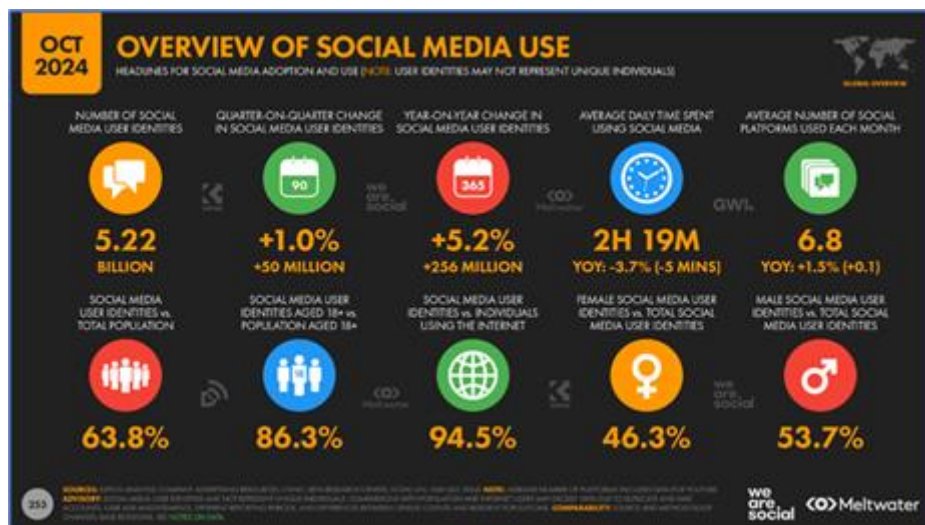


Figura 1 - Visão geral do uso de redes sociais, baseada nos dados do DataReportal, referentes a outubro de 2024. Fonte: <https://datareportal.com/global-digital-overview> (acesso em 9 de janeiro de 2025).

Este crescimento da tecnologia, que impulsionou a ideia predominante de 'Conteúdo Gerado pelos Utilizadores' (Mohammad, 2018; Nanli et al., 2012), em conjunto com o aumento da comunicação online (através de e-mails, sites, fóruns e chats) (Park & Woo, 2019), confere uma importância crescente (Mäntylä et al., 2018; Pang & Lee, 2008) à análise de sentimentos.

Segundo Mäntylä et al. (2018), até 2016 tinham sido publicados quase 7.000 artigos sobre este tema, dos quais 99% após 2004. A sua importância e interesse crescente é tão grande que várias empresas oferecem serviços de extração de opiniões (B. Liu & Zhang, 2012).

A Figura 2 mostra o interesse global pelo termo 'Sentiment Analysis' entre 2004 e 2024.

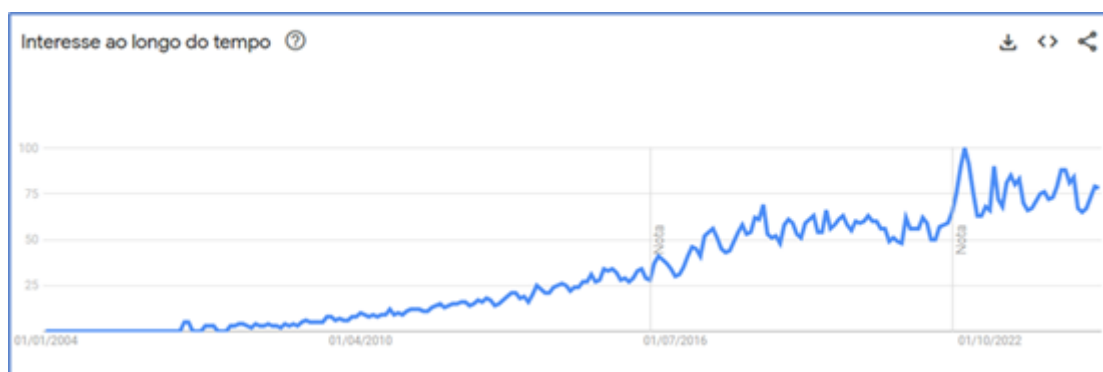


Figura 2 - Evolução do interesse global pelo termo "Sentiment Analysis" de 2004 a 2024. Fonte: Google Trends (acesso em 9 de janeiro de 2025).

De acordo com o estudo de Shivaprasad & Shetty (2017), a análise de sentimentos pode ser realizada em três níveis distintos, conforme resumido na Tabela 1.

Tabela 1 - Níveis de análise de sentimentos. Fonte: Elaboração própria com base no estudo de (Shivaprasad & Shetty, 2017).

Nível	Descrição
Documento	Avalia o documento como um todo, classificando o sentimento global (positivo ou negativo). Adequado quando o texto aborda uma única entidade.
Frase	Analisa cada frase individualmente, distinguindo informação subjetiva de objetiva e atribuindo polaridade (positivo, negativo ou neutro). Assume que cada frase contém apenas uma opinião.
Entidade e Aspeto	Identifica sentimentos associados a entidades específicas e aos seus aspetos ou características. Permite capturar polaridades distintas para diferentes aspetos de uma mesma entidade.

A polaridade associada à opinião pode ser classificada em diferentes categorias, conforme resumido na Tabela 2.

Tabela 2 - Categorias de polaridade. Fonte: Elaboração própria com base nos estudos de (Pang & Lee, 2005; Turney, 2002; Wilson et al., 2005).

Categoria	Descrição
Binária	Classifica o sentimento em duas categorias: positiva ou negativa. (Pang & Lee, 2005; Turney, 2002).
Multiclasse	Também designada por classificação por escalas, distribui os sentimentos em diferentes níveis (por ex. 1 a 5 estrelas) (Pang & Lee, 2005).
Mista	Identifica sentimentos positivos, negativos, neutros ou ambas as polaridades na mesma expressão, capturando ambiguidades e variações contextuais. A etiqueta <i>mista</i> é atribuída a expressões com polaridade positiva e negativa em simultâneo (por ex: “ <i>Além disso, os políticos referem-se ao bem e ao mal apenas para fins de intimidação e exagero</i> ”) (Wilson et al., 2005).

Têm sido realizados vários estudos no domínio da análise de sentimentos que abordam diferentes métodos, técnicas e domínios de aplicação, como por exemplo a deteção de fenómenos linguísticos complexos, incluindo concordância e discordância, sarcasmo, e análise de opiniões expressas em diferentes domínios, como redes sociais, notícias,

avaliações de produtos e críticas de filmes (González-Ibáñez et al., 2011; Hillard et al., 2003; Pang & Lee, 2004, 2005; Turney, 2002; Wilson et al., 2005).

Nesta investigação, a análise de sentimento foi abordada em quatro grandes categorias de metodologias: métodos baseados em léxicos, aprendizagem computacional tradicional, aprendizagem profunda e modelos baseados em transformadores. Apesar de Mao et al. (2024) apresentarem uma classificação ligeiramente diferente (incluindo abordagens híbridas), a sua taxonomia forneceu o enquadramento geral que fundamenta esta divisão. As secções seguintes descrevem em detalhe cada uma destas abordagens.

2.2 Métodos baseados em léxicos

A polaridade de um texto pode ser determinada a partir da orientação semântica das palavras ou expressões que o compõem. Esta técnica utiliza dicionários de termos positivos e negativos. São construídos manualmente ou de forma automática, com valores de sentimento atribuídos a cada termo (Jurek et al., 2015; Rizinski et al., 2024; Taboada et al., 2011).

Alguns léxicos são amplamente utilizados, como são o caso de *SentiWordNet* (Sebastiani et al., 2010), o *AFINN* (Nielsen, 2011), o *Loughran-McDonald Financial Lexicon* (Loughran & McDonald, 2011), e o *VADER* (C. Hutto & Gilbert, 2014).

O *SentiWordNet* 3.0 é uma versão melhorada do *SentiWordNet* apresentada por Esuli & Sebastiani (2006), que por sua vez é uma extensão do *WordNet*, resultante da anotação automática de todos os *synsets* (conjuntos de sinónimos com o mesmo significado) (Sebastiani et al., 2010). O *WordNet* reúne mais de 118.000 palavras distintas e mais de 90.000 sinónimos. É utilizado para atribuir uma pontuação de sentimento aos termos subjetivos presentes no texto. Cada *synset* tem um valor de positividade e negatividade. A pontuação de cada termo resulta da diferença entre estes dois valores e, para o texto, as polaridades individuais são combinadas de forma a obter a pontuação global de sentimento do documento (Taj et al., 2019).

O *AFINN* foi pensado para textos curtos e informais. A primeira versão surgiu em 2009 e inclui milhares de palavras e expressões classificadas com valores de valência que variam de -5 (muito negativo) a +5 (muito positivo). Este vocabulário foi construído

manualmente, recorrendo a listas de palavras afetivas pré-existentes e a dicionários online. O objetivo foi captar melhor a linguagem típica de *microblogs* (Nielsen, 2011).

O *Loughran–McDonald Financial Lexicon* é um léxico de análise de sentimento desenvolvido especificamente para documentos financeiros. É amplamente utilizado em estudos de finanças. Inclui milhares de termos organizados em seis categorias (positivo, negativo, incerteza, litigioso, restritivo e supérfluo). Foi desenvolvido para ultrapassar as limitações de léxicos mais genéricos, que por vezes interpretam de forma incorreta termos comuns neste domínio (Loughran & McDonald, 2011).

O *VADER (Valence Aware Dictionary and Sentiment Reasoner)* atribui um valor de sentimento a cada texto. Trata-se de um léxico baseado em regras, desenvolvido a partir de publicações em redes sociais. Pode ser aplicado para identificar o sentimento em dados não rotulados, que depois podem ser classificados por métodos de aprendizagem computacional (C. J. Hutto & Gilbert, 2014; Marutho et al., 2022).

A literatura apresenta diversos estudos que utilizaram os léxicos para a análise de sentimentos. Destacam-se os trabalhos resumidos na Tabela 3.

Tabela 3 - Resumo de estudos com abordagem lexical na análise de sentimentos. Fonte: Elaboração própria.

Autores/Ano	Objetivo	Dados	Resultados
Ray & Chakrabarti (2017)	Analisar sentimentos de utilizadores do <i>Twitter</i> sobre produtos usando abordagem lexical.	<i>Tweets</i> sobre <i>'iPhone'</i>	Análise a nível de documento, frase e aspeto; identificação de opiniões positivas, negativas e neutras para diferentes características do produto (ex.: bateria, preço, qualidade da câmara).
Sohangir, Petty, et al. (2018)	Avaliar a eficácia de métodos baseados em léxico (<i>VADER</i> , <i>SentiWordNet</i> , <i>TextBlob</i>) em comparação com algoritmos de aprendizagem supervisionada.	<i>StockTwits</i>	<i>VADER</i> obteve a maior exatidão (94,4%) e <i>AUC</i> (0,861), superando <i>SentiWordNet</i> , <i>TextBlob</i> e modelos de aprendizagem supervisionada (<i>Logistic Regression</i> , <i>Naïve Bayes</i> , <i>SVM</i>).
Taj et al. (2019)	Determinar a polaridade de artigos de notícias usando abordagem baseada em léxico.	<i>BBC News Datasets</i>	Negócios e desporto com maioria de artigos positivos; entretenimento e tecnologia com predominância de negativos; política equilibrada entre positivo e negativo.

Al-Shabi (2020)	Avaliar o desempenho dos principais léxicos utilizados na análise de sentimentos no Twitter (<i>VADER</i> , <i>SentiWordNet</i> , <i>SentiStrength</i> , <i>Liu e Hu Opinion Lexicon</i> e <i>AFINN-111</i>).	<i>Stanford Twitter Sentiment Test Set</i> e <i>Sandars Twitter Dataset</i>	<i>VADER</i> obteve a melhor exatidão em ambos os conjuntos de dados (72% e 65%), destacando-se na classificação positiva e negativa; <i>AFINN-111</i> teve melhor desempenho na classificação negativa; <i>SentiWordNet</i> apresentou a menor exatidão.
Marutho et al. (2022)	Otimizar análise de sentimentos usando léxico <i>VADER</i> como extração de características para aprendizagem automática.	<i>Amazon Product Review</i> e <i>SemEval 2016 Task 5</i>	Melhor desempenho: <i>TF-IDF + SVM</i> (92,6% para dados não rotulados; 80,3% para rotulados). Combinação de <i>VADER</i> e aprendizagem automática aumentou a exatidão.
Mehenaoui et al. (2024)	Comparar o desempenho de quatro léxicos de análise de sentimentos (<i>VADER</i> , <i>TextBlob</i> , <i>SentiWordNet</i> , <i>SpaCy</i>).	Comentários do website <i>Mark My Professor</i>	<i>TextBlob</i> obteve a melhor exatidão (92,34%) e <i>F1-score</i> (89,23%); <i>SpaCy</i> e <i>VADER</i> tiveram desempenhos próximos (77,62% e 77,06% de exatidão); <i>SentiWordNet</i> apresentou o pior desempenho (71,40%).

Embora de fácil utilização, os métodos apresentados na Tabela 3 revelam limitações na detecção de sarcasmo, ironia e variações contextuais. São frequentemente superados por métodos modernos baseados em aprendizagem profunda, como os modelos transformadores (Rizinski et al., 2024).

2.3 Aprendizagem computacional tradicional

As abordagens baseadas em léxicos são muitas vezes ultrapassadas por algoritmos de aprendizagem computacional. Alguns estudos demonstram que métodos como *Support Vector Machines (SVM)* e *Naïve Bayes* apresentam melhor desempenho (Kolchyna et al., 2015).

O método *Naïve Bayes* classifica dados com base em probabilidades e no teorema de *Bayes*. É eficiente mesmo com grandes volumes de dados e prevê a classe com maior probabilidade (Danyal et al., 2023).

O algoritmo *K-Nearest Neighbor (KNN)* classifica um objeto atribuindo-lhe a categoria predominante entre os seus vizinhos mais próximos (Radhiya et al., 2022). Este método de classificação representa os dados num espaço multidimensional, onde cada dimensão corresponde a uma característica do objeto.

Uma árvore de decisão é um modelo hierárquico de aprendizagem supervisionada, no qual os dados são divididos recursivamente em nós de decisão com base em testes realizados em atributos. A estrutura de uma árvore de decisão é semelhante a um fluxograma, onde cada nó interno realiza um teste num atributo, cada ramo representa os resultados desse teste e cada nó final define o rótulo da classe (Bayhaqy et al., 2018).

Para separar os dados em diferentes classes, o *Support Vector Machine (SVM)* procura um hiperplano num espaço de alta dimensão. A qualidade do hiperplano depende da distância ao ponto de treino mais próximo (Abdelwahab et al., 2015). Pode ser aplicado tanto a tarefas de classificação como de regressão.

A Tabela 4 apresenta alguns dos estudos que aplicam algoritmos de aprendizagem computacional tradicional à análise de sentimentos.

Tabela 4 - Resumo de estudos com aplicação de aprendizagem computacional tradicional na análise de sentimentos.
Fonte: Elaboração própria.

Autores/Ano	Objetivo	Dados	Resultados
Abdelwahab et al. (2015)	Avaliar o impacto do tamanho do conjunto de treino no desempenho de <i>SVM</i> e <i>Naïve Bayes</i> em análise de sentimentos.	Conjunto do <i>SemEval 2014</i> : 4.269 <i>tweets</i> para treino e 782 para teste.	<i>SVM</i> obteve exatidão de 73,27%–76,47% e <i>Naïve Bayes</i> 73%–75%, com pouca variação ao aumentar o treino.
Bayhaqy et al. (2018)	Comparar o desempenho de três algoritmos de classificação (<i>Decision Tree</i> , <i>K-NN</i> e <i>Naïve Bayes</i>) na análise de sentimentos.	<i>Tweets</i> sobre <i>Tokopedia</i> e <i>Bukalapak</i> (Indonésia)	<i>Naïve Bayes</i> obteve o melhor desempenho (exatidão 77%, precisão 88,50%, sensibilidade 64%), superando <i>Decision Tree</i> e <i>K-NN</i> .
Tarimer et al. (2019)	Criar e comparar modelos de análise de sentimentos utilizando técnicas de representação vetorial e algoritmos de aprendizagem computacional.	2.000 críticas de filmes do <i>IMDB</i> e 3.200 <i>tweets</i> sobre empresas tecnológicas	No <i>IMDB</i> , <i>Decision Tree</i> obteve melhor exatidão (94,00%), seguido de <i>SVM</i> (85,50%) e <i>Naïve Bayes</i> (73,20%). Nos dados do <i>Twitter</i> , <i>Decision Tree</i> liderou (82,76%), seguido de

			<i>Naïve Bayes</i> (75,44%) e <i>SVM</i> (72,50%).
Wongkar & Angdresey (2019)	Analisar sentimentos em tweets sobre os candidatos presidenciais da Indonésia em 2019, comparando <i>Naïve Bayes</i> , <i>SVM</i> e <i>KNN</i> .	443 <i>tweets</i> recolhidos entre janeiro e maio de 2019	<i>Naïve Bayes</i> apresentou melhor exatidão (80,90%), seguido de <i>KNN</i> (75,58%) e <i>SVM</i> (63,99%).
Tusar & Islam (2021)	Comparar técnicas de <i>NLP</i> (<i>Bag-of-Words</i> e <i>TF-IDF</i>) e algoritmos de aprendizagem computacional (<i>SVM</i> , <i>Logistic Regression</i> , <i>Multinomial Naïve Bayes</i> , <i>Random Forest</i>) na análise de sentimentos.	14.640 <i>tweets</i> do conjunto de dados <i>Twitter US Airline Sentiment</i>	<i>SVM</i> e <i>Logistic Regression</i> obtiveram a melhor exatidão (77%) com <i>Bag-of-Words</i> , superando <i>Naïve Bayes</i> (74%) e <i>Random Forest</i> (74%).
Radhiya et al. (2022)	Desenvolver um modelo de análise de sentimentos para críticas gastronômicas usando <i>KNN</i> com <i>TF-IDF</i> .	1.953 críticas em indonésio de dois restaurantes em Bandung.	Melhor desempenho com <i>TF-IDF</i> e <i>KNN</i> , sem remoção de <i>stopwords</i> : exatidão 93,09% e <i>F1-score</i> 93,08%.
Danyal et al. (2023)	Comparar o desempenho de <i>Linear SVM</i> e <i>Multinomial Naïve Bayes</i> com outros algoritmos de referência (<i>LR</i> , <i>RFC</i> , <i>DT</i> , <i>K-NN</i> , <i>GBC</i> , <i>PAC</i>) na análise de sentimentos.	<i>IMDB Dataset</i> de 50,000 críticas e <i>Sentiment Polarity Dataset v2.0</i>	No <i>IMDB</i> , <i>PAC</i> obteve melhor exatidão (90,27%), seguido de <i>Linear SVM</i> e <i>LR</i> (89,96%). No <i>Sentiment Polarity</i> , <i>Multinomial Naïve Bayes</i> liderou com 71,04%.

Os estudos presentes na Tabela 4 mostram que o desempenho dos algoritmos de aprendizagem computacional tradicional varia consoante os dados utilizados, não existindo um classificador superior em todos os cenários.

2.4 Aprendizagem profunda

Os modelos de aprendizagem profunda, baseados em redes neuronais, permitiram avanços significativos na análise de sentimentos. Estes métodos alcançaram desempenhos superiores aos métodos tradicionais (Ahmad & Umar, 2023). Estas abordagens recorrem a redes neuronais com múltiplas camadas. Estas camadas têm capacidade para extrair

características complexas dos dados. Os três principais tipos de redes neuronais são: *Feed-forward Neural Networks (FNN)*, *Convolutional Neural Networks (CNN)* e *Recurrent Neural Networks (RNN)* (Pashaei & Pashaei, 2021).

Um *Multilayer Perceptron (MLP)* é um tipo de rede *FNN*. Neste tipo de redes as informações fluem numa única direção, da camada de entrada para a camada de saída. São compostas por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Os neurónios estão organizados em camadas, com conexões das camadas inferiores para as superiores. Não existem ligações na mesma camada. O número de neurónios na camada de entrada corresponde ao número de variáveis independentes (atributos usados como entrada do modelo). O número de neurónios na camada de saída corresponde ao número de variáveis dependentes (valores que o modelo pretende prever) (Ramchoun et al., 2016), podendo estas representar classes, no caso de classificação, ou valores contínuos, no caso de regressão.

Uma *Rede Neuronal Convolutiva (CNN)* é um tipo de rede especializada para lidar com formatos espaciais. Como por exemplo imagens 2D. As *CNN* são utilizadas para tarefas de reconhecimento de padrões (ex: caracteres manuscritos) (Lecun et al., 1998).

A *Rede Neuronal Recorrente (RNN)* tem a capacidade de modelar a dependência estrutural de textos curtos ou frases. No entanto, enfrentam dificuldades em modelar dependências de longo prazo (relações entre palavras que estão muito distantes entre si na sequência). É o problema do *vanishing gradient* (Haque et al., 2019). Esta limitação origina que os gradientes fiquem muito pequenos e desta forma não são propagados corretamente ao longo de longas sequências.

O *Long Short-Term Memory (LSTM)* é uma versão melhorada da *RNN* tradicional tendo como objetivo mitigar o problema *vanishing gradient*. Esta rede tem a capacidade de lidar com dependências de longo prazo. Alguns exemplos são a previsão da próxima palavra numa frase, resposta a perguntas, tradução automática, legendas de imagens, entre outras (Haque et al., 2019).

A Tabela 5 apresenta uma síntese de estudos que aplicam modelos baseados em redes neuronais na análise de sentimentos.

Tabela 5 - Resumo de estudos com aplicação de aprendizagem profunda na análise de sentimentos. Fonte: Elaboração própria.

Autores/Ano	Objetivo	Dados	Resultados
Kim (2014)	Avaliar o desempenho de <i>CNNs</i> simples, com vetores de palavras pré-treinados em <i>word2vec</i> , em tarefas de classificação de frases, incluindo análise de sentimentos.	<i>Stanford Sentiment Treebank, Movie Reviews, Subjectivity dataset, TREC question dataset</i> e comentários de produtos.	<i>CNN</i> com vetores pré-treinados obteve resultados de topo em 4 de 7 tarefas.
Yin et al. (2017)	Comparar <i>CNN, GRU</i> e <i>LSTM</i> em várias tarefas de <i>NLP</i> , incluindo classificação de sentimentos.	<i>Stanford Sentiment Treebank (SST)</i> .	<i>GRU</i> obteve melhor exatidão (86,32%), seguido de <i>LSTM</i> (84,51%) e <i>CNN</i> (82,38%).
Haque et al. (2019)	Comparar <i>CNN, LSTM</i> e <i>LSTM-CNN</i> na classificação de sentimentos.	<i>IMDb Movie Reviews</i> .	<i>CNN</i> obteve melhor desempenho geral (exatidão 90%), superando <i>LSTM</i> (88%) e <i>LSTM-CNN</i> (89%).
F. Huang et al. (2022)	Propor o modelo <i>AEC-LSTM</i> , que combina <i>LSTM</i> melhorado com inteligência emocional (<i>ELSTM</i>).	Quatro conjuntos de dados de redes sociais (<i>Twitter</i> e <i>microblogs</i>).	Em dados reais, o <i>AEC-LSTM</i> melhora a classificação de sentimentos e supera os principais métodos de aprendizagem profunda.
Airlangga (2024)	Comparar <i>MLP, CNN, LSTM, BiLSTM, GRU</i> e modelos com atenção na detecção de <i>spam</i> .	1,956 comentários do <i>YouTube</i> .	<i>LSTM</i> obteve a maior exatidão (95,65%), seguido de perto por <i>CNN</i> (94,37%), <i>BiLSTM</i> (95,14%) e modelo com atenção (95,14%). <i>GRU</i> (91,30%) e <i>MLP</i> (54,99%) tiveram desempenho inferior.
Shiri et al. (2025)	Comparar <i>CNN, RNN, LSTM, BiLSTM, GRU, BiGRU, TCN</i> e modelos baseados em transformadores na análise de sentimentos.	<i>IMDb Movie Reviews</i> .	O modelo baseado em transformadores obteve o melhor desempenho (exatidão 88,03%), superando <i>CNN</i> (85,90%), <i>LSTM</i> (87,53%), <i>BiLSTM</i> (87,45%), <i>GRU</i> (87,55%), <i>BiGRU</i> (87,97%), <i>TCN</i> (84,42%) e <i>RNN</i> (59,03%).

A Tabela 5 mostra que as arquiteturas baseadas em aprendizagem profunda apresentam, em geral, um bom desempenho na tarefa de análise de sentimento.

2.5 Modelos baseados em transformadores

Apesar dos progressos das *LSTM*, persistiam ainda algumas limitações (Al-Selwi et al., 2023). Os modelos baseados em transformadores ajudaram a ultrapassar limitações na captura de dependências de longo alcance em tarefas de processamento de linguagem natural (Pipalia et al., 2020).

Introduzido pela Google no artigo '*Attention is All You Need*' (Vaswani et al., 2017), os modelos baseados em transformadores substituíram a recorrência e a convolução por um mecanismo de atenção, permitindo captar dependências de longo alcance de forma eficiente. O mecanismo de atenção permite atribuir maior peso às palavras mais relevantes de uma frase (Jaggi et al., 2021). A aprendizagem por transferência, que reutiliza conhecimento prévio para melhorar o desempenho noutra tarefa, tem demonstrado grande sucesso em variadas aplicações, incluindo o processamento de linguagem natural (Radhakrishnan et al., 2023).

Baseado exclusivamente na camada de *encoders* da arquitetura dos transformadores, o *BERT* (*Bidirectional Encoder Representations from Transformers*) foi desenvolvido por investigadores da Google em 2018 (Devlin et al., 2018). Segundo Gupta (2024), o *BERT* impulsionou a adoção da aprendizagem por transferência como alternativa aos modelos especializados (modelos treinados de raiz para uma tarefa específica). O autor sublinha que esta técnica se estabeleceu como uma base sólida para novos desenvolvimentos, eliminando a necessidade de projetar arquiteturas do zero.

Diversos estudos comparativos têm analisado o desempenho de modelos baseados em transformadores na tarefa de análise de sentimentos. A Tabela 6 apresenta um resumo de alguns estudos.

Tabela 6 - Resumo de estudos comparativos sobre modelos baseados em transformadores aplicados à análise de sentimentos. Fonte: Elaboração própria.

Autores/Ano	Objetivo	Dados	Resultados
Pipalia et al. (2020)	Comparar <i>BERT</i> e <i>XLNet</i> em análise de sentimentos.	<i>IMDb Reviews</i> .	Ambos modelos atingiram elevada exatidão, com ligeira vantagem do <i>XLNet</i> sobre o <i>BERT</i> .
Ilgun & Kilic (2021)	Comparar modelos clássicos (<i>LR</i> , <i>SVM</i> , <i>KNN</i>) e <i>BERT</i> combinado com esses classificadores, para análise de sentimento.	<i>IMDb Reviews</i> ; <i>tweets</i> (vários temas); <i>tweets</i> do debate <i>GOP 2016</i> .	<i>BERT + SVM</i> e <i>BERT + LR</i> superaram consistentemente os modelos clássicos.
Singh & Jain (2021)	Avaliar <i>BERT</i> , <i>RoBERTa</i> , <i>DistilBERT</i> , <i>XLNet</i> em análise de sentimentos.	Títulos de notícias.	<i>BERT</i> foi o melhor (MCC=90,1%, F1=93,6%, Exatidão=93,6%), superando <i>RoBERTa</i> , <i>DistilBERT</i> e <i>XLNet</i> .
Bozanta et al. (2021)	Comparar o desempenho de classificadores tradicionais (<i>LR</i> , <i>RF</i>), modelos de aprendizagem profunda (<i>LSTM</i> , <i>GRU</i>) e <i>transformer</i> pré-treinados (<i>BERT</i> , <i>DistilBERT</i> , <i>RoBERTa</i> , <i>XLNet</i>) na análise de sentimentos.	<i>StockTwits</i>	<i>RoBERTa</i> obteve o melhor desempenho médio (F1=0,88 para SPY), superando outros transformadores, modelos <i>DL</i> e abordagens tradicionais. <i>BERT</i> apresentou resultados próximos, enquanto <i>RF</i> superou <i>LSTM</i> e <i>GRU</i> .
Chinnalagu & Durairaj (2022)	Comparar <i>BERT</i> e modelos de aprendizagem profunda (<i>fastText-BiLSTM</i> , <i>fastText-Trigram</i> , <i>LSVM-Trigram</i>) em análise de sentimentos.	Grandes conjuntos de avaliações de clientes recolhidos de redes sociais.	O <i>BERT</i> apresentou melhor desempenho em todas as métricas face aos modelos <i>DL</i> testados. O híbrido <i>fastText-BiLSTM</i> superou outros modelos não baseados em transformadores.
Areshey & Mathkour (2024)	Avaliar e comparar <i>BERT</i> , <i>RoBERTa</i> , <i>ALBERT</i> , <i>DistilBERT</i> e <i>XLNet</i> em classificação de sentimentos.	<i>Yelp Open Dataset</i> .	<i>RoBERTa</i> obteve a melhor exatidão (98,30%), seguido de <i>XLNet</i> (98,20%), <i>BERT</i> (97,40%), <i>ALBERT</i> (97,20%) e <i>DistilBERT</i> (96,00%).

Os estudos presentes na Tabela 6 demonstram a superioridade dos modelos baseados em transformadores face a abordagens tradicionais.

2.6 Técnicas de pré-processamento

Os dados gerados nas redes sociais apresentam frequentemente ruído, erros de sintaxe e problemas gramaticais (Sharma & Goyal, 2023). O pré-processamento é um passo essencial para obter bons resultados e é mencionado na maioria dos estudos. Contudo, poucos trabalhos analisam o impacto de cada técnica básica aplicada a dados textuais (Angiani et al., 2016).

No estudo de Kaur & Sharma (2023), as técnicas de pré-processamento visaram a remoção de dados irrelevantes, palavras sem valor semântico e outros elementos indesejados. De acordo com Siino et al. (2024), o pré-processamento envolve a limpeza e preparação dos dados, que depois são segmentados em unidades linguísticas, desempenhando um papel crucial antes e durante a seleção de características. Neste estudo são identificadas várias estratégias de pré-processamento. Sendo uma lista extensa, destacam-se as seguintes:

(i) remoção de ruído, (ii) substituição de abreviações e gírias, (iii) expansão de contrações, (iv) tratamento de pontuação repetida, (v) remoção de pontuação, (vi) remoção de números, (vii) conversão para minúsculas, (viii) remoção de palavras sem significado semântico, (ix) correção ortográfica, (x) marcação de classes gramaticais, (xi) lematização, (xii) radicalização, (xiii) redução de caracteres repetidos, (xiv) normalização de *emojicons*, (xv) tratamento de negações, (xvi) segmentação de palavras.

Em modelos modernos baseados em transformadores, a escolha das técnicas de pré-processamento pode ter um impacto significativo, que é confirmado no estudo apresentado por Siino et al. (2024). Foi demonstrado que um pré-processamento bem aplicado permitiu que modelos simples, como o *Naïve Bayes*, superassem transformadores em tarefas de classificação. Verificou-se também que os modelos de aprendizagem profunda são os menos sensíveis à estratégia utilizada, enquanto os transformadores são os mais afetados pela aplicação destas técnicas.

2.7 Técnicas de balanceamento de dados

Dados do mundo real tendem a apresentar desequilíbrio de classes, com predominância de certas classes sobre outras (Prusa et al., 2015). Conjuntos de dados desequilibrados são um grande desafio em processamento de linguagem natural e afetam o desempenho dos modelos (Kedas et al., 2022). O desequilíbrio de classes favorece a classe majoritária e afeta negativamente o desempenho dos classificadores (Hasib et al., 2020). Os modelos tendem a ter bom desempenho nas classes com muitos exemplos e fraco nas classes minoritárias (Gao et al., 2025).

As classes minoritárias podem conter informação crítica. Ao não atribuir o mesmo peso a estes exemplos poderá dar origem a modelos enviesados, métricas desajustadas e previsões pouco fiáveis (Yu & Zhou, 2021). Para mitigar este problema, usam-se técnicas de amostragem que aumentam os exemplos da classe minoritária (por duplicação ou geração sintética) ou reduzem os da classe majoritária (eliminando registos redundantes). A redução da classe majoritária poderá ser útil em conjuntos de dados muito grandes, pois ajuda a reduzir os custos computacionais (Prusa et al., 2015). Ambas as abordagens procuram equilibrar a distribuição das classes no conjunto de treino e melhorar a capacidade de generalização dos modelos (Yu & Zhou, 2021).

O estudo de Kedas et al. (2022) demonstra que os modelos com dados balanceados superaram os restantes, com melhorias significativas face aos originais.

2.8 Técnicas de representação textual

As máquinas não interpretam texto como os humanos. Para que uma máquina consiga extrair significado de um texto, os dados precisam de ser convertidos em vetores numéricos (Siebers et al., 2022).

Técnicas como *Bag-of-Words (BoW)* e *Term Frequency-Inverse Document Frequency (TF-IDF)* representam a frequência e a importância das palavras numa frase. No *BoW*, cada documento é representado por um vetor onde cada posição indica quantas vezes uma palavra do dicionário aparece no texto. O *TF-IDF* ajusta as frequências, reduz o peso das palavras comuns e realça as menos frequentes (Zhixiang et al., 2013). Apesar da sua utilidade, estes métodos discretos baseados na frequência de palavras têm limitações claras pois perdem o contexto (Pak et al., 2024). Não capturam relações de significado

entre palavras e geram matrizes muito grandes quando aplicados a textos extensos (Khomsah et al., 2022).

Modelos como *Word2Vec* (Mikolov et al., 2013), *FastText* (Bojanowski et al., 2017) e *GloVe* (Pennington et al., 2014) aprendem representações densas de palavras projetando-as num espaço vetorial contínuo (Pak et al., 2024). As palavras com significados semelhantes ficam próximas umas das outras. O *Word2Vec* transforma palavras em vetores densos usando dois modelos: *Continuous Bag-of-Words (CBOW)* e *Skip-gram*. O *FastText* considera sub-palavras através de *n-grams*. Desta forma fica mais robusto para lidar com termos raros ou desconhecidos (Khomsah et al., 2022). Já o *GloVe* cria vetores numéricos com base na frequência de coocorrência das palavras e na proximidade de significado (Çano & Morisio, 2019).

A forma como o texto é representado afeta diretamente o desempenho dos modelos. Estudos mostram que representações baseadas em vetores de palavras tendem a superar métodos discretos em tarefas de análise de sentimentos, pois captam melhor as relações de significado e o contexto das palavras (Khomsah et al., 2022).

O estudo de Asudani et al. (2023) apresenta uma revisão sobre três tipos de modelos de representação textual: convencionais, distribucionais e contextuais. Os resultados mostram que os modelos contextuais, como o *BERT*, alcançam melhor desempenho em tarefas exigentes, como classificação de texto, análise de sentimentos e reconhecimento de entidades nomeadas. No entanto, os modelos estáticos mantêm relevância, por exigirem menos recursos computacionais e continuarem eficazes.

2.9 Aplicação no contexto financeiro

Na última década, a análise de sentimento tem assumido uma relevância cada vez maior no contexto financeiro (Chaturvedi et al., 2018; Du et al., 2024; Xing et al., 2018).

As redes sociais têm fornecido uma grande quantidade de dados. Estes dados podem ser utilizados para compreender o comportamento dos investidores e o que pensam sobre o mercado (Nousi & Tjortjis, 2021; Park & Woo, 2019). Estas plataformas, específicas do domínio financeiro (como o *Twitter* e *Stocktwits*), tem alterado a forma de analisar o sentimento público. Segundo Nousi & Tjortjis (2021), estas plataformas de *microblogging* oferecem informação valiosa sobre as emoções das pessoas. O sentimento

público positivo que os investidores demonstram em relação a uma empresa geralmente resulta no aumento dos preços dessas ações. Sentimentos negativos tendem a provocar o efeito contrário. Pagolu et al. (2016) confirmam que notícias e comentários positivos sobre uma empresa estimulam o investimento e levam à valorização das suas ações. Também Alshahrani et al. (2018) afirmam que as reações emocionais dos investidores, como reação a notícias e eventos, podem ter um impacto significativo nas suas decisões de compra e venda de ações.

A análise de sentimentos, neste domínio, desempenha um papel crucial. Permite prever retornos, volatilidade e desenvolver estratégias de negociação e gestão de risco (Deveikyte et al., 2022).

Abordagens que recorrem a Modelos de Linguagem de Grande Escala adaptados ao setor financeiro têm demonstrado avanços significativos na extração de informação e na previsão de movimentos de mercado, conforme evidenciado no estudo de revisão de (Darwish et al., 2025).

A Tabela 7 apresenta alguns exemplos representativos dessas abordagens.

Tabela 7 - Estudos de abordagens específicas de análise de sentimentos adaptados ao domínio financeiro. Fonte: Elaboração própria.

Autores/Ano	Abordagem	Descrição
Schumaker & Chen (2009)	BoW, Sintagmas Nominais, Entidades Nomeadas	Estuda várias representações textuais aplicadas a notícias financeiras para prever movimentos no mercado de ações.
Li & Shah (2017)	<i>SOWE</i>	Primeiro modelo de representação textual especificamente aplicado ao contexto do mercado de ações.
Gutiérrez-Fandiño et al. (2022)	<i>FinEAS</i>	Modelo pré-treinado em domínios gerais e posteriormente ajustado para o setor financeiro, focado em análise de sentimentos.
A. H. Huang et al. (2023)	<i>FinBERT</i>	Modelo de linguagem baseado no <i>BERT</i> , adaptado para o domínio financeiro e tarefas de análise de sentimentos.
Rodriguez Inserte et al. (2023)	Adaptação de <i>LLMs</i>	Investigação sobre a adaptação de Modelos de Linguagem de Grande Escala, treinados com grandes volumes de dados, para o domínio financeiro.
Konstantinidis et al. (2024)	<i>FinLlama</i>	Versão especializada de modelo <i>LLaMA</i> para a área financeira, com aplicações em <i>trading</i> algorítmico (negociação de ativos financeiros efetuada automaticamente)

por sistemas computacionais que seguem regras ou modelos predefinidos).

Os estudos presentes na Tabela 7 mostram a evolução de representações textuais específicas para modelos de linguagem de grande escala adaptados ao domínio financeiro.

2.10 Plataforma StockTwits

O *StockTwits* é uma rede social exclusivamente dedicada ao mercado de ações (Oliveira et al., 2013). Comparado ao *Twitter*, o *StockTwits* é mais informativo para dados financeiros (Divernois & Filipović, 2024; Ghoshal & Roberts, 2016). Direcionada para a comunicação sobre mercados e ações, serve como um espaço para utilizadores que possuem um conhecimento mais especializado sobre o mercado financeiro (Nousi & Tjortjis, 2021) partilharem ideias e informações.

Os utilizadores do *StockTwits* têm a possibilidade de publicar mensagens sobre qualquer ação cotada. Para isso utilizam o símbolo '\$' seguido do identificador da ação, como se ilustra na Figura 3.

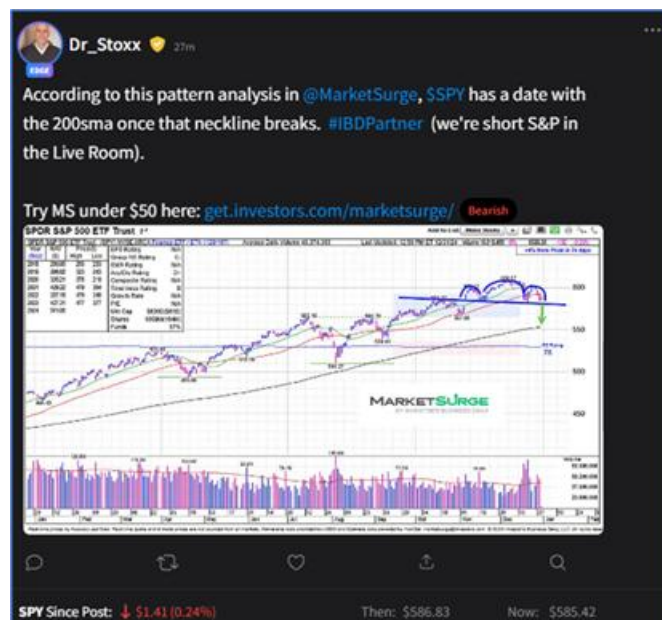


Figura 3 - Exemplo de mensagem. Fonte: StockTwits (acesso em 9 de janeiro de 2025).

A plataforma permite aos utilizadores marcar mensagens como '*bullish*' (otimista) ou '*bearish*' (pessimista). Estes dois termos são vulgarmente utilizados neste domínio, e são presença constante nas mensagens publicadas pelos utilizadores. Desconhece-se a origem

exata destes termos. No entanto existe uma ideia generalizada que está relacionada com a forma como ambos os animais atacam. O touro investe de baixo para cima (subida). O urso ataca de cima para baixo (descida).

A facilidade de obtenção destes comentários, através de *APIs*, é também uma mais-valia, já que simplifica muito a tarefa (Divernois & Filipović, 2024).

Vários estudos têm explorado a análise de sentimento financeiro aplicada ao mercado de ações, recorrendo ao conjunto de dados do *StockTwits*. A Tabela 8 sintetiza alguns dos estudos mais relevantes.

Tabela 8 - Estudos que utilizam dados do *StockTwits* na análise de sentimentos e previsão do mercado financeiro.
Fonte: Elaboração própria.

Autores/Ano	Abordagem	Descrição
Oliveira et al. (2013)	Avaliar previsibilidade de variáveis de mercado com dados do <i>StockTwits</i>	Analizou volume de mensagens e indicadores de sentimento para prever retornos, volatilidade e volume de negociação. Concluiu que o volume de mensagens melhora previsões de volume de negociação, mas não encontrou evidência de previsibilidade de retornos com sentimento.
Batra & Daudpota (2018)	Integrar análise de sentimentos com dados de mercado para prever preços	Recolheu mensagens sobre a <i>Apple</i> no <i>StockTwits</i> , aplicou <i>SVM</i> para classificar sentimento e combinou com dados do <i>Yahoo Finance</i> para prever movimentos diários. Os resultados mostram relação positiva entre opinião e mercado, com 76,65% de exatidão.
Nousi & Tjortjís (2021)	Prever movimentos de ações usando análise de sentimentos	Propôs metodologia baseada em <i>SVM</i> e <i>Regressão Logística</i> aplicadas a dados do <i>Twitter</i> e <i>StockTwits</i> . Estudo de caso com ações da <i>Microsoft</i> . No <i>StockTwits</i> , obteve melhor desempenho com <i>TextBlob</i> e <i>SVM</i> .
Bozanta et al. (2021)	Comparar modelos de análise de sentimentos no <i>StockTwits</i>	Comparou classificadores tradicionais, modelos de aprendizagem profunda e modelos baseados em transformadores (<i>BERT</i> , <i>RoBERTa</i> , <i>XLNet</i>) para classificar mensagens como <i>bullish</i> ou <i>bearish</i> . <i>RoBERTa</i> obteve o melhor F1-score médio.
J. Liu et al. (2023)	Prever movimentos do mercado usando sentimento do <i>StockTwits</i>	Aplicou o modelo <i>FinBERT</i> para classificar comentários do <i>StockTwits</i> sobre o <i>ETF SPDR S&P 500 (SPY)</i> . Calculou índices de sentimento e combinou-os com dados históricos para prever movimentos intradiários. Com o <i>SVM</i> obteve F1-score de 65,30%, superior a outros métodos testados.

Divernois & Filipović (2024)	Classificar sentimento e analisar relação com retornos de ações	Desenvolveu um classificador de mensagens do <i>StockTwits</i> em <i>bullish</i> , <i>bearish</i> e neutro. Criou uma métrica de polaridade diária e analisou a sua relação com retornos. A polaridade está associada a retornos contemporâneos e, em eventos de pico de volume de mensagens, prevê retornos anormais.
Delgadillo et al. (2024)	Desenvolver LLM especializado para finanças (<i>FinSoSent</i>)	Criou modelo de linguagem pré-treinado em corpora financeiros, afinado com dados do <i>StockTwits</i> e outras fontes. Superou modelos de referência (incluindo <i>FinBERT</i>) em tarefas de classificação de sentimento no domínio financeiro.

A Tabela 8 mostra a crescente utilização dos dados do *StockTwits*, evoluindo de métodos tradicionais para modelos de linguagem especializados.

2.11 Teoria do Mercado Eficiente

A Teoria do Mercado Eficiente (EMH), que sustenta que os preços das ações refletem toda a informação disponível (Fama, 1970; Komariah et al., 2016), é um conceito fundamental no mercado financeiro, apresentado por Fama (1970). Segundo M. P. Baker & Wurgler (2003), na teoria financeira clássica o sentimento dos investidores não tem qualquer influência nos preços das ações nem nos retornos.

A Teoria do Mercado Eficiente é classificada em três formas distintas, de acordo com o tipo de informação que os preços refletem (Komariah et al., 2016; Niroomand et al., 2020):

- **Forma fraca:** afirma que os preços futuros não podem ser previstos apenas com base nos preços passados.
- **Forma semiforte:** indica que os preços das ações se ajustam rapidamente à divulgação de novas informações públicas.
- **Forma forte:** afirma que os preços refletem totalmente toda a informação, pública e privada.

As emoções e opiniões dos investidores também desempenham um papel importante na dinâmica do mercado, uma vez que diversos estudos mostram que o sentimento expresso

em plataformas sociais tem impacto nos preços das ações (Bollen et al., 2011; Jaggi et al., 2021). Este contraponto teórico, entre o conceito de eficiência total e influência comportamental, pode desafiar a visão estritamente racional da EMH, uma vez que há evidência de que os preços de mercado não seguem totalmente um passeio aleatório (Bollen et al., 2011).

As opiniões públicas expressas em *microblogs* podem influenciar as tendências do mercado, sugerindo que o mercado nem sempre é estritamente racional e que fatores como o sentimento público podem afetar os preços (Komariah et al., 2016; Papaioannou et al., 2013). Os preços dos ativos não dependem inteiramente de informações desatualizadas e reagem a novas informações, como é o caso de notícias financeiras (Gite et al., 2021).

Apesar da Teoria do Mercado Eficiente ter uma longa história no campo das finanças, existem alguns estudos que sugerem que as três formas podem não descrever o comportamento dos mercados na sua totalidade. Deng (2007) afirma que a EMH “*não é uma aproximação razoável ao mercado real*” e que possui “*conteúdo científico próximo de zero*”. Niroomand et al. (2020) afirmam que “*as regras de médias móveis superam a estratégia buy and hold em mais de 10% ao ano em todo o período e em cada sub-período, considerando tanto o risco como os custos de transação*”, ou seja, é possível obter retornos consistentes, utilizando a análise técnica. Estas afirmações contradizem a hipótese de que os preços refletem toda a informação disponível.

Além das análises técnica e fundamental, a análise de sentimento tem emergido como uma ferramenta complementar no estudo de possíveis ineficiências de mercado (Pagliaro, 2025).

2.12 Mercado NASDAQ

O *NASDAQ* (*National Association of Securities Dealers Automated Quotations*) é um mercado para a negociação de valores mobiliários. É também o principal índice de referência para as ações tecnológicas nos Estados Unidos. Criado pela *National Association of Securities Dealers* (também conhecido por *FINRA - Financial Industry Regulatory Authority* (L. Manning, 2024)), foi idealizado para que os investidores negociassem através de um sistema totalmente automatizado (Hayes, 2025).

As operações começaram a 8 de fevereiro de 1971, tornando-se no primeiro mercado do mundo com uma estrutura inteiramente eletrônica. Foi um passo decisivo na modernização dos mercados financeiros. Hoje, de acordo com *World Federation of Exchanges (2025)* o *NASDAQ* é a segunda maior bolsa em capitalização bolsista a nível global, com mais de 3.000 ações listadas (Royal, 2025) e acolhendo algumas das maiores empresas de tecnologia e biotecnologia, como *Google (Alphabet)*, *Apple*, *Microsoft*, *Amazon*, *Intel* e *Oracle* (B. Baker, 2025), conforme apresentado na Tabela 9.

Tabela 9 - As 10 maiores bolsas de valores do mundo por capitalização bolsista, incluindo o número de empresas listada. Fonte: Elaboração própria. com base nos dados da *World Federation of Exchanges* de maio de 2025 (acesso em 24 de Agosto de 2025).

Bolsa	Capitalização de Mercado (USD)	Empresas Listadas
New York Stock Exchange	\$ 31,7 trilhões	2.126
Nasdaq	\$ 29,9 trilhões	3.285
Shanghai Stock Exchange	\$ 7,3 trilhões	2.284
Japan Exchange Group	\$ 6,9 trilhões	3.958
Euronext	\$ 6,0 trilhões	1.808
London Stock Exchange	\$ 5,9 trilhões	1.025
Hong Kong Exchanges and Clearing	\$ 5,2 trilhões	2.633
National Stock Exchange of India	\$ 5,2 trilhões	2.735
Shenzhen Stock Exchange	\$ 4,6 trilhões	2.872
TMX Group	\$ 3,8 trilhões	3.581

Dentro do *NASDAQ*, estão incluídas as empresas do grupo *FAANG*. Trata-se de um acrónimo que representa as maiores e mais influentes empresas de tecnologia do mundo (Jaggi et al., 2021).

O “F” corresponde ao *Facebook*, atualmente *Meta*, focada no *metaverso* e redes sociais. O primeiro “A” é a *Apple*, conhecida pelos seus produtos inovadores como o *iPhone* e o *MacBook*. O segundo “A” refere-se à *Amazon*, líder global em comércio eletrónico e serviços de *cloud computing*. O “N” representa a *Netflix*, uma das principais plataformas de *streaming* de vídeo. O “G” corresponde à *Google*, parte da *holding Alphabet*, líder em pesquisa *online* e publicidade digital. A estas empresas junta-se a *Microsoft*, reconhecida pelo *software* e soluções empresariais.

A Tabela 10 apresenta os principais indicadores financeiros das empresas deste estudo.

Tabela 10 - Dados financeiros das ações das empresas FAANG mais Microsoft, incluindo capitalização de mercado, variação das ações (52 semanas) e P/E Ratio. Fonte: Elaboração própria com base nos dados do site Nasdaq.com, (acesso em 1 de janeiro de 2025).

Empresa	Capitalização de Mercado (USD)	Varição (52 semanas)	P/E Ratio
Meta	\$ 1,478,113,280,957	\$ 638,40 / \$ 340,01	51,68
Apple	\$ 3,785,304,395,660	\$ 260,10 / \$ 164,08	40,92
Amazon	\$ 2,306,888,265,045	\$ 233,00 / \$144,05	113,67
Netflix	\$ 381,001,966,170	\$ 941,75 / \$ 461,86	74,21
Google	\$ 2,331,176,040,000	\$ 202,88 / \$ 131,55	36,55
Microsoft	\$ 3,133,802,247,084	\$ 468,35 / \$ 366,50	40,80

A metodologia adotada na presente dissertação foi pensada e executada na tentativa de dar resposta às perguntas de investigação e cumprir os objetivos identificados, seguindo os princípios do *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*), que se mantém como um referencial importante para estruturar e gerir projetos de descoberta de conhecimento em dados (Martinez-Plumed et al., 2021).

A Figura 4 mostra a forma como o *framework* foi utilizado e adaptado à metodologia adotada no presente estudo.

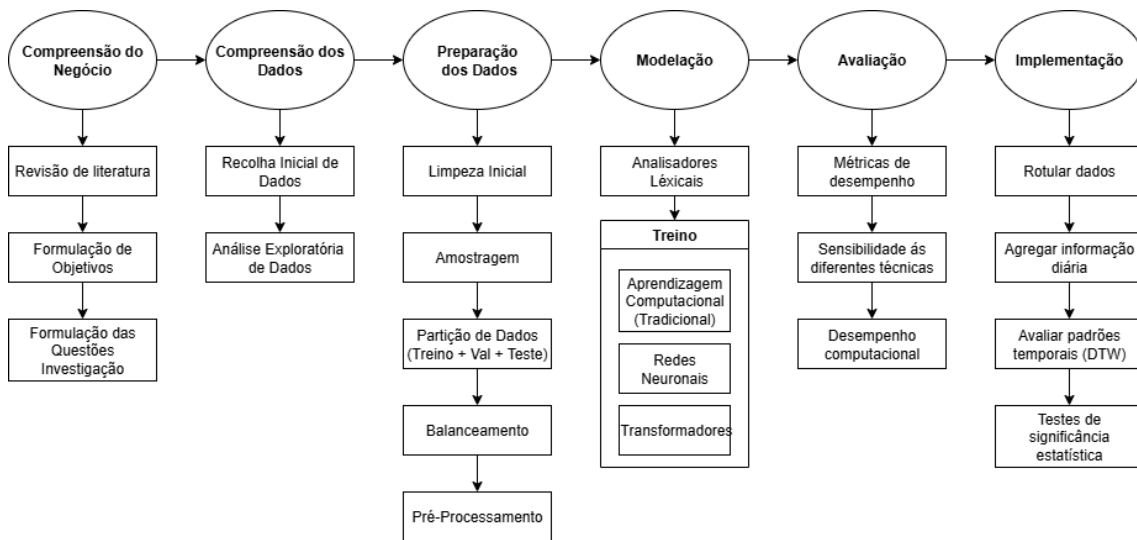


Figura 4 - Adaptação do modelo *CRISP-DM* ao presente estudo. Fonte: Elaboração própria.

3.1 Compreensão do negócio

A primeira fase da metodologia *CRISP-DM* corresponde à revisão da literatura efetuada no Capítulo 2, que teve como objetivo principal compreender o negócio, identificar o estado da arte e conhecimento sobre diferentes abordagens de processamento de linguagem natural utilizadas na análise de sentimentos no contexto financeiro. Esta fase permitiu formular os objetivos e definir as questões de investigação.

3.2 Compreensão dos dados

A segunda fase da metodologia *CRISP-DM* tem como finalidade assegurar que os dados recolhidos são adequados e compreendidos para suportar a fase de modelação, tendo sido efetuada em duas tarefas principais. Primeiro, a recolha a partir do *StockTwits*,

EODHD.com (End of Day Historical Data) e *Yahoo Finance*. Por fim, a análise exploratória, que permitiu aprofundar o conhecimento sobre a informação recolhida.

3.2.1 Recolha inicial dos dados

A recolha dos dados dos comentários do *StockTwits* foi feita através da *API REST*¹. O período analisado foi de janeiro de 2023 a dezembro de 2024. Para garantir a cobertura integral do período em análise, a recolha de comentários foi efetuada desde o final de 2022 até ao início de 2025, sendo os valores residuais utilizados apenas para assegurar a continuidade e consistência da série temporal. Este intervalo de tempo permitiu trabalhar com informação recente e evitar distorções causadas pela pandemia de *COVID-19*. A versão gratuita da *API* impunha um limite de apenas 30 mensagens por chamada. Esta situação obrigou a procurar uma solução para conseguir um maior volume de dados históricos. Após pesquisa na bibliografia, optou-se por seguir a estratégia proposta por Divernois & Filipović (2024). Esta abordagem permitiu carregar mensagens mais antigas de forma incremental. A recolha decorreu entre 3 de dezembro de 2024 e 25 de janeiro de 2025.

A recolha de notícias financeiras foi feita através da *API REST* disponibilizada pela plataforma *EODHD*². Este é um serviço especializado em fornecer dados históricos e em tempo real. É relevante salientar que este serviço disponibiliza todas as notícias previamente rotuladas com o respetivo sentimento. O período de recolha foi de janeiro de 2023 a dezembro de 2024, garantindo o alinhamento temporal com os dados do *StockTwits*. A versão gratuita da *API* limitava o acesso a quatro chamadas diárias, com um máximo de 1.000 registos. Para ultrapassar esta limitação, foi efetuada uma recolha faseada. Foi executada ao longo de vários dias, assegurando a obtenção de todos os registos existentes.

Os dados relativos à informação financeira das ações foram obtidos através do *Yahoo Finance*³. Esta plataforma é amplamente reconhecida pelo fornecimento de informações financeiras, e tem sido utilizada em diversos estudos académicos, como evidenciado nos trabalhos de Jaggi et al. (2021), S. Mokhtari et al. (2021) e M. Mokhtari et al. (2023).

¹ <https://api.stocktwits.com>

² <https://eodhd.com/financial-apis/stock-market-financial-news-api>

³ <https://finance.yahoo.com>

3.2.2 Análise exploratória de dados

A análise exploratória de dados tem como objetivo examinar e compreender a informação recolhida, identificar padrões relevantes e detetar anomalias que permitissem resolver eventuais problemas de qualidade de dados.

3.2.2.1 Comentários do *StockTwits*

O processo começou com a avaliação da distribuição dos comentários por ação. A Figura 5 mostra essa distribuição.

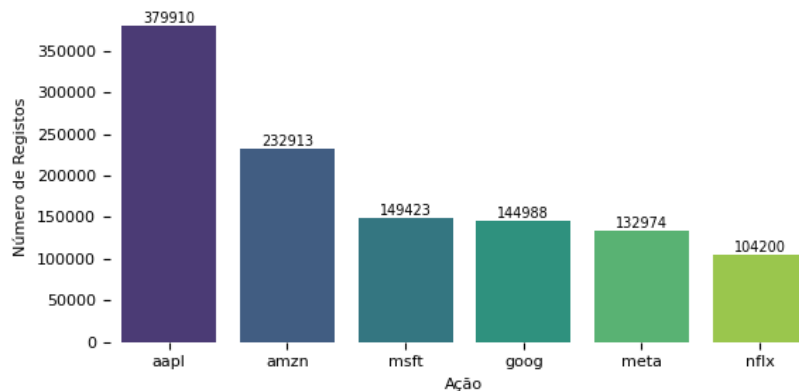


Figura 5 - Distribuição do número de comentários por ação. Fonte: Elaboração própria a partir dos dados do *StockTwits*.

A Figura 6 mostra a distribuição do número de comentários por ação ao longo dos anos. Os valores de 2022 e 2025 são apenas residuais, incluídos para manter a continuidade da série temporal. A Figura 7 destaca a evolução mensal.

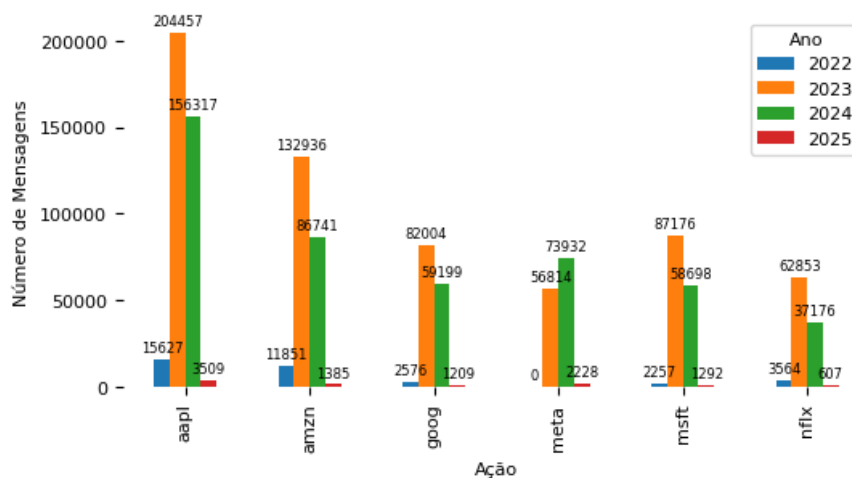


Figura 6 - Distribuição do número de comentários por ação e por ano. Fonte: Elaboração própria a partir dos dados do *StockTwits*.

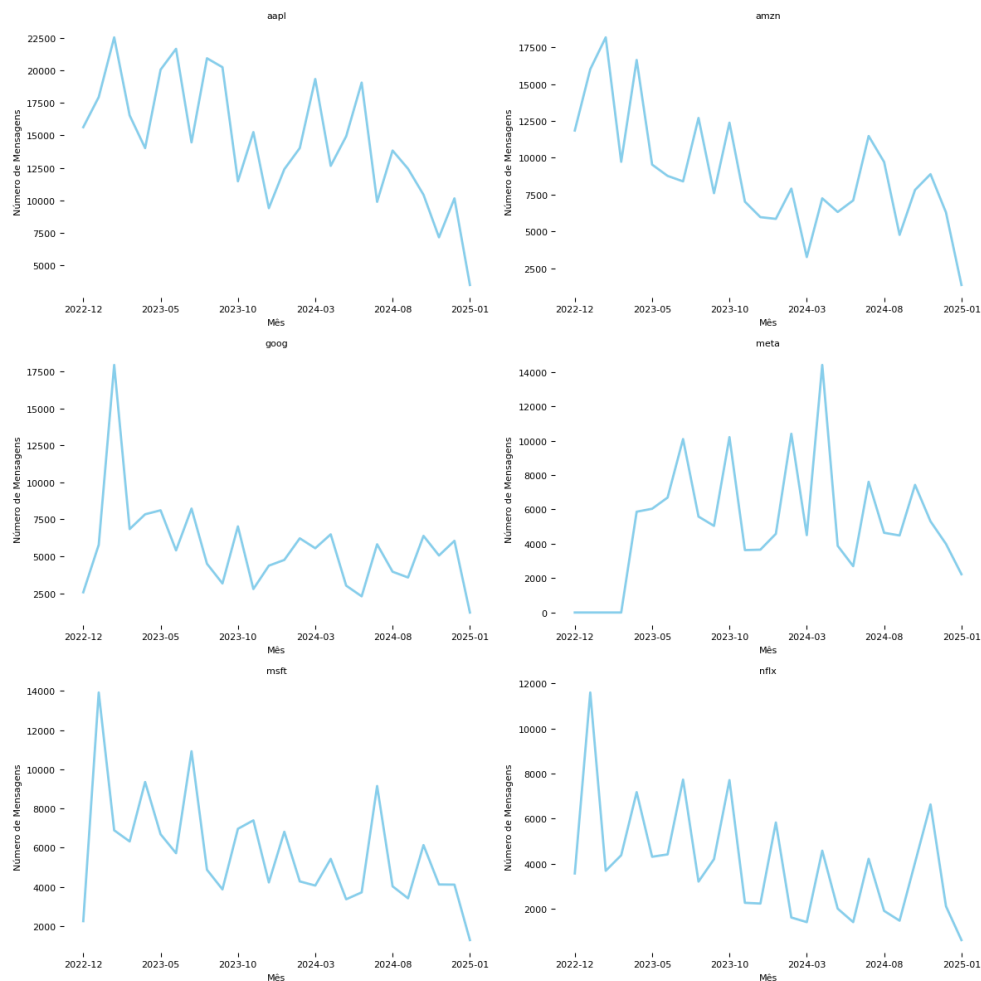


Figura 7 - Distribuição do número de comentários por ação e por mês. Fonte: Elaboração própria a partir dos dados do StockTwits.

A Figura 8 mostra a distribuição global dos sentimentos presentes no conjunto de dados, informação que já se encontrava disponível nos próprios comentários através da classificação atribuída pelos utilizadores.

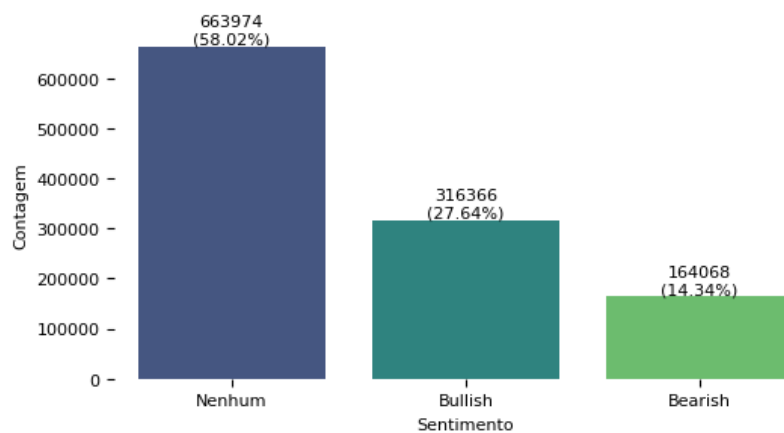


Figura 8 - Distribuição do sentimento nos comentários. Fonte: Elaboração própria a partir dos dados do StockTwits.

3.2.2.2 Notícias financeiras do EODHD

De forma semelhante ao procedimento adotado para os dados do *StockTwits*, realizou-se uma análise exploratória das notícias financeiras recolhidas. O primeiro passo foi avaliar a distribuição do número total de notícias por ação, que pode ser observado na Figura 9.

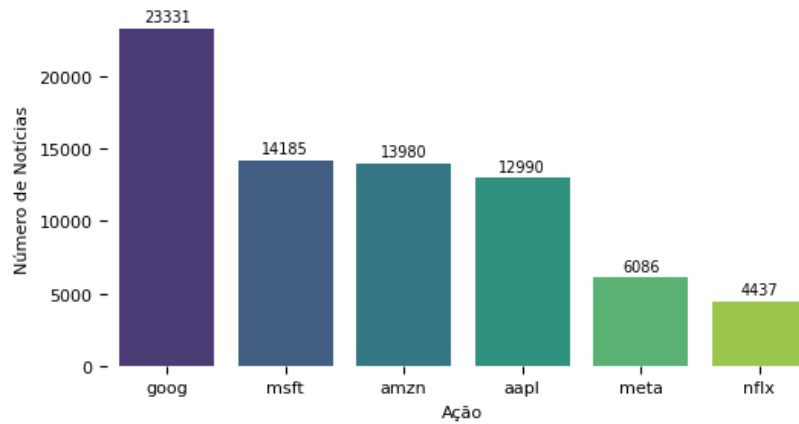


Figura 9 - Número de notícias por ação. Fonte: Elaboração própria a partir dos dados do EODHD.

A Figura 10 mostra o número de notícias por ação ao longo dos anos, enquanto a Figura 11 apresenta o mesmo indicador numa escala mensal.

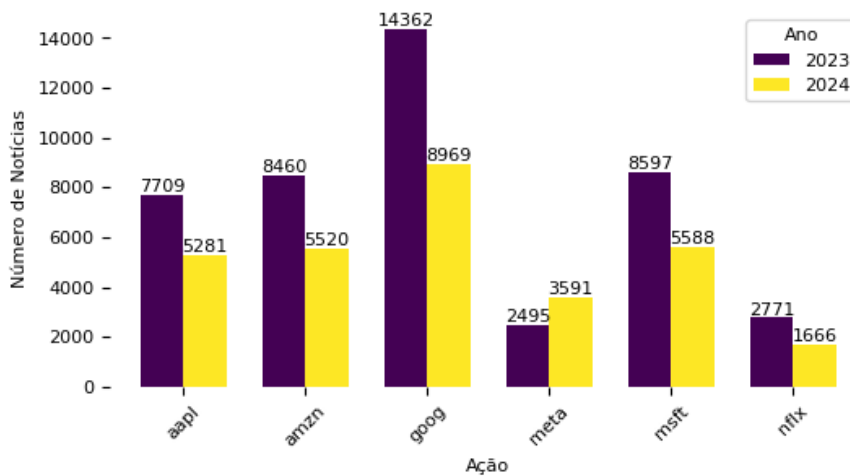


Figura 10 - Número de notícias por ação por ano. Fonte: Elaboração própria a partir dos dados do EODHD.

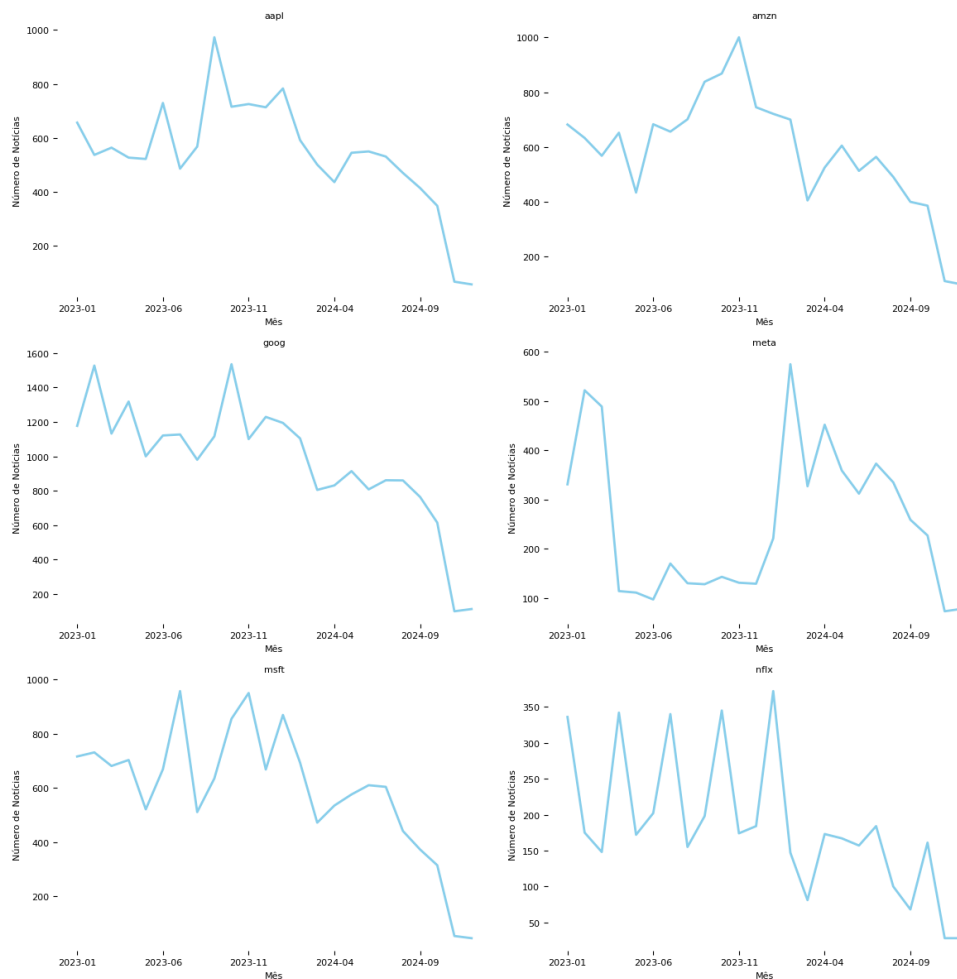


Figura 11 - Número de notícias por ação por mês. Fonte: Elaboração própria a partir dos dados do EODHD.

A Figura 12 mostra a distribuição dos sentimentos expressos nas notícias, informação já disponibilizada pelo próprio fornecedor de dados (EODHD).

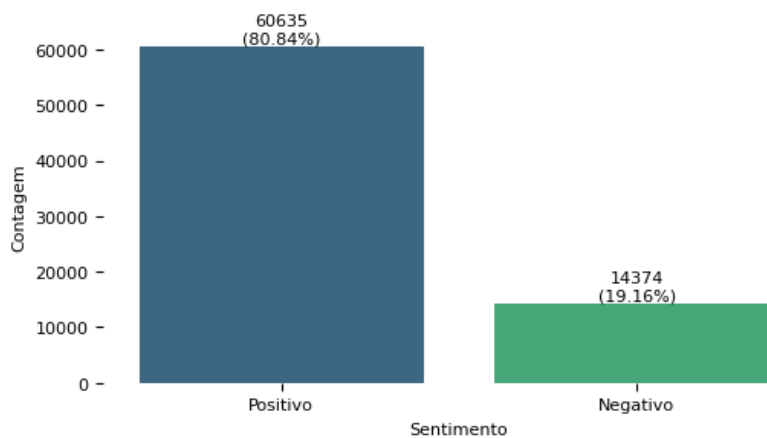


Figura 12 - Distribuição do sentimento nas notícias. Fonte: Elaboração própria a partir dos dados do EODHD.

3.2.2.3 Informação financeira das ações do *Yahoo Finance*

Foi também realizada uma análise exploratória focada na evolução do comportamento das ações ao longo do tempo, que pode ser observada na Figura 13.

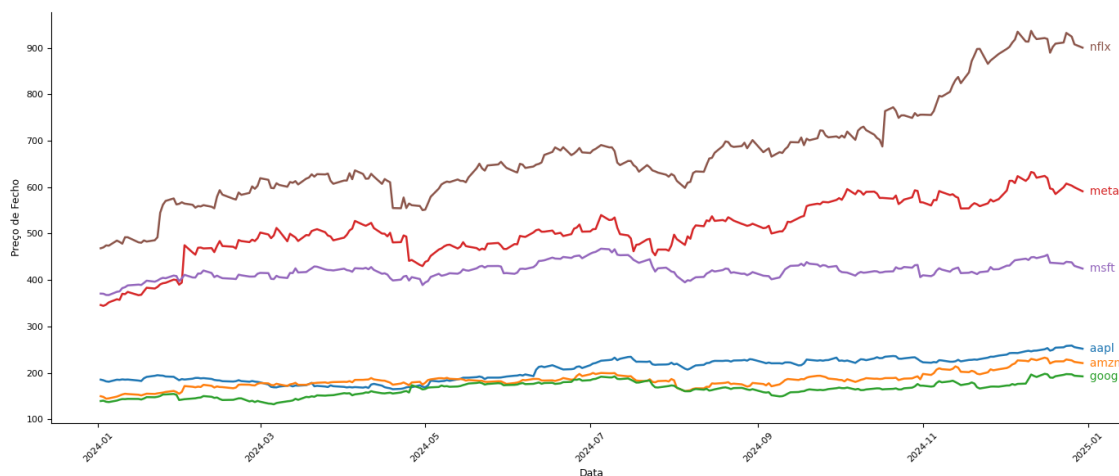


Figura 13 - Evolução do preço de fecho das ações. Fonte: Elaboração própria a partir dos dados do Yahoo Finance.

3.3 Preparação dos dados

Foi nesta etapa que se deu início ao desenvolvimento do sistema de processamento concebido para este estudo. O sistema proposto foi idealizado e implementado para que cada tarefa fosse executada de forma sequencial. Os dados de saída de uma etapa servem como entrada para a seguinte. No final de cada processamento é automaticamente gerado um ficheiro de metadados, onde ficam registadas as principais métricas e indicadores de execução. O objetivo deste sistema foi criar um mecanismo automatizado para executar, auditar e avaliar as diferentes etapas envolvidas.

3.3.1 Limpeza dos dados

A primeira tarefa deste fluxo consistiu na limpeza inicial dos dados recolhidos, através da aplicação de conjuntos de processamentos com o objetivo de garantir a qualidade da informação. Foram executadas tarefas como a eliminação de mensagens nulas e de registos duplicados, a eliminação de colunas irrelevantes, a normalização dos rótulos de sentimento, a limpeza textual (incluindo hiperligações, código *HTML*, menções e etiquetas) bem como a eliminação de mensagens sem rótulo definido. Foram também eliminadas mensagens com menos de 20 caracteres, uma vez que apenas comentários com o mínimo de informação relevante foram considerados úteis para o treino dos modelos.

O fluxo desta etapa está representado na Figura 14, enquanto a Tabela 11 resume os principais efeitos do processo de limpeza.

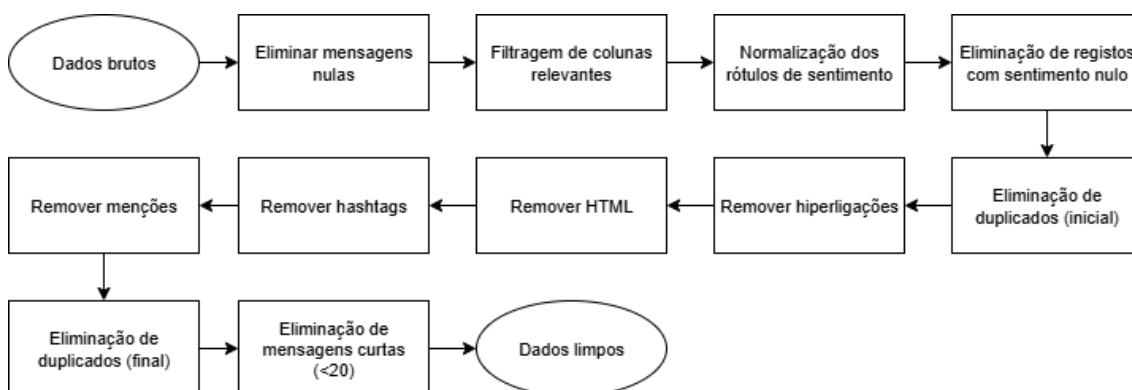


Figura 14 - Representação do fluxo de limpeza inicial dos dados. Fonte: Elaboração própria.

Tabela 11 - Comparação dos principais indicadores antes e depois da limpeza dos dados. Fonte: Elaboração própria.

Métrica	Antes	Depois
Número total de registos	1 144 408	295 617
Mensagens Bullish	316 366	191 752
Mensagens Bearish	164 068	103 865
Mensagens com sentimento nulo	663 974	0
Mensagens duplicadas	9 748	0
Mensagens nulas	795	0

3.3.2 Amostragem

Nesta fase procedeu-se à amostragem aleatória do conjunto de dados processado na etapa anterior. Esta tarefa teve como principal objetivo reduzir o volume de dados para um tamanho computacionalmente aceitável, tendo em conta o número de execuções previstas. Para o efeito foi definido um limite de 20.000 comentários do *StockTwits*, valor baseado em estudos que adotaram amostras de dimensão semelhante. Por exemplo, Bozanta et al. (2021) construíram conjuntos de 20.000 mensagens por ação para avaliar modelos de aprendizagem computacional, aprendizagem profunda e modelos baseados em transformadores no domínio financeiro. O conjunto *SemEval-2013* para *Twitter Sentiment Analysis* inclui cerca de 20.000 mensagens anotadas (Saif et al., 2013), servindo de avaliação comparativa para muitas pesquisas. O estudo de Mosbach et al.

(2021) demonstra que, mesmo com conjuntos de dados pequenos, é possível obter resultados robustos desde que sejam aplicadas técnicas de treino adequadas.

A amostragem foi realizada de forma estratificada (preservando a distribuição original das classes de sentimento). A proporção da classe $c \in \{\text{bullish}, \text{bearish}\}$ na amostra estratificada p_c^{amostra} foi preservada de forma aproximada, mantendo a distribuição do conjunto original:

$$p_c^{\text{amostra}} \approx p_c^{\text{original}} = \frac{n_c}{n} \quad (1)$$

onde n_c representa o número de mensagens da classe c no conjunto original, e n é o número total de mensagens antes da amostragem.

Importa salientar que, para avaliações mais robustas e que não se limitem a uma mera comparação entre abordagens, mas que tenham como objetivo alcançar o desempenho máximo, principalmente no que diz respeito a modelos baseados em redes neuronais ou em transformadores, o número de registos necessário deverá ser consideravelmente superior. Estes modelos exigem volumes de dados muito maiores, já que dependem da deteção de padrões mais complexos. Segundo Alzubaidi et al. (2023), a escassez de dados é um dos principais desafios no treino de modelos de aprendizagem profunda, quando se pretende alcançar um desempenho máximo.

A Tabela 12 apresenta os principais indicadores antes e depois da amostragem aleatória dos dados do *StockTwits*.

Tabela 12 - Comparação dos dados após a limpeza inicial e após a amostragem aleatória. Fonte: Elaboração própria.

Métrica	Antes	Depois
Número total de registos	295 617	20 000
Mensagens Bullish	191 752	12 973
Mensagens Bearish	103 865	7 027
Número total de tokens	5 311 719	359 467

3.3.3 Partição dos dados

Após a amostragem dos dados, realizada na etapa anterior, estes foram divididos em três subconjuntos: treino, validação e teste. O objetivo desta partição foi garantir uma separação entre os dados usados para treinar os modelos, aqueles destinados à afinação

de hiper-parâmetros e critérios de paragem, e os reservados para a avaliação final do desempenho do modelo (Xu & Goodacre, 2018). A divisão foi feita de forma estratificada (preservando a proporção original das classes em todos os subconjuntos). Para cada conjunto de dados, o número de observações foi determinado segundo as proporções:

$$\alpha_{\text{train}} = 0,70, \quad \alpha_{\text{val}} = 0,10, \quad \alpha_{\text{test}} = 0,20 \quad (2)$$

onde α_k representa a fração atribuída ao subconjunto $k \in \{\text{train, val, test}\}$. Mantendo a estratificação das classes, o tamanho de cada subconjunto é então dado por:

$$n_k = \alpha_k n \quad (3)$$

Assim, $n_{\text{train}} = 0,70 \times 20\,000 = 14\,000$ (arredondado para 13 999 após estratificação), $n_{\text{val}} = 2\,000$ e $n_{\text{test}} = 4\,001$.

A Tabela 13 apresenta os principais indicadores dos subconjuntos gerados após a partição dos dados.

Tabela 13 - Estatísticas dos subconjuntos gerados após a partição dos dados. Fonte: Elaboração própria.

Métrica	Treino	Validação	Teste	Total
Número total de registos	13 999	2 000	4 001	20 000
Mensagens Bullish	9 080	1 297	2 596	12 973
Mensagens Bearish	4 919	703	1 405	7 027
Número total de tokens	251 202	36 126	72 139	359 467

3.3.4 Balanceamento dos dados

Nesta etapa aplicou-se a técnica de balanceamento com o objetivo de reduzir a assimetria evidente entre as classes presentes no conjunto de treino, já que poderia prejudicar o desempenho dos modelos, favorecendo a classe maioritária (Branco et al., 2017; Carvalho et al., 2025). Foram criadas duas versões do conjunto de treino, definidas como:

$$n_{\text{bull}}^{\text{train}} > n_{\text{bear}}^{\text{train}} \quad (4)$$

Define-se o conjunto subamostrado D_{USM} como:

$$D_{\text{USM}} = D_{\text{bear}}^{\text{train}} \cup D_{\text{bull}}^{\text{train}*} \quad (5)$$

Onde:

$$D_{\text{bull}}^{\text{train}*} \subset D_{\text{bull}}^{\text{train}} \quad \text{tal que} \quad |D_{\text{bull}}^{\text{train}*}| = |D_{\text{bear}}^{\text{train}}| \quad (6)$$

A primeira versão corresponde ao cenário com os dados originais, sem aplicação de qualquer técnica de balanceamento, identificada como "*NOB*". Nesta configuração, mantém-se a distribuição inicial das classes. A segunda versão, identificada como "*USM*", deu origem a uma redução aleatória do número de exemplos da classe majoritária (*Bullish*) para igualar ao da classe minoritária (*Bearish*).

A Tabela 14 apresenta uma comparação dos principais indicadores resultante da aplicação da técnica de balanceamento (subamostragem da classe majoritária).

Tabela 14 - Comparação após aplicação das técnicas de balanceamento no conjunto de treino. Fonte: Elaboração própria.

Métrica	Sem balanceamento (NOB)	Subamostragem da classe majoritária (USM)
Número total de registos	13 999	9 838
Mensagens Bullish	9 080	4 919
Mensagens Bearish	4 919	4 919
Número total de tokens	251 202	172 432

3.3.5 Técnicas de pré-processamento

Nesta etapa, foi efetuada uma seleção das técnicas de pré-processamento, definida com base em dois critérios principais: a frequência de utilização evidenciada na literatura e a viabilidade computacional disponível para explorar todas as combinações possíveis. Esta escolha seguiu fundamentação de trabalhos anteriores, como o de Siino et al. (2024), que identificam as técnicas mais utilizadas em análise de sentimento. Tendo por base este estudo, optou-se por três técnicas principais: (i) conversão para minúsculas (*LOW*), (ii) remoção de palavras sem significado semântico (*RSW*) e (iii) radicalização (*STM*). A estas acrescentou-se a opção de ausência de pré-processamento (*DON*).

A técnica de conversão para minúsculas (*lowercase*), consiste na transformação de todas as letras do texto para minúsculas (Siino et al., 2024). A Tabela 15 mostra um exemplo do efeito desta técnica.

Tabela 15 - Exemplo do efeito da técnica de conversão para minúsculas (LOW). Fonte: Elaboração própria.

Antes	Depois
Key News Update Apples anticipated event unveils new iPhones and Apple Watch customers expect new features	key news update apples anticipated event unveils new iphones and apple watch customers expect new features

A remoção de palavras sem significado semântico (*stopwords*), implica a eliminação de termos muito comuns que tendem a não acrescentar valor informativo relevante (Siino et al., 2024). A Tabela 16 apresenta um exemplo desta técnica.

Tabela 16 - Exemplo do efeito da técnica de eliminação de palavras sem significado semântico (RSW). Fonte: Elaboração própria.

Antes	Depois
how can investors do any due diligence or get an accurate picture of this company without subscriber data It is a freaking subscription streaming service A subscription streamer not providing subscription metrics	investors due diligence get accurate picture company without subscriber data freaking subscription streaming service subscription streamer providing subscription metrics

Por fim, a radicalização (*stemming*), é o processo de reduzir palavras derivadas à sua raiz comum (Siino et al., 2024). A Tabela 17 ilustra o efeito desta técnica.

Tabela 17 - Exemplo do efeito da técnica de radicalização (STM). Fonte: Elaboração própria.

Antes	Depois
This decline reminds e of when everyone thought ChatGPT and copilot AI will steal most of their seach and enterprise business only to realize Google isnt going anywhere let it decline and buy large rinse and repeat	thi declin remind e of when everyon thought chatgpt and copilot ai will steal most of their seach and enterpris busi onli to realiz googl isnt go anywher let it declin and buy larg rins and repeat

Pretendeu-se também avaliar o efeito das combinações das diversas técnicas, considerando todas as permutações possíveis quanto à ordem de aplicação. À semelhança do trabalho de Siino et al. (2024), foram geradas todas as permutações possíveis dessas três técnicas. A Tabela 18 apresenta um exemplo do efeito combinado das técnicas de eliminação de palavras com pouco significado semântico e radicalização. A Tabela 19

ilustra o efeito da aplicação das mesmas técnicas, mas na ordem inversa. A Tabela 20 apresenta um exemplo da aplicação sequencial das três técnicas.

Tabela 18 - Exemplo do efeito combinado das técnicas eliminação de palavras com pouco significado semântico (RSW) e radicalização (STM). Fonte: Elaboração própria.

Antes	Depois
JOLTS data was so so but Jobs data Friday will be a key stat to watch Consolidating nicely with RSI at before next move higher Will see sideways price action this week as traders vacay in Hamptons and machines take over trading with some junior traders tasked with keeping things moving until monday	jolt data job data friday key stat watch consolid nice rsi next move higher see sideways price action week trader vacay hampton machin take trade junior trader task keep thing move monday

Tabela 19 - Exemplo do efeito combinado das técnicas radicalização (STM) e remoção de palavras com pouco significado semântico (RSW). Fonte: Elaboração própria.

Antes	Depois
JOLTS data was so so but Jobs data Friday will be a key stat to watch Consolidating nicely with RSI at before next move higher Will see sideways price action this week as traders vacay in Hamptons and machines take over trading with some junior traders tasked with keeping things moving until monday	jolt data wa job data friday key stat watch consolid nice rsi befor next move higher see sideways price action thi week trader vacay hampton machin take trade junior trader task keep thing move monday

Tabela 20 - Exemplo do efeito combinado das técnicas de conversão para minúsculas (LOW) e remoção de palavras com pouco significado semântico (RSW) e radicalização (STM). Fonte: Elaboração própria.

Antes	Depois
no follow through after the gap up over resistance Still think there is one more blowoff top to before correction Market in neutral till CPI just killing theta on options so staying patient	follow gap resist still think one blowoff top correct market neutral till cpi kill theta option stay patient

Foram criadas 32 versões do conjunto de dados original, resultantes da aplicação de diferentes combinações de técnicas de pré-processamento sobre os dados balanceados e

não balanceados, 16 para cada cenário de balanceamento. De forma semelhante, Siino et al. (2024) geraram 16 versões do seu conjunto de dados. O fluxo de execução correspondente à etapa de preparação dos dados pode ser representado graficamente através da Figura 15.

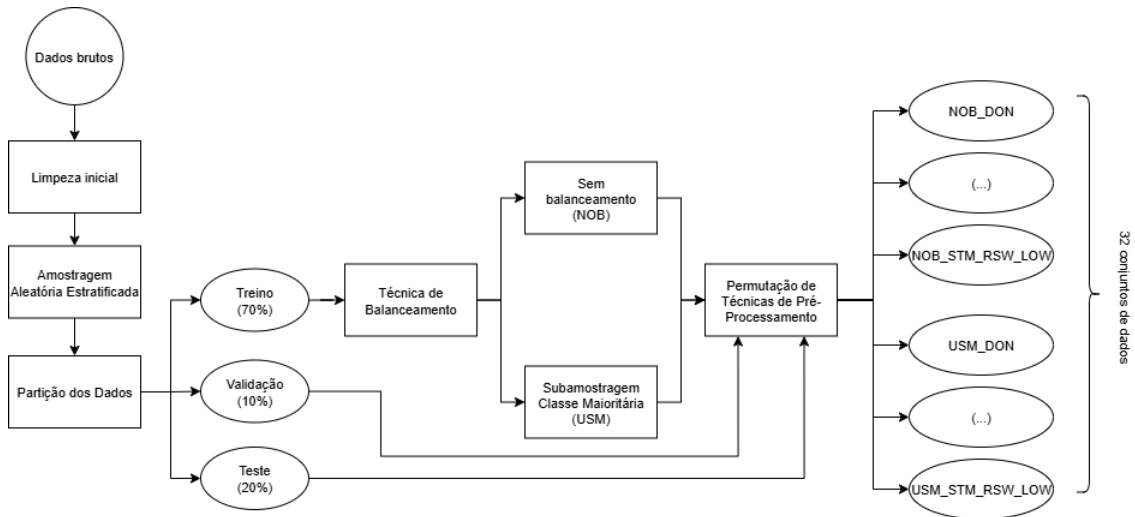


Figura 15 - Representação do fluxo de preparação dos dados. Fonte: Elaboração própria.

3.4 Modelação

O processo de modelação envolveu a aplicação de diferentes abordagens de análise de sentimentos aos comentários do *StockTwits*. Foram exploradas quatro grandes categorias: analisadores baseados em léxicos, algoritmos de aprendizagem computacional tradicional, redes neurais e modelos baseados em transformadores.

3.4.1 Analisadores baseados em léxicos

A seleção destes analisadores teve como critério garantir diversidade metodológica, abrangendo ferramentas genéricas, como *TextBlob*⁴, *AFINN* (Nielsen, 2011) e *SentiWordNet* (Sebastiani et al., 2010), especializadas em redes sociais, como o *VADER* (C. J. Hutto & Gilbert, 2014) e orientadas para o domínio financeiro, como o *FinVADER*⁵, bem como a sua relevância na literatura científica, conforme evidenciado nos estudos de Denecke & Reichenpfader (2023), Mehenaoui et al. (2024) e Ribeiro et al. (2015). A Tabela 21 apresenta um resumo das principais características de cada ferramenta.

⁴ <https://textblob.readthedocs.io/en/dev/index.html>

⁵ <https://github.com/PetrKorab/FinVADER>

Tabela 21 - Resumo das ferramentas de análise de sentimento baseadas em léxicos. Fonte: Elaboração própria.

Ferramenta	Descrição
<i>VADER</i>	Ferramenta baseada em regras e léxico, otimizada para identificar sentimentos em textos de redes sociais.
<i>SentiWordNet</i>	Recurso lexical que atribui a cada <i>synset</i> do <i>WordNet</i> três pontuações: positividade, negatividade e objetividade.
<i>AFINN</i>	Lista de palavras com valência classificada por valores inteiros de -5 (negativo) a +5 (positivo).
<i>TextBlob</i>	Biblioteca <i>Python</i> com interface simples para tarefas comuns de NLP, como marcação gramatical, extração de grupos nominais, análise de sentimento e classificação.
<i>FinVADER</i>	Variante do <i>VADER</i> que incorpora dois léxicos financeiros: <i>SentiBignomics</i> (cerca de 7300 termos com polaridade em [-1, 1]) e lista de <i>Henry</i> (189 palavras frequentes em comunicados financeiros).

Cada um destes métodos foi executado sobre as 32 versões de conjuntos de dados preparados na etapa anterior. A avaliação foi feita no conjunto de teste correspondente a 20% dos dados. A Figura 16 ilustra o fluxo de execução da aplicação de analisadores baseados em léxicos.



Figura 16 - Representação do fluxo de processamento da categoria de analisadores baseados em léxicos. Fonte: Elaboração própria.

3.4.2 Modelos de aprendizagem computacional tradicional

Esta seleção teve como objetivo combinar diferentes classificadores clássicos com várias abordagens de representação textual. A escolha dos classificadores teve como critério a sua relevância na literatura, conforme evidenciado nos estudos de Medhat et al. (2014) e Tan et al. (2023), e a diversidade, incluindo modelos probabilísticos como o *Naïve Bayes*

(Narayanan et al., 2013), lineares como o *Support Vector Machines* e *Logistic Regression* (Pang et al., 2002), baseados em árvores como *Decision Trees* (Sreenivas et al., 2023), e técnicas de conjunto, como o *LightGBM* e *XGBoost* (Mishra, 2021).

No que respeita às técnicas de representação textual, pretendeu-se testar abordagens baseadas na frequência de termos, como o *BOW* e *TF-IDF* (Bhargavi, 2025), e métodos mais contextuais, como *Word2Vec* e *FastText* (Khomsah et al., 2022; Lu, 2025; Zhan, 2025).

A Tabela 22 e a Tabela 23 apresentam, respetivamente, um resumo dos algoritmos de aprendizagem computacional tradicional e das técnicas de representação textual utilizados neste estudo.

Tabela 22 - Resumo dos algoritmos de aprendizagem computacional tradicional. Fonte: Elaboração própria.

Algoritmo	Descrição
<i>Decision Tree</i>	Estrutura em forma de diagrama de fluxo, em que cada nó interno representa um teste a um atributo, cada ramo indica o resultado desse teste e cada nó terminal corresponde a uma classe ou decisão final.
<i>Naive Bayes</i>	Conjunto de algoritmos de aprendizagem supervisionada baseados no teorema de <i>Bayes</i> , assumindo independência condicional entre cada par de atributos dado o valor da variável de classe.
<i>Logistic Regression</i>	Técnica estatística de classificação que estima a probabilidade de ocorrência de uma classe com base em variáveis explicativas contínuas e/ou binárias, utilizando a função logística (<i>sigmóide</i>). A probabilidade estimada é posteriormente convertida num rótulo discreto.
<i>SVM</i>	Algoritmo supervisionado para classificação e regressão, que procura encontrar o hiperplano que melhor separa os dados de diferentes classes.
<i>LightGBM</i>	Biblioteca de <i>gradient boosting</i> distribuído baseada em árvores de decisão, desenvolvida pela Microsoft, otimizada para alto desempenho e escalabilidade em tarefas como classificação, ordenação e regressão.
<i>XGBoost</i>	Biblioteca otimizada de <i>gradient boosting</i> distribuído que implementa construção paralela de árvores de decisão (<i>GBDT/GBM</i>), oferecendo alta eficiência e precisão em problemas de ciência de dados.

Tabela 23 - Resumo das técnicas de representação textual. Fonte: Elaboração própria.

Técnica	Descrição
<i>Bag of Words (BOW)</i>	Modelo de representação textual que utiliza uma coleção não ordenada (“ <i>saco</i> ”) de palavras. Ignora a ordem e a estrutura gramatical, mas registra a frequência das ocorrências. Amplamente usado em processamento de linguagem natural e recuperação de informação.
<i>Term Frequency–Inverse Document Frequency (TF-IDF)</i>	Medida estatística que avalia a importância de uma palavra num documento relativamente a um conjunto de documentos (<i>corpus</i>). Muito utilizada em processamento de linguagem natural e recuperação de informação.
<i>Word2Vec</i>	Técnica que gera representações vetoriais de palavras a partir do contexto em que ocorrem. Captura relações semânticas e é capaz de identificar sinónimos ou sugerir palavras relacionadas.
<i>FastText</i>	Biblioteca desenvolvida pelo <i>Facebook AI Research</i> para criar representações vetoriais de palavras e classificação de texto. Suporta aprendizagem supervisionada e não supervisionada, com modelos pré-treinados para 294 línguas.

Nas situações em que foram utilizadas representações esparsas baseadas em frequência de palavras (*Bag of Words* e *TF-IDF*), foi realizada uma pesquisa em grelha (*grid search*, que consiste em testar todas as combinações possíveis de valores (Bergstra et al., 2012)), sobre diferentes combinações de *n*-gramas (*unigramas* e *bigramas*), avaliadas no conjunto de validação. Um *n*-grama corresponde a uma sequência de *n* unidades linguísticas consecutivas. Os *unigramas* ($n = 1$) representam palavras isoladas, enquanto os *bigramas* ($n = 2$) correspondem a pares de palavras consecutivas (C. Manning & Schütze, 1999). Para os modelos que recorrem a representações densas, isto é, representações vetoriais contínuas, distribuídas e de comprimento fixo (Almeida & Xexéo, 2023), em que cada palavra é representada por valores reais, como são o caso do *Word2Vec* e *FastText* (Khomsah et al., 2022), cada mensagem foi representada pela média dos vetores das suas palavras.

Ambos os modelos de representação densa foram treinados a partir do *corpus* rotulado neste estudo, com vetores de dimensão 300, baseado no estudo de Mikolov et al. (2013), janela de contexto de 5 palavras, frequência mínima de 5 ocorrências por palavra (permite eliminar termos raros e pouco informativos, reduzindo o ruído e o custo computacional) e arquitetura *skip-gram* (arquitetura que tenta prever as palavras vizinhas a partir de uma palavra-alvo, dentro de uma janela de contexto definida (Mikolov et al., 2013)), ao longo de 10 épocas. A Tabela 24 apresenta um resumo dos principais parâmetros utilizados no treino destes modelos.

Tabela 24 - Parâmetros utilizados no treino dos modelos de representação densa. Fonte: Elaboração própria.

Parâmetro	Word2Vec	FastText
Dimensão dos vetores	300	300
Janela de contexto	5	5
Frequência mínima	5	5
Arquitetura	<i>Skip-gram</i>	<i>Skip-gram</i>
Número de épocas	10	10

Foi realizada otimização de hiperparâmetros dos classificadores no conjunto de validação. Esta técnica permite avaliar exaustivamente todas as combinações possíveis de valores (Bischi et al., 2021). A Tabela 25 apresenta os hiperparâmetros e respectivos intervalos de valores na grelha de otimização (*grid search*) para os classificadores. Já a Tabela 26 resume os parâmetros aplicados nos métodos *BoW* e *TF-IDF*.

A Figura 17 apresenta o fluxo de processamento desenvolvido para a aplicação dos modelos de aprendizagem computacional tradicional neste trabalho.

Tabela 25 - Hiperparâmetros utilizados nos classificadores de aprendizagem computacional. Fonte: Elaboração própria.

Classificador	Hiperparâmetros	Parâmetros adicionais
<i>Decision Tree</i>	criterion = ["gini", "entropy"], max_depth = [5, 10, None] min_samples_split = [2, 5] class_weight = [None, "balanced"]	Pesquisa de ngram_range: (1,1) e (1,2)
<i>Naive Bayes (Multinomial)</i>	alpha = [0.01, 0.1, 0.5, 1.0] fit_prior = [True, False]	Representações esparsas Pesquisa de ngram_range: (1,1) e (1,2)
<i>Naive Bayes (Gaussian)</i>	N/A	Representações densas Sem hiperparâmetros configuráveis

<i>Logistic Regression</i>	C = [0.01, 0.1, 1.0, 10.0] penalty = ["l2"] fit_intercept = [True, False]	solver = "lbfgs", max_iter = 1000 class_weight = "balanced" Pesquisa de ngram_range: (1,1) e (1,2)
<i>Support Vector Machine</i>	C = [0.01, 0.1, 1.0, 10.0] loss = ["hinge", "squared_hinge"] fit_intercept = [True, False]	Pesquisa de ngram_range: (1,1) e (1,2)
<i>LightGBM</i>	num_leaves = [15, 31] learning_rate = [0.05, 0.1] max_depth = [5, 10]	n_estimators = 200, class_weight = "balanced" Pesquisa de ngram_range: (1,1) e (1,2)
<i>XGBoost</i>	n_estimators = [100, 200] learning_rate = [0.05, 0.1] max_depth = [5, 10]	eval_metric = "logloss" Pesquisa de ngram_range: (1,1) e (1,2)

Tabela 26 - Parâmetros utilizados nos vetorizadores textuais. Fonte: Elaboração própria.

Vetorizador	Parâmetros principais
<i>Bag of Words</i>	max_features = 1000, min_df = 10, max_df = 0.8
<i>TF-IDF</i>	max_features = 1000, min_df = 10, max_df = 0.8

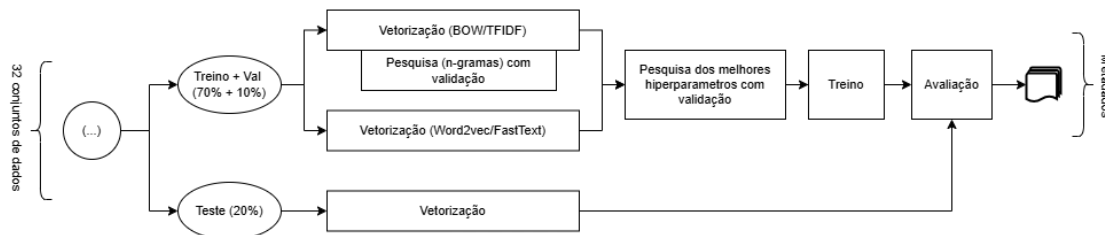


Figura 17 - Representação do fluxo de processamento da categoria de modelos de aprendizagem computacional. Fonte: Elaboração própria.

3.4.3 Modelos baseados em redes neurais

A seleção destes modelos teve como objetivo garantir diversidade de arquiteturas com capacidades de modelação distintas. As arquiteturas de *perceptron* multicamada não sequencial, como o *MLP* (Rumelhart et al., 1986) constituem um modelo de referência, as arquiteturas com capacidades sequenciais, como o *GRU* (Cho et al., 2014) e *BiLSTM* (Graves & Schmidhuber, 2005) permitem modelar dependências temporais e relações entre palavras em sequência, as arquiteturas com capacidades de extração local, como as *CNN* (Lecun et al., 1998) identificam padrões relevantes em pequenas janelas de texto, e

a arquitetura híbrida *CNN-BiLSTM* (Bharal & V Vamsi Krishna, 2021) combina ambas as abordagens para capturar simultaneamente padrões locais e dependências de longo alcance, sendo todas estas variantes amplamente reconhecidas na literatura, conforme evidenciado nos estudos de Aftab et al. (2023), Islam et al. (2024), Mabrouk et al. (2020), Tan et al. (2023) e L. Zhang et al. (2018). A Tabela 27 apresenta um resumo das arquiteturas de redes neuronais exploradas neste estudo.

Tabela 27 - Resumo das arquiteturas de redes neuronais utilizadas. Fonte: Elaboração própria.

Arquitetura	Descrição
<i>Multilayer Perceptron (MLP)</i>	Rede neuronal <i>feed-forward</i> (fluxo unidirecional, da entrada até à saída, sem ciclos de retroalimentação) composta por neurónios totalmente conectados com funções de ativação não lineares, organizados em camadas. Capaz de distinguir dados não linearmente separáveis.
<i>Gated Recurrent Unit (GRU)</i>	Variante das redes <i>LSTM</i> que utiliza mecanismos de portas para controlar a entrada e esquecimento de informação, mas sem célula de memória separada, resultando em menos parâmetros.
<i>Bidirectional Long Short-Term Memory (BiLSTM)</i>	Rede neuronal recorrente que ultrapassa limitações das <i>RNN</i> tradicionais, lidando com o desvanecimento do gradiente e captando dependências de longo prazo em ambos os sentidos (passado/futuro e futuro/passado).
<i>Convolutional Neural Networks (CNNs)</i>	Redes <i>feed-forward</i> que aprendem a extrair características através de filtros otimizados. Muito usadas em processamento de imagem, visão computacional e, mais recentemente, texto, embora atualmente muitas vezes substituídas por transformadores.
<i>CNN-BiLSTM</i>	Arquitetura híbrida que combina <i>CNN</i> para extração local de características com <i>BiLSTM</i> para modelação de dependências sequenciais nos dois sentidos temporais.

Os comentários foram convertidos em representações densas a partir de vetores pré-treinados (*Word2Vec* e *FastText*). Considerando a diversidade de modelos desenvolvidos foram realizadas duas abordagens distintas para a conversão das mensagens em representações numéricas. Para o modelo não sequencial *MLP*, que não capta dependências temporais entre palavras, utilizou-se a média dos vetores. Para os restantes

modelos efetuaram-se sequências com preenchimento (*padding*) de forma a preservar a ordem e garantir o mesmo comprimento entre entradas.

O conjunto de dados de validação serviu para regularização e controlo do sobreajustamento (*overfitting*: situação em que o modelo aprende demasiado bem os exemplos do conjunto de treino, ficando incapaz de generalizar para novos dados), com o mecanismo de paragem antecipada (*early stopping*: técnica de regularização que interrompe o treino de um modelo quando este deixa de melhorar no conjunto de validação), com paciência de três épocas e um máximo de vinte épocas de treino. Esta configuração foi escolhida por se tratar de modelos com requisitos computacionais significativamente inferiores aos transformadores, o que possibilitou treinos mais prolongados sem comprometer a eficiência.

A Tabela 28 apresenta um resumo dos principais hiperparâmetros utilizados no treino dos modelos baseados em redes neuronais. A Figura 18 ilustra o fluxo de processamento desenvolvido para esta fase do estudo.

Tabela 28 - Hiperparâmetros dos modelos baseados em redes neuronais. Fonte: Elaboração própria.

Modelo	Dim. Oculta	Camadas	Dropout	Tam. Batch	Ép./ Pac.	Otimiz.	Tx. Aprend.
MLP	64	2	0,3	32	20 / 3	AdamW	0,001
BiLSTM	64	2	0,3	32	20 / 3	AdamW	0,001
GRU	64	1	0,3	32	20 / 3	AdamW	0,001
CNN	100 filt. (3,4,5)	3 filt.	0,3	32	20 / 3	AdamW	0,001
CNN-BiLSTM	100 + 64 LSTM	2 LSTM	0,3	32	20 / 3	AdamW	0,001

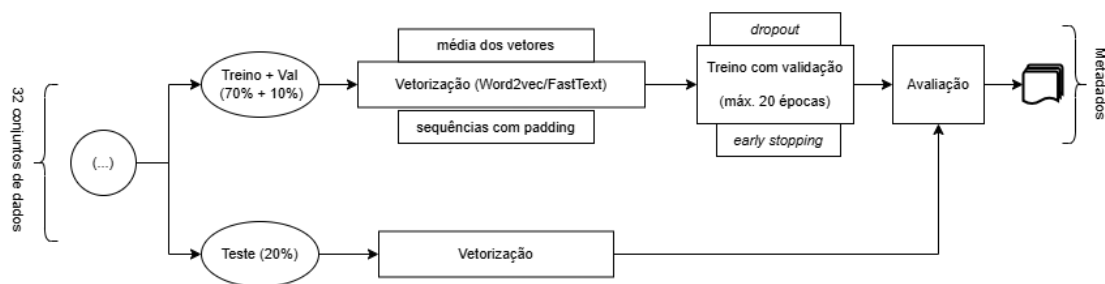


Figura 18 - Representação do fluxo de processamento da categoria de modelos baseados em redes neuronais. Fonte: Elaboração própria.

3.4.4 Modelos baseados em transformadores

A escolha recaiu sobre arquiteturas disponibilizadas na biblioteca *Hugging Face*⁶, pela sua simplicidade de utilização e pela possibilidade de incorporar diversidade metodológica e diferentes perspectivas: modelos fundacionais pré-treinados em dados genéricos, como o *BERT* (Devlin et al., 2018), versões derivadas mais leves, como o *DistilBERT* (Sanh et al., 2019), modelos especializados em redes sociais, como o *BERTweet* (Nguyen et al., 2020) e modelos adaptados ao domínio financeiro, como o *RoBERTaStockTwits* (F. Cao et al., 2022) e *FinBERT* (Araci, 2019), e com relevância na literatura, conforme evidenciado nos estudos de Casola et al. (2022), Min et al. (2021), Qiu et al. (2020) e Rogers et al. (2020).

O processo incluiu o ajuste fino (*fine-tuning*) destes cinco modelos, ou seja, a adaptação dos modelos ao conjunto de dados deste estudo.

A comparação entre modelos já previamente ajustados ao domínio com modelos genéricos, sem ajuste, permite avaliar em que medida a afinação de modelos pré-treinados com um número limitado de registos pode ser suficiente para os tornar competitivos em tarefas específicas. Mesmo no caso dos modelos que tinham sido previamente ajustados a domínios semelhantes (financeiro e redes sociais), optou-se por realizar afinação adicional para adequá-los ao *corpus* disponível.

A Tabela 29 apresenta um resumo dos modelos baseados em transformadores utilizados neste estudo.

Tabela 29 - Resumo dos modelos baseados em transformadores utilizados. Fonte: Elaboração própria.

Modelo	Descrição
<i>BERT</i>	Transformador bidirecional pré-treinado com texto não rotulado, que prevê <i>tokens</i> ocultos e identifica se uma frase sucede logicamente a outra. Consiste em mascarar aleatoriamente as palavras, aprendendo com base no contexto à esquerda e à direita. Pode ser adaptado a outras tarefas de processamento de linguagem natural através de ajuste fino.

⁶ <https://huggingface.co/>

<i>DistilBERT</i>	Versão reduzida do <i>BERT</i> obtida por destilação de conhecimento, com menor dimensão, inferência mais rápida e menor necessidade de recursos computacionais, preservando boa parte do desempenho do modelo original.
<i>RoBERTaStockTwits</i>	Versão afinada do <i>RoBERTa-base</i> , treinada especificamente para análise de sentimentos em textos financeiros da plataforma <i>StockTwits</i> . Afinada com 3,2 milhões de comentários.
<i>BERTweet</i>	Primeiro modelo linguístico de grande escala pré-treinado para <i>tweets</i> em inglês, seguindo o procedimento de pré-treino do <i>RoBERTa</i> . Treinado com 850 milhões de <i>tweets</i> (cerca de 16 mil milhões de <i>tokens</i> , 80GB).
<i>FinBERT</i>	Modelo <i>BERT</i> ajustado ao domínio financeiro, treinado para análise de sentimentos em textos financeiros. Utiliza vasto <i>corpus</i> especializado e afinado com o <i>Financial PhraseBank</i> .

A Tabela 30 apresenta um resumo das principais características da arquitetura dos diversos modelos.

Tabela 30 - Resumo dos parâmetros dos modelos baseados em transformadores. Fonte: Elaboração própria a partir dos dados da plataforma Hugging Face .

Modelo	Camadas Ocultas	Cabeças Atenção	Tam. Oculto	Nº de Parâm.
<i>bert-base-uncased</i>	12	12	768	110M
<i>vinai/bertweet-base</i>	12	12	768	135M
<i>distilbert-base-uncased</i>	6	12	768	66M
<i>ProsusAI/finbert</i>	12	12	768	110M
<i>roberta-base-stocktwits-finetuned</i>	12	12	768	125M

Os textos foram transformados em sequências numéricas correspondentes a unidades linguísticas (*tokens*) limitadas a um comprimento máximo de 128 (dado o reduzido comprimento das mensagens no conjunto de treino). O treino foi efetuado com a biblioteca *Hugging Face Transformers*⁷, realizando avaliação em três épocas e com paragem antecipada (*early stopping*) se não ocorresse melhoria no desempenho durante uma época consecutiva. A escolha do número de épocas, tamanho do lote (*batch size*) e

⁷ <https://github.com/huggingface/transformers>

da taxa de aprendizagem (*learning rate*) baseou-se no estudo de Devlin et al. (2018), tendo ainda sido tido em linha de conta a elevada exigência computacional dos modelos baseados em transformadores e o número de execuções em múltiplos conjuntos de dados, que impõe a necessidade de limitar o número de épocas de treino.

A Tabela 31 apresenta os principais parâmetros utilizados no processo de ajuste fino dos modelos baseados em transformadores.

Tabela 31 - Parâmetros de ajuste fino dos modelos baseados em transformadores. Fonte: Elaboração própria.

Modelo	Tam. Do Batch	Épocas (máx.)	Comp. Seq.	Acum. grad.	Otimiz.	Tx. De Aprend.
<i>BERT</i>	32	3	128	2	AdamW	2e-5
<i>BERTweet</i>	32	3	128	2	AdamW	2e-5
<i>FinBERT</i>	32	3	128	2	AdamW	2e-5
<i>DistilBERT</i>	32	3	128	2	AdamW	2e-5
<i>RoBERTaStockTwits</i>	32	3	128	2	AdamW	2e-5

A Figura 19 ilustra o fluxo de processamento.

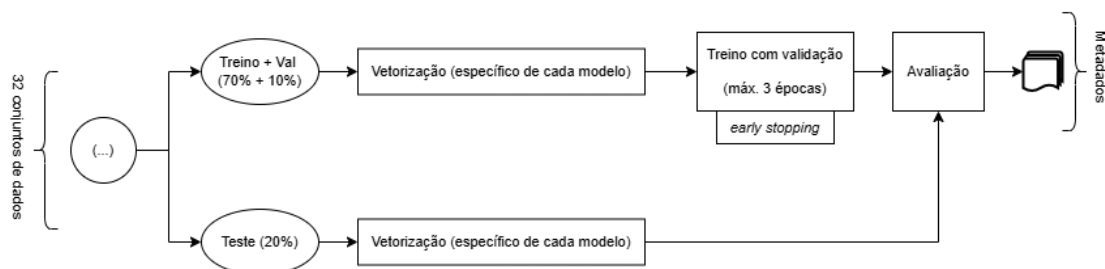


Figura 19 - Representação do fluxo de processamento da categoria de modelos baseados em transformadores. Fonte: Elaboração própria.

3.5 Avaliação

A avaliação dos modelos foi realizada com base no conjunto de teste, e foram recolhidas, em cada processamento, as métricas apresentadas na Tabela 32. Estas métricas são frequentemente utilizadas na literatura (Sathyanarayanan, 2024).

Tabela 32 - Resumo das métricas de avaliação utilizadas. Fonte: Elaboração própria.

Métrica	Descrição	Fórmula
Matriz de Confusão	Estrutura matricial que resume o desempenho do modelo através das contagens de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN).	$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$
Exatidão (<i>Accuracy</i>)	Proporção de previsões corretas relativamente ao total de exemplos avaliados.	$\frac{TP + TN}{TP + TN + FP + FN}$
Precisão (<i>Precision</i>)	Proporção de exemplos classificados como positivos que são, de facto, positivos.	$\frac{TP}{TP + FP}$
Sensibilidade (<i>TPR</i>)	Proporção de exemplos positivos corretamente identificados pelo modelo.	$\frac{TP}{TP + FN}$
Especificidade (<i>TNR</i>)	Proporção de exemplos negativos corretamente classificados.	$\frac{TN}{TN + FP}$
Medida-F1 (<i>F1-score</i>)	Média harmónica entre precisão e sensibilidade, equilibrando ambas as métricas.	$2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

Com o objetivo de aumentar a fiabilidade estatística dos resultados, a execução dos diversos fluxos de processamento foi repetida três vezes, utilizando diferentes sementes de aleatoriedade. Esta prática segue a recomendação de Nunez et al. (2021), que sugerem o uso de múltiplas sementes para aumentar a robustez e a clareza dos resultados.

As principais métricas (exatidão, precisão, sensibilidade, especificidade, medida F1, tempo de execução e utilização de memória GPU) foram posteriormente agregadas por média, sendo calculado o respetivo desvio padrão.

Outra das avaliações presentes neste estudo será analisar a sensibilidade dos modelos a três fatores: (i) técnicas de balanceamento, comparando a diferença absoluta entre as exatidões máximas obtidas nos cenários sem balanceamento e com subamostragem da classe maioritária; (ii) técnicas de pré-processamento, calculando a diferença entre a maior e a menor exatidão máxima alcançadas pelo modelo quando aplicadas diferentes transformações textuais, para cada cenário de balanceamento; (iii) representações textuais, medindo a variação entre a melhor e a pior exatidão máxima obtidas para cada modelo quando utilizadas diferentes formas de representação do texto, para cada cenário de balanceamento. Em todos as situações, a variação é expressa como a diferença

absoluta entre os valores máximo e mínimo de exatidão máxima para o modelo em questão. A métrica escolhida para esta análise de sensibilidade foi a exatidão, por ser a mais frequentemente adotada nos estudos de análise de sentimento (Bordoloi & Biswas, 2023; Giachanou & Crestani, 2017), e pela necessidade de utilizar uma métrica padronizada que assegure comparabilidade entre as abordagens nos diversos cenários testados. Siino et al. (2024) realizaram uma análise de sensibilidade de diversos modelos de classificação, comparando a variação da exatidão média, entre o melhor e pior resultado, após a aplicação de diferentes técnicas de pré-processamento, isoladas ou em combinações. A análise de sensibilidade é uma abordagem utilizada para avaliar o impacto de diferentes condições no desempenho de modelos (Ankenbrand et al., 2021; Dimopoulos & Bakas, 2019). Permite estudar como a incerteza na saída de um modelo pode ser atribuída a diferentes fontes de incerteza nas suas entradas (Iooss & Lemaître, 2014; Saltelli et al., 2021), fornecendo uma medida direta do impacto de cada fator no desempenho global.

Foi também analisada a eficiência computacional dos modelos baseados em redes neurais e transformadores. Mediu-se a utilização média da memória da GPU (*VRAM*) durante o treino, calculada como a média dos valores registados em todas as execuções de cada modelo.

Apesar do elevado desempenho, os modelos baseados em transformadores exigem muito mais *VRAM* no ajuste fino do que redes convolucionais ou recorrentes, dificultando a afinação mesmo em GPUs potentes (Ardakani et al., 2024).

3.6 Implementação

A segunda fase do trabalho corresponde à etapa de Implementação do processo *CRISP-DM*. O objetivo desta fase consiste em utilizar o modelo com o melhor desempenho na fase de Modelação, para rotular os comentários do *StockTwits* que ainda não possuíam classificação.

De forma a analisar a evolução do sentimento em conjunto com o mercado, os comentários rotulados foram agregados em séries temporais de sentimento diário. A construção destas séries permite acompanhar a evolução do sentimento ao longo do tempo e compará-la diretamente com a evolução dos preços das ações. A construção destas

séries temporais segue, com adaptações, a lógica de Bollen et al. (2011). Neste estudo, os autores agregaram mensagens do *Twitter* por dia e calcularam a proporção de mensagens positivas. Posteriormente compararam com a evolução média do preço das ações. Esta análise tem por objetivo perceber se existe uma relação entre o sentimento e o comportamento de mercado, sendo efetuado em duas etapas distintas.

Na primeira etapa, aplica-se o algoritmo *Dynamic Time Warping (DTW)*, um algoritmo amplamente reconhecido em diversas áreas (Senin, 2008) para determinar, para cada ação, o desfasamento temporal ótimo entre o sentimento com a evolução do preço.

A utilização do algoritmo *Dynamic Time Warping* proposta neste trabalho implementa a abordagem de Y. Zhang et al. (2023). Os autores aplicaram o DTW para identificar desfasamentos temporais em séries financeiras. O método permite comparar séries mesmo quando não estão alinhadas no tempo, determinando assim um atraso ótimo.

Numa segunda etapa, e com base naquele desfasamento ótimo, realiza-se uma análise do alinhamento entre variações direcionais de sentimento e de preço, utilizando janelas deslizantes. A significância estatística destes alinhamentos é avaliada com um teste de hipóteses não paramétrico baseado em permutação aleatória. De acordo com Holt & Sullivan (2023), um teste de permutação avalia a hipótese nula comparando os resultados observados com a distribuição gerada por permutações aleatórias, permitindo calcular o *p-valor* sem assumir uma distribuição específica dos dados.

3.6.1 Seleção do melhor modelo e classificação dos comentários

O melhor modelo, definido pelo valor mais elevado de exatidão média, é aplicado a todo o corpus do *StockTwits*, atribuindo rótulos aos comentários ainda não classificados.

$$y_{\text{tweet}} \in \{bullish, bearish\} \quad (7)$$

3.6.2 Construção das séries temporais de sentimento agregado

O objetivo deste passo é transformar os comentários do *StockTwits* e das notícias financeiras em séries temporais diárias por ação, composta por duas variáveis principais: (i) o índice de sentimento agregado e (ii) o número total de comentários/notícias.

3.6.2.1 Agregação diária dos comentários

Para cada ação i e para cada dia civil t , calcula-se:

$$n_{i,t}^{\text{bull}} = |\{\text{comentários } \textit{bullish}\}| \quad (8) \quad n_{i,t}^{\text{bear}} = |\{\text{comentários } \textit{bearish}\}| \quad (9)$$

$$C_{i,t}^{\text{tw}} = n_{i,t}^{\text{bull}} + n_{i,t}^{\text{bear}} \quad (10)$$

3.6.2.2 Índice de sentimento agregado dos comentários

O sentimento diário dos comentários é medido pela proporção de comentários *bullish* face ao total de comentários do dia:

$$S_{i,t}^{\text{tw}} = \frac{n_{i,t}^{\text{bull}}}{C_{i,t}^{\text{tw}}} \in [0,1] \quad (11)$$

Este índice representa o grau de otimismo presente nos comentários para a ação i no dia t , onde:

- $S_{i,t}^{\text{tw}} = 1$: todos os comentários são *bullish*;
- $S_{i,t}^{\text{tw}} = 0,5$: há equilíbrio entre *bullish* e *bearish*;
- $S_{i,t}^{\text{tw}} = 0$: todos os comentários são *bearish*.

Nos dias em que possam não existir comentários para determinada ação ($C_{i,t}^{\text{tw}} = 0$), o valor do índice de sentimento é preenchido com o valor observado no dia útil imediatamente anterior (no caso dos comentários, todas as ações tiveram comentários em todos os dias).

3.6.2.3 Agregação diária das notícias

De forma análoga ao que foi realizado para os comentários do *StockTwits*, o conjunto de notícias financeiras é processado para obter, por ação i e por dia t :

$$n_{i,t}^{\text{bull}} = |\{\text{notícias } \textit{bullish}\}| \quad (12) \quad n_{i,t}^{\text{bear}} = |\{\text{notícias } \textit{bearish}\}| \quad (13)$$

$$C_{i,t}^{\text{nw}} = n_{i,t}^{\text{bull}} + n_{i,t}^{\text{bear}} \quad (14)$$

3.6.2.4 Índice de sentimento agregado das notícias

$$S_{i,t}^{\text{nw}} = \frac{n_{i,t}^{\text{bull}}}{C_{i,t}^{\text{nw}}} \in [0,1] \quad (15)$$

Esta métrica representa a dominância do sentimento positivo nas notícias para a ação i na data t , interpretando-se da seguinte forma:

- $S_{i,t}^{nw} = 1$: todas as notícias são *bullish*;
- $S_{i,t}^{nw} = 0,5$: há equilíbrio entre *bullish* e *bearish*;
- $S_{i,t}^{nw} = 0$: todas as notícias são *bearish*.

Nos casos em que possam não existir notícias para uma determinada ação no dia t ($C_{i,t}^{nw} = 0$), o valor do índice de sentimento é preenchido com o valor observado no dia útil imediatamente anterior (ocorreu em 93 dias).

3.6.2.5 Construção da série de preços

Para cada ação i e para cada dia civil t no intervalo temporal considerado, construiu-se uma série contínua de preços com base no preço de fecho ajustado ($P_{i,t}$):

$$P_{i,t} = \text{Preço de fecho ajustado} \quad (16)$$

Os dias sem preços de fecho são preenchidos por propagação do último valor disponível anterior para cada ação (ocorreu em 113 dias). Os preços foram normalizados por ação com recurso à normalização *Min-Max* (técnica que altera a escala de um atributo e desloca os seus valores de forma que fiquem compreendidos no intervalo $[0,1]$) (de Amorim et al., 2023).

$$P_{i,t}^{norm} = \frac{P_{i,t} - \min(P_i)}{\max(P_i) - \min(P_i)} \in [0,1] \quad (17)$$

Após a aplicação da normalização *Min-Max*, os valores da série $P_{i,t}^{norm}$ assumem o seguinte significado:

- $P_{i,t}^{norm} = 1$: corresponde ao maior preço de fecho ajustado;
- $P_{i,t}^{norm} = 0,5$: corresponde a um preço de fecho ajustado situado entre o mínimo e o máximo;
- $P_{i,t}^{norm} = 0$: corresponde ao menor preço de fecho ajustado.

3.6.3 Detecção do desfasamento ótimo

Nesta etapa procura-se identificar, para cada ação, o desfasamento ótimo, tanto para os comentários τ_i^{tw} como para as notícias τ_i^{nw} . O método adotado baseia-se no algoritmo *Dynamic Time Warping*. Este algoritmo permite comparar séries temporais mesmo quando não estão alinhadas. Para enquadrar este processo, consideram-se os seguintes elementos:

- a) **Teste de desfasamento:** De acordo com John & Ferbinteanu (2021), o conceito de desfasamento (*lag*) refere-se ao atraso temporal entre duas séries temporais, representando o atraso com que uma série reage em relação à outra. De acordo com Y. Zhang et al. (2023), o conceito de desfasamento corresponde ao número de períodos de atraso entre uma série temporal e outros fatores explicativos. Naquele estudo é aplicado o *DTW* para identificar atrasos entre séries temporais financeiras. Neste trabalho avaliou-se se o sentimento tem impacto imediato ou se os seus efeitos surgem com desfasamento face à variação dos preços. Sendo assim, para cada ação, avaliam-se três cenários: sentimento no próprio dia ($\tau = 0$), com um dia de antecedência ($\tau = 1$) ou com dois dias de antecedência ($\tau = 2$).
- b) **Aplicação de janelas deslizantes:** A utilização de janelas deslizantes permite modelar características estatísticas de nível micro presentes nas séries temporais (Tao et al., 2023), que poderiam vir a ficar diluídas numa análise realizada apenas sobre a série completa. Sendo assim, para cada valor de τ , desloca-se a série de sentimento $S_{i,t-\tau}$. Em cada janela deslizante, calcula-se a distância *DTW* entre a série de sentimentos e a dos preços normalizados.
- c) **Cálculo do desfasamento ótimo dos comentários:** o valor ótimo τ_i^{tw} é aquele que minimiza a distância *DTW* média ao longo de todas as janelas deslizantes:

$$\tau_i^{tw} = \min_{\tau \in \{0,1,2\}} \bar{d}_{DTW}(S_{i,t-\tau}^{tw}, P_{i,t}^{norm}) \quad (18)$$

- d) **Cálculo do desfasamento ótimo das notícias:** da mesma forma, o desfasamento ótimo τ_i^{nw} corresponde ao valor de τ que minimiza a distância *DTW* média sobre as janelas deslizantes:

$$\tau_i^{nw} = \min_{\tau \in \{0,1,2\}} \bar{d}_{DTW}(S_{i,t-\tau}^{nw}, P_{i,t}^{norm}) \quad (19)$$

3.6.4 Alinhamento direcional em janelas deslizantes

Para cada ação i , e tendo em conta o desfasamento τ_i previamente identificado como ótimo, construiu-se uma versão deslocada da série de sentimento $S_{i,t-\tau_i}$, que foi comparada com a série de preços normalizados $P_{i,t}^{norm}$. Este procedimento foi aplicado separadamente às séries de sentimento provenientes dos comentários ($S_{i,t}^{tw}$) e das notícias ($S_{i,t}^{nw}$).

Define-se uma janela deslizante de tamanho fixo de 4 dias. Isto significa que numa semana de 5 dias uteis obtêm-se três janelas completas, cada uma contendo três variações consecutivas (domingo a quarta, segunda a quinta e terça a sexta), permitindo assim captar padrões de curto prazo.

Para cada uma destas janelas, calcula-se a direção das variações diárias em ambas as séries, considerando apenas subidas e descidas e ignorando variações nulas. Uma janela é considerada alinhada quando:

$$\text{sign}(S_{t+1} - S_t) = \text{sign}(P_{t+1} - P_t), \quad \forall t \in \text{janela} \quad (20)$$

e quando não existem variações nulas na janela. Em termos práticos, o alinhamento verifica-se quando a sequência de subidas e descidas ao longo da janela coincide exatamente em ambas as séries.

A taxa de alinhamento observada é definida como o rácio entre o número de janelas perfeitamente alinhadas (A) e o número total de janelas válidas (T):

$$\text{Taxa de Alinhamento} = \frac{A}{T} \quad (21)$$

Este cálculo do alinhamento direcional baseia-se no conceito de *Directional Symmetry*⁸, que é a medida estatística que avalia se o modelo prevê corretamente a direção da variação de uma série temporal de um período para o seguinte, seja de subida ou de descida (Drossu & Obradovic, 1997). Os autores L. Cao & Tay (2001) mediram a percentagem de vezes em que a variação prevista e a variação real têm o mesmo sinal. No estudo original, esta métrica foi aplicada ponto a ponto. No presente trabalho foi calculado em janelas deslizantes de quatro dias, considerando alinhada apenas a sequência em que todas as variações coincidem.

3.6.5 Teste de significância estatística com permutações aleatórias

Para avaliar se o número de janelas alinhadas poderia ser resultado do acaso, aplicou-se um teste de hipóteses não paramétrico por permutação (Good, 2005; Holt & Sullivan, 2023).

⁸ [https://en.wikipedia.org/wiki/Directional_symmetry_\(time_series\)](https://en.wikipedia.org/wiki/Directional_symmetry_(time_series))

- **Hipótese nula (H_0):** o alinhamento entre a direção do sentimento e a direção do preço é resultado do acaso, ou seja, não existem padrões temporais entre as duas séries;
- **Hipótese alternativa (H_1):** o alinhamento entre sentimento (comentários e notícias) e o preço das ações é superior ao esperado por mero acaso, sugerindo uma relação não aleatória entre as duas séries.

O teste é implementado da seguinte forma:

- Mantendo fixos os preços, reordena-se aleatoriamente a série de sentimento, tendo em conta o desfasamento τ_i previamente identificado.
- Calcula-se o número de janelas com variação direcional perfeitamente alinhada, tal como no cenário original.
- Repete-se este processo 1000 vezes, gerando uma distribuição do número de alinhamentos esperados ao acaso. Este valor encontra-se em consonância com a recomendação de Marozzi (2004), que indica que cerca de 1000 permutações são substancialmente suficientes para testes de permutação realizados ao nível de significância de 5%.

Com base na distribuição obtida, calcula-se o *valor-p* como a proporção de permutações em que o número de janelas alinhadas foi maior ou igual ao observado:

$$p = \frac{1}{N} \sum_{j=1}^N 1(A_j^{\text{perm}} \geq A^{\text{obs}}) \quad (22)$$

Onde A^{obs} representa o número de alinhamentos observados, e A_j^{perm} representa o número de alinhamentos na permutação j .

Um valor de $p\text{-value} < 0,05$ indica que o alinhamento observado é estatisticamente significativo, sugerindo que a direção do sentimento está relacionada com a direção do preço de fecho de forma não aleatória. Este limiar é amplamente utilizado na literatura para rejeitar a hipótese nula e considerar o resultado estatisticamente significativo (Di Leo & Sardanelli, 2020).

CAPÍTULO IV – ANÁLISE E DISCUSSÃO DE RESULTADOS

Para o ambiente de desenvolvimento, foi utilizado um computador com *Windows 11 Home Edition*, com um processador *Intel Core i7-13620H* (13.^a geração), 16 GB de memória *RAM* e uma placa gráfica *NVIDIA GeForce RTX 4050*, com 6 GB de memória *VRAM*, compatível com *CUDA 11.8*.

Com o objetivo de aumentar a fiabilidade estatística dos resultados, o fluxo de processamento foi executado três vezes, utilizando diferentes sementes de aleatoriedade. No total, foram realizadas 4.224 execuções. O tempo acumulado de execução foi de 181.248,84 segundos, aproximadamente 50 horas de execução. Os detalhes relativos ao número de execuções e ao tempo total de execução obtidos para cada abordagem está resumido na Tabela 33, Tabela 34, Tabela 35 e Tabela 36.

Tabela 33 - Número de Execuções e Tempo Total de Execução para Analisadores Lexicais. Fonte: Elaboração própria.

Tipo de Analisador	Num Exec	Tempo Exec (s)
<i>Afinn</i>	96	98.84
<i>Finvader</i>	96	7550.96
<i>Sentiwordnet</i>	96	1084.12
<i>Textblob</i>	96	104.79
<i>Vader</i>	96	59.51
Total	480	8898.22

Tabela 34 - Número de Execuções e Tempo Total de Execução para Modelos de Aprendizagem Computacional Tradicional. Fonte: Elaboração própria.

Tipo de Analisador	Num Exec	Tempo Exec (s)
<i>DecisionTree</i>	384	16910.30
<i>LightGBM</i>	384	2164.73
<i>LogisticRegression</i>	384	123.01
<i>NaiveBayes</i>	384	31.07
<i>SVM</i>	384	1243.11
<i>XGBoost</i>	384	16714.49
Total	2304	37186.71

Tabela 35 - Número de Execuções e Tempo Total de Execução para Modelos Baseados em Redes Neurais. Fonte: Elaboração própria.

Tipo de Modelo	Num Exec	Tempo Exec (s)
<i>BiLSTM</i>	192	8037.39
<i>CNN</i>	192	3045.61
<i>CNN-BiLSTM</i>	192	7998.52
<i>GRU</i>	192	3149.60
<i>MLP</i>	192	1521.77
Total	960	8898.22

Tabela 36 - Número de Execuções e Tempo Total de Execução para Modelos Baseados em Transformadores. Fonte: Elaboração própria.

Tipo de Modelo	Num Exec	Tempo Exec (s)
<i>BERT</i>	96	25871.27
<i>BERTweet</i>	96	28004.91
<i>DistilBERT</i>	96	13115.99
<i>FinBERT</i>	96	25714.80
<i>RoBERTa-StockTwits</i>	96	18704.05
Total	480	111411.02

4.1 Comparação de desempenho dos classificadores

Esta secção apresenta a análise comparativa dos diversos modelos agrupados em quatro grandes categorias: analisadores baseados em léxicos, modelos de aprendizagem computacional tradicional, modelos baseados em redes neuronais e modelos baseados em transformadores.

4.1.1 Analisadores baseados em léxicos

No caso dos analisadores lexicais, não foi realizado treino sobre os dados. Estas técnicas aplicam as regras diretamente ao texto, sem qualquer ajuste adicional. Como os resultados obtidos são idênticos em ambos os cenários, não se aplica a análise de sensibilidade a técnicas de balanceamento.

A Figura 20 mostra, para cada analisador, o valor mais elevado registado. A Figura 21 apresenta um mapa de calor com os valores de exatidão obtidos por cada analisador lexical em função da técnica de pré-processamento aplicada. A Figura 22 evidencia a distribuição da exatidão média dos analisadores ao longo de diferentes execuções.

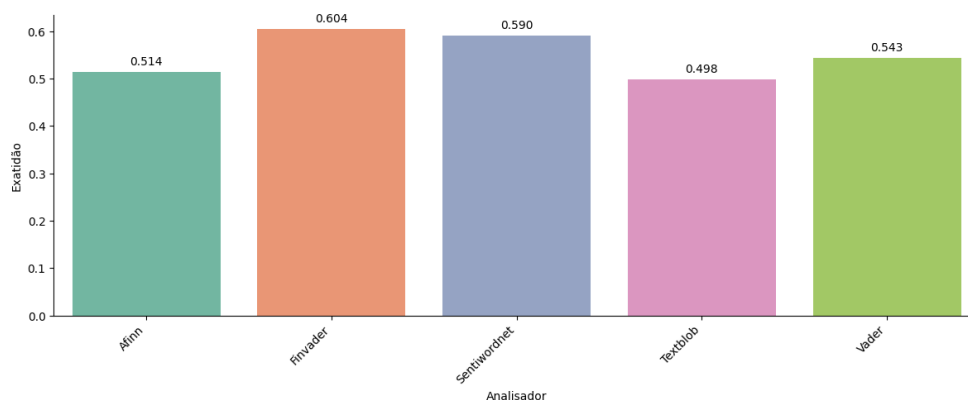


Figura 20 - Exatidão máxima por analisador baseado em léxicos. Fonte: Elaboração própria.

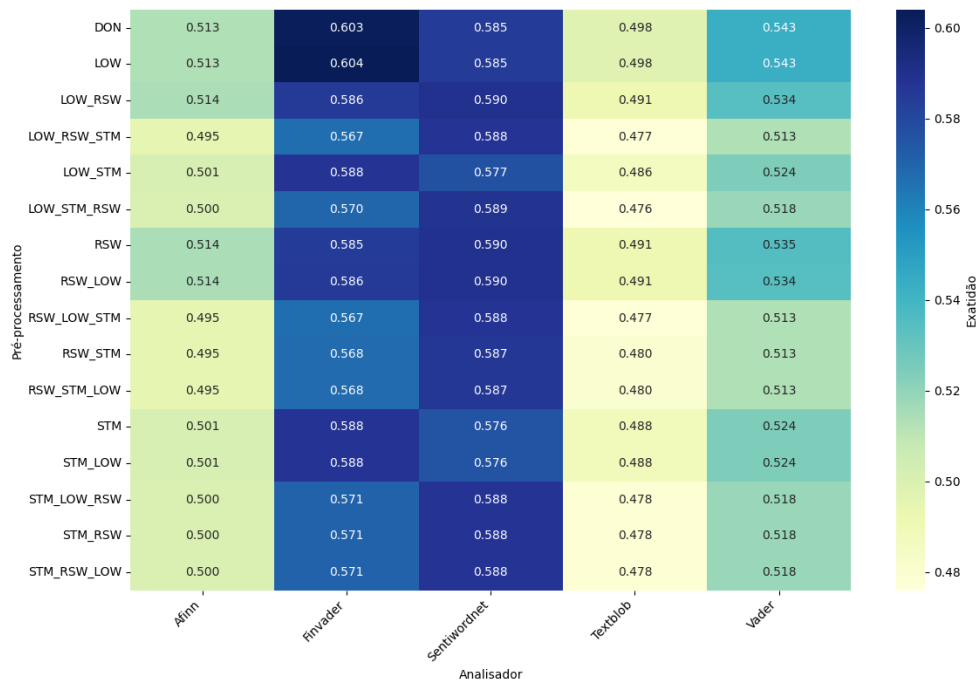


Figura 21 - Exatidão máxima por analisador baseado em léxicos e técnica de pré-processamento. Fonte: Elaboração própria.

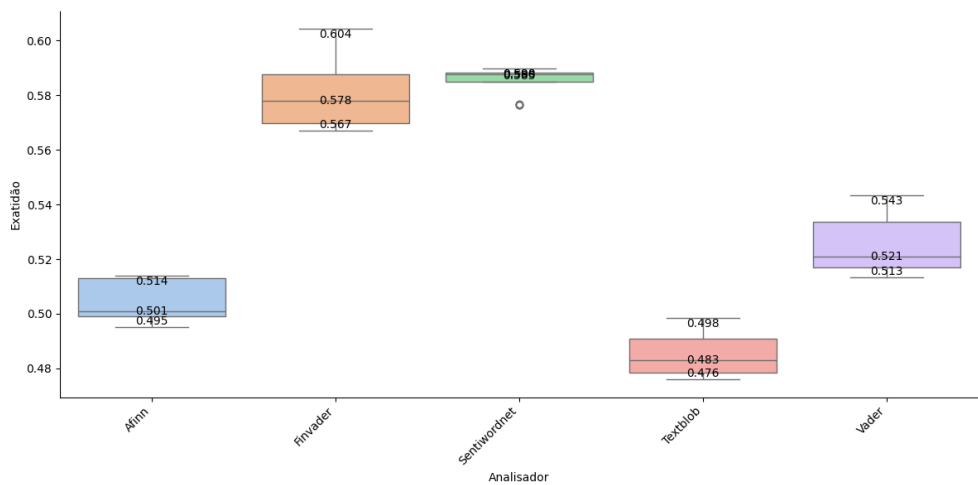


Figura 22 - Distribuição da exatidão dos analisadores baseados em léxicos, nas diferentes execuções. Fonte: Elaboração própria.

Os resultados obtidos demonstram que o *FinVader* foi a abordagem com melhor desempenho global, alcançando uma exatidão máxima de **0.604**, sendo seguido por *SentiWordNet* (**0.590**) e *Vader* (**0.543**). Por outro lado, o *TextBlob* registou o desempenho mais baixo, com uma exatidão máxima de **0.498**.

4.1.1.1 Análise de sensibilidade à técnica de pré-processamento

A Tabela 37 apresenta, para cada analisador lexical, a técnica de pré-processamento com a qual foi obtida a exatidão mais elevada, a técnica com o valor mais baixo e a variação absoluta entre ambos. A Figura 23 mostra essas variações sob a forma de gráfico de barras.

Tabela 37 - Variação da exatidão dos analisadores baseados em léxicos entre a melhor e a pior técnica de pré-processamento. Fonte: Elaboração própria.

Analisador	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>Finvader</i>	LOW	0.604	LOW_RSW_STM, RSW_LOW_STM	0.567	0.037
<i>Vader</i>	DON	0.543	LOW_RSW_STM, RSW_LOW_STM, RSW_STM, RSW_STM_LOW	0.513	0.030
<i>TextBlob</i>	DON, LOW	0.498	LOW_STM_RSW	0.476	0.022
<i>Afinn</i>	LOW_RSW, RSW, RSW_LOW	0.514	LOW_RSW_STM, RSW_LOW_STM, RSW_STM, RSW_STM_LOW	0.495	0.019
<i>SentiWordNet</i>	LOW_RSW, RSW, RSW_LOW	0.590	STM, STM_LOW	0.576	0.013

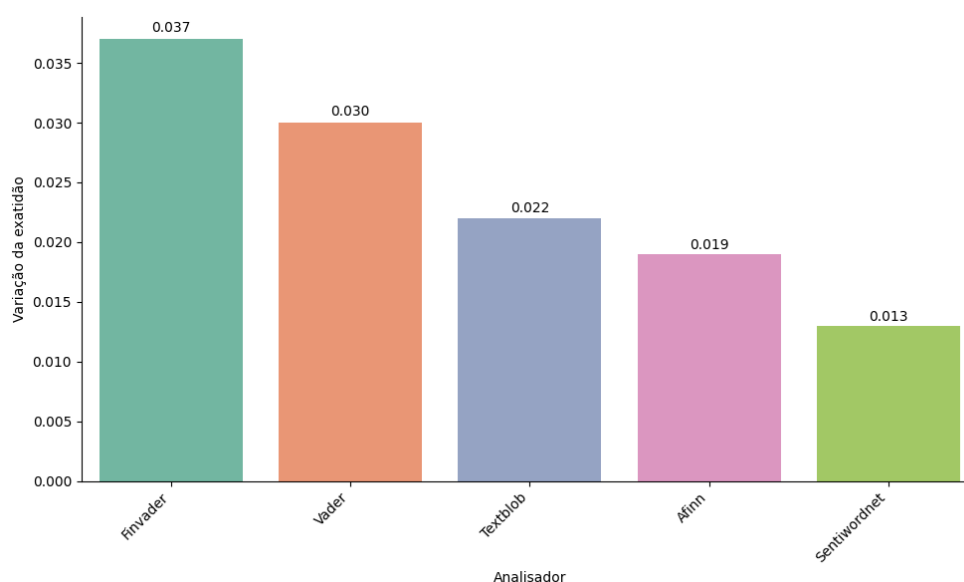


Figura 23 - Variação da exatidão dos analisadores baseados em léxicos por técnica de pré-processamento. Fonte: Elaboração própria.

Os resultados obtidos demonstram que os piores desempenhos ocorreram com combinações de várias técnicas de pré-processamento, sobretudo nos modelos *FinVader* e *Vader*, que registaram variações de exatidão de **0.037** e **0.030**, respectivamente.

4.1.1.2 Melhores combinações

A Tabela 38 apresenta o resumo das melhores combinações para cada analisador baseado em léxicos, considerando a técnica de pré-processamento, a técnica de balanceamento e o respectivo valor de exatidão média obtido.

Tabela 38 - Comparação das melhores combinações de pré-processamento e desempenho para cada analisador baseado em léxicos. Fonte: Elaboração própria.

Analisador	Melhor estratégia de pré-processamento	Melhor exatidão média
<i>FinVADER</i>	LOW	0.604
<i>SentiWordNet</i>	LOW_RSW, RSW, RSW_LOW	0.590
<i>VADER</i>	DON	0.543
<i>AFINN</i>	LOW_RSW, RSW, RSW_LOW	0.514
<i>TextBlob</i>	DON, LOW	0.498

4.1.2 Modelos de aprendizagem computacional tradicional

A Figura 24 ilustra a melhor exatidão alcançada por cada modelo, considerando a técnica de balanceamento aplicada. A Figura 25 representa, sob a forma de mapa de calor, a distribuição dos valores de exatidão por modelo e por técnica de pré-processamento. A Figura 26 evidencia a distribuição da exatidão média dos modelos ao longo de diferentes execuções.

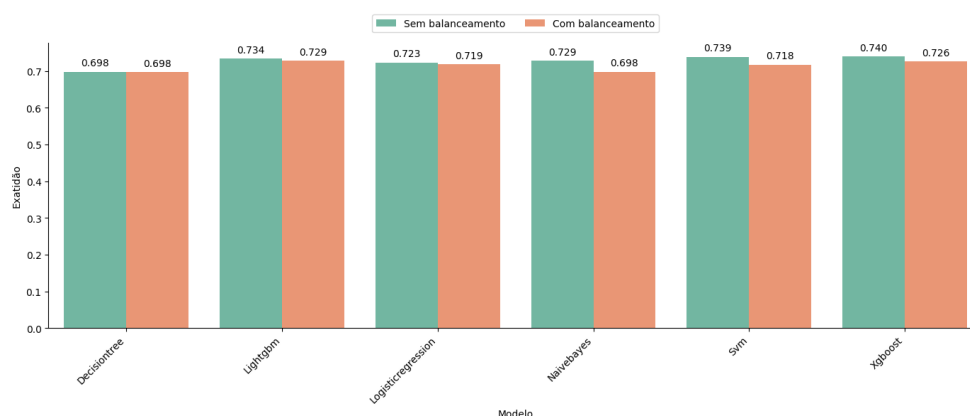


Figura 24 - Exatidão máxima por modelo de aprendizagem computacional tradicional, com e sem aplicação de técnicas de balanceamento. Fonte: Elaboração própria.

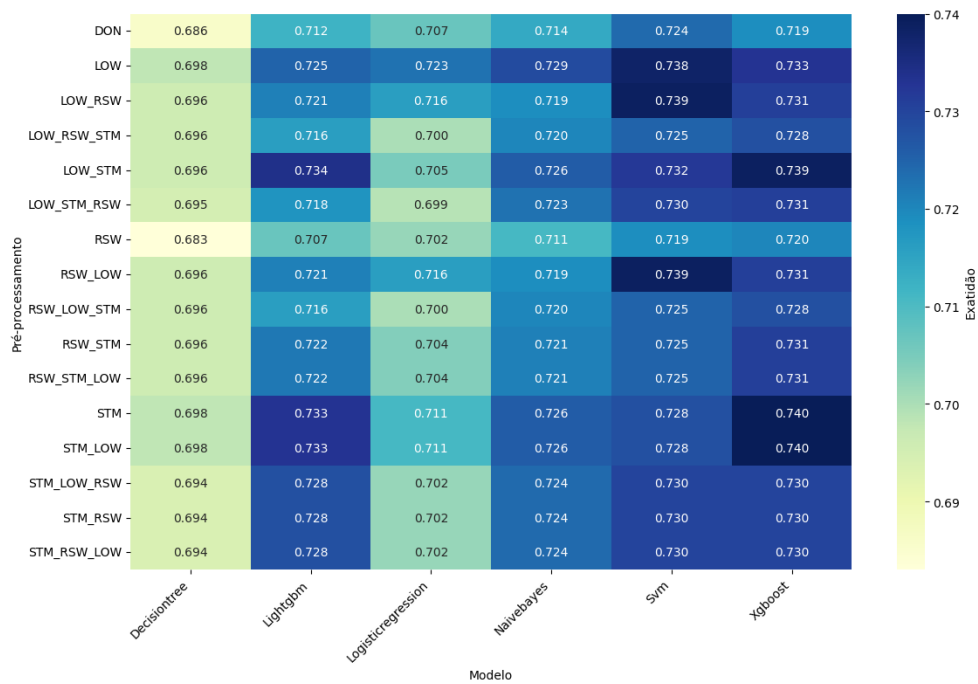


Figura 25 - Exatidão máxima por modelo de aprendizagem computacional tradicional e técnica de pré-processamento, com e sem balanceamento. Fonte: Elaboração própria.

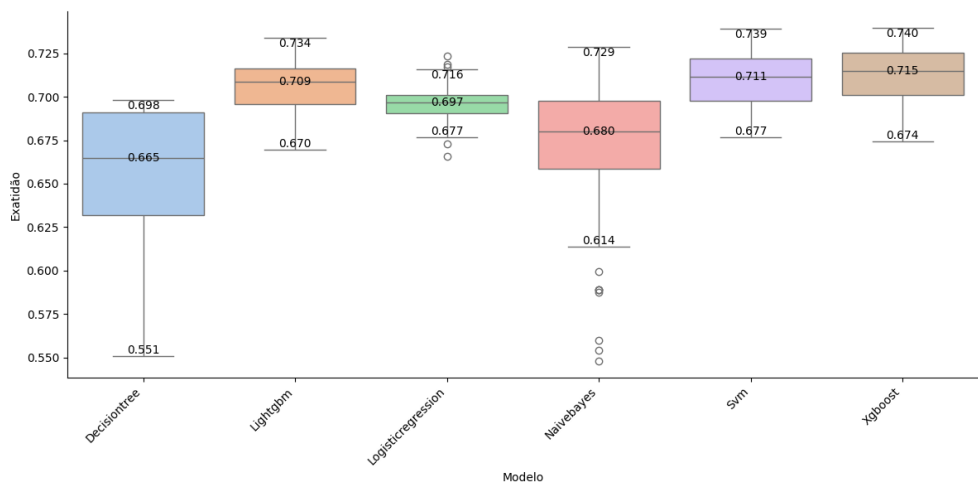


Figura 26 - Distribuição da exatidão dos modelos de aprendizagem computacional tradicional, nas diferentes execuções, com e sem balanceamento. Fonte: Elaboração própria.

Os resultados obtidos demonstram que o *XGBoost* obteve a melhor exatidão máxima (**0.740**), seguido do *SVM* (**0.739**), ambos com desempenho estável e baixa dispersão. A *Decision Tree* alcançou **0.698**, mas com elevada variabilidade e menor consistência.

4.1.2.1 Análise de sensibilidade à técnica de balanceamento

A Tabela 39 apresenta, para cada modelo de aprendizagem computacional tradicional, os valores de exatidão máxima obtidos nos cenários com e sem aplicação da técnica de

balanceamento, bem como a variação absoluta entre ambos. A Figura 27 representa essas variações através de um gráfico de barras.

Tabela 39 - Comparação da exatidão máxima dos modelos de aprendizagem computacional tradicional, com e sem balanceamento dos dados, incluindo a variação observada. Fonte: Elaboração própria.

Modelo	Sem balanceamento	Com balanceamento	Variação
<i>NaiveBayes</i>	0.729	0.698	0.031
<i>SVM</i>	0.739	0.718	0.021
<i>XGBoost</i>	0.740	0.726	0.014
<i>LightGBM</i>	0.734	0.729	0.005
<i>LogisticRegression</i>	0.723	0.719	0.004
<i>DecisionTree</i>	0.698	0.698	0.000

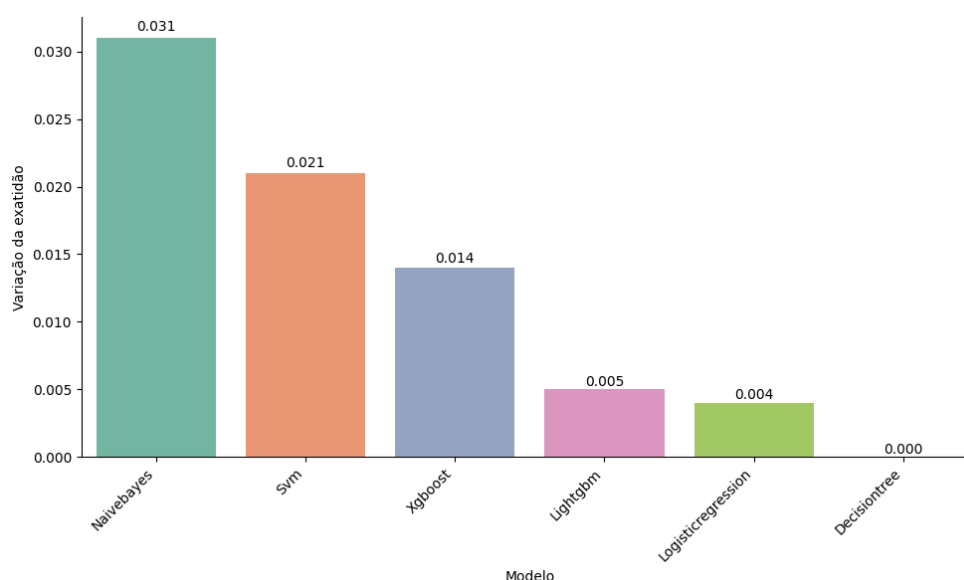


Figura 27 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, com e sem balanceamento dos dados. Fonte: Elaboração própria.

Os resultados obtidos demonstram que o *Naive Bayes* foi o mais sensível ao balanceamento, com variação de exatidão de **0.031**. A *Decision Tree* manteve variação nula (**0.000**), revelando grande robustez.

4.1.2.2 Análise de sensibilidade à técnica de pré-processamento

A Tabela 40 e Tabela 41 apresentam, para cada modelo de aprendizagem computacional tradicional, a combinação de técnicas de pré-processamento com a qual foi obtida a maior e menor exatidão máxima, os respectivos valores e a variação absoluta entre ambos, nos

conjuntos de dados sem e com balanceamento, respetivamente. A Figura 28 e Figura 29 representam graficamente essas variações.

Tabela 40 - Comparação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (sem balanceamento dos dados). Fonte: Elaboração própria.

Modelo	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>LightGBM</i>	LOW_STM	0.734	RSW	0.707	0.027
<i>LogisticRegression</i>	LOW	0.723	LOW_RSW_STM, LOW_STM_RSW, RSW_LOW_STM	0.699	0.024
<i>XGBoost</i>	STM, STM_LOW	0.740	DON	0.719	0.021
<i>SVM</i>	LOW_RSW, RSW_LOW	0.739	RSW	0.719	0.020
<i>NaiveBayes</i>	LOW	0.729	RSW	0.711	0.018
<i>DecisionTree</i>	LOW	0.698	RSW	0.683	0.015

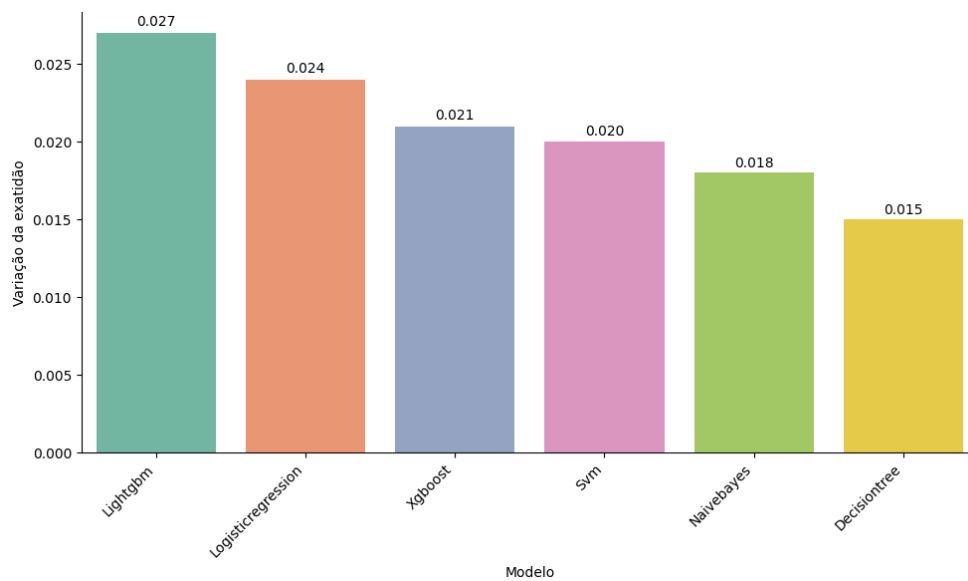


Figura 28 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre combinações de técnicas de pré-processamento, sem balanceamento dos dados. Fonte: Elaboração própria.

Tabela 41 - Comparação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (com balanceamento dos dados). Fonte: Elaboração própria.

Modelo	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>NaiveBayes</i>	LOW_STM, STM, STM_LOW	0.698	RSW	0.663	0.035

<i>LightGBM</i>	LOW_STM	0.729	RSW	0.705	0.024
<i>XGBoost</i>	LOW_STM	0.726	RSW	0.704	0.022
<i>DecisionTree</i>	STM, STM_LOW	0.698	RSW	0.678	0.020
<i>LogisticRegression</i>	LOW	0.719	LOW_STM_RSW, RSW	0.699	0.020
<i>SVM</i>	LOW	0.718	RSW	0.701	0.017

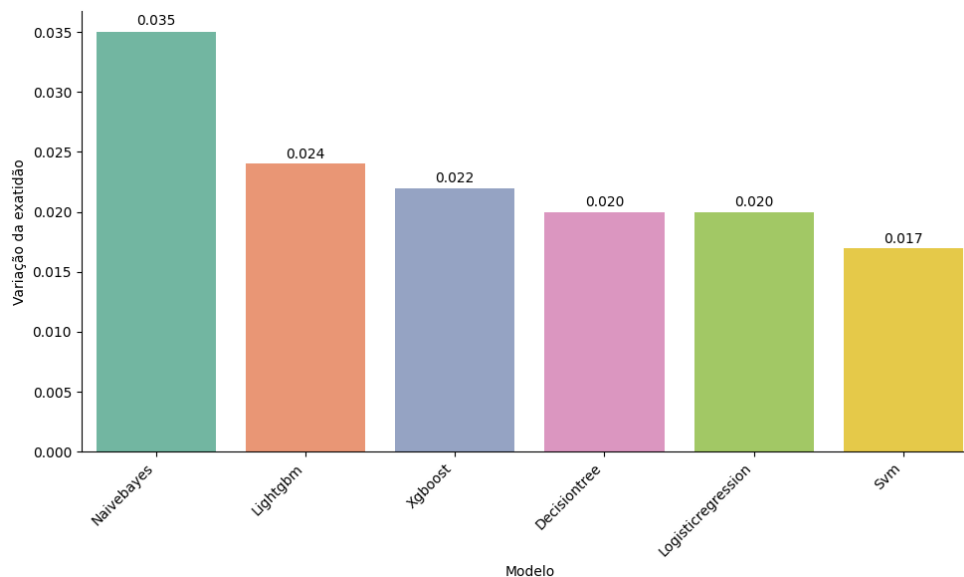


Figura 29 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre combinações de técnicas de pré-processamento, com balanceamento dos dados. Fonte: Elaboração própria.

Os resultados obtidos demonstram que, sem balanceamento, o *LightGBM* apresentou maior variação (**0.027**), seguido da *Logistic Regression* (**0.024**). Com balanceamento, o destaque foi para o *Naive Bayes*, com variação de **0.035**. De forma geral, técnicas simples como conversão para minúsculas e combinações com radicalização obtiveram melhores resultados, enquanto remoção de palavras com pouco significado semântico esteve frequentemente associada a pior desempenho.

4.1.2.3 Análise de sensibilidade à técnica de representação textual

A Tabela 42 e a Figura 30 apresentam a variação da exatidão dos modelos de aprendizagem computacional tradicional consoante a representação textual utilizada, no cenário sem balanceamento. De forma equivalente, a Tabela 43 e a Figura 32 mostram os mesmos resultados com aplicação de balanceamento dos dados. A Figura 31 e a Figura 33 complementam as análises anteriores ao ilustrar a exatidão máxima alcançada por cada modelo de aprendizagem computacional tradicional para cada uma das representações textuais testadas.

Tabela 42 - Exatidão dos modelos de aprendizagem computacional tradicional para cada representação textual, incluindo a melhor e a pior exatidão máxima registadas e a respetiva variação (sem balanceamento dos dados).

Fonte: Elaboração própria.

Modelo	BOW	FastText	TFIDF	Word2Vec	Melhor	Pior	Variação
<i>NaiveBayes</i>	0.729	0.687	0.718	0.678	0.729	0.678	0.051
<i>DecisionTree</i>	0.696	0.651	0.698	0.655	0.698	0.651	0.047
<i>LogisticRegression</i>	0.706	0.723	0.711	0.717	0.723	0.706	0.017
<i>SVM</i>	0.724	0.738	0.732	0.739	0.739	0.724	0.015
<i>LightGBM</i>	0.734	0.720	0.723	0.722	0.734	0.720	0.014
<i>XGBoost</i>	0.740	0.732	0.733	0.731	0.740	0.731	0.009

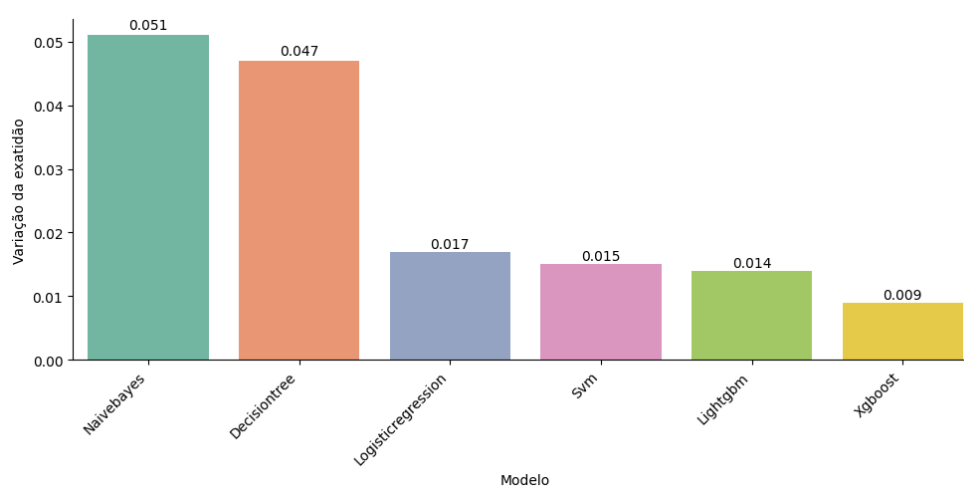


Figura 30 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre diferentes representações textuais, sem balanceamento dos dados. Fonte: Elaboração própria.

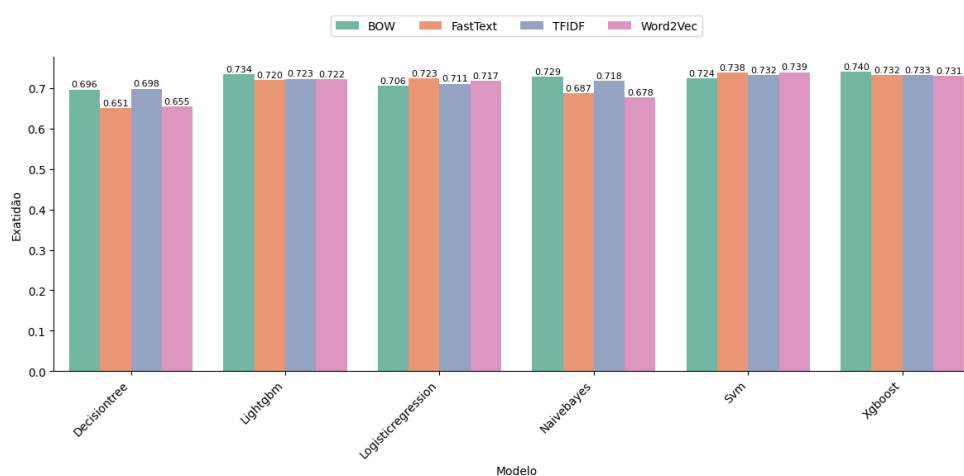


Figura 31 - Exatidão máxima dos modelos de aprendizagem computacional tradicional para cada representação textual (sem balanceamento dos dados). Fonte: Elaboração própria.

Tabela 43 - Exatidão dos modelos de aprendizagem computacional tradicional para cada representação textual, incluindo a melhor e a pior exatidão registradas e a respectiva variação (com balanceamento de dados). Fonte: Elaboração própria.

Modelo	BOW	FastText	TFIDF	Word2Vec	Melhor	Pior	Variação
<i>DecisionTree</i>	0.691	0.608	0.698	0.600	0.698	0.600	0.098
<i>NaiveBayes</i>	0.691	0.666	0.698	0.645	0.698	0.645	0.053
<i>LightGBM</i>	0.729	0.697	0.722	0.702	0.729	0.697	0.032
<i>XGBoost</i>	0.726	0.698	0.723	0.702	0.726	0.698	0.028
<i>LogisticRegression</i>	0.697	0.719	0.696	0.715	0.719	0.696	0.023
<i>SVM</i>	0.701	0.718	0.706	0.713	0.718	0.701	0.017

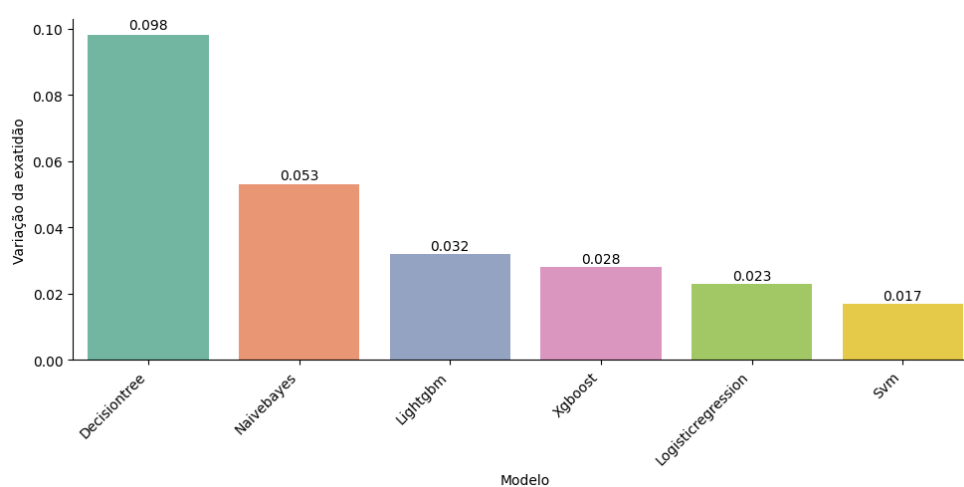


Figura 32 - Variação da exatidão máxima dos modelos de aprendizagem computacional tradicional, entre diferentes representações textuais, com balanceamento dos dados. Fonte: Elaboração própria.

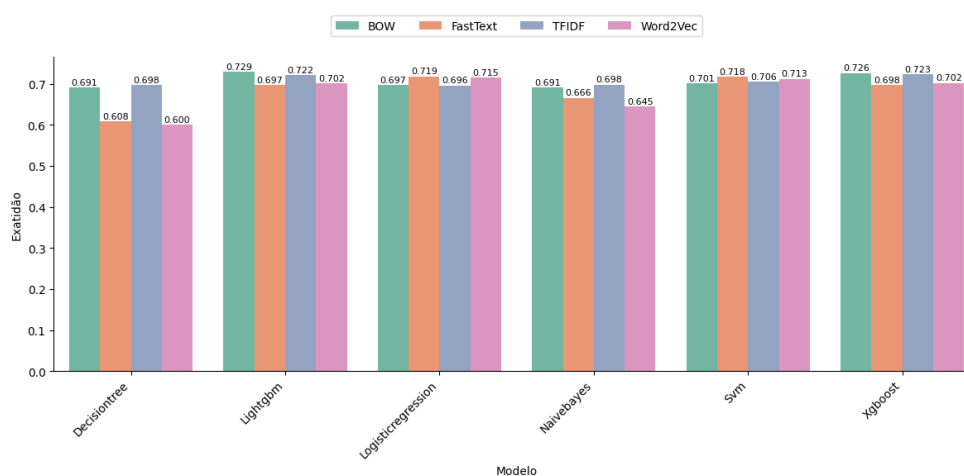


Figura 33 - Exatidão máxima dos modelos de aprendizagem computacional tradicional para cada representação textual (com balanceamento dos dados). Fonte: Elaboração própria.

Os resultados obtidos demonstram que *Decision Tree* e *Naive Bayes* foram os mais sensíveis, enquanto *XGBoost* e *SVM* apresentaram maior robustez. De forma geral, representações esparsas, baseadas em frequências de palavras como *BOW* e *TF-IDF* obtiveram melhores resultados do que representações densas neste tipo de classificadores.

4.1.2.4 Melhores combinações

A Tabela 44 resume as melhores combinações encontradas para cada modelo de aprendizagem computacional, considerando em simultâneo a estratégia de pré-processamento, a técnica de balanceamento, a representação textual e o valor de exatidão média obtido.

Tabela 44 - Comparação das melhores combinações de pré-processamento, balanceamento, representação textual e desempenho para cada modelo de aprendizagem computacional. Fonte: Elaboração própria.

Modelo	Melhor estratégia pré-processamento	Melhor estratégia balanceamento	Melhor estratégia repres. textual	Melhor exatidão média
<i>XGBoost</i>	STM, STM_LOW	NOB	BOW	0.740
<i>SVM</i>	LOW_RSW, RSW_LOW	NOB	Word2Vec	0.739
<i>LightGBM</i>	LOW_STM	NOB	BOW	0.734
<i>NaiveBayes</i>	LOW	NOB	BOW	0.729
<i>LogisticRegression</i>	LOW	NOB	FastText	0.723
<i>DecisionTree</i>	LOW, STM, STM_LOW	NOB, USM	TFIDF	0.698

4.1.3 Modelos baseados em redes neuronais

A Figura 34 mostra a melhor exatidão alcançada por cada modelo, em função da técnica de balanceamento utilizada. A Figura 35 representa os mesmos dados sob a forma de mapa de calor, permitindo visualizar a distribuição dos resultados por modelo e por técnica de pré-processamento. A Figura 36 evidencia a distribuição da exatidão média dos modelos ao longo de diferentes execuções.

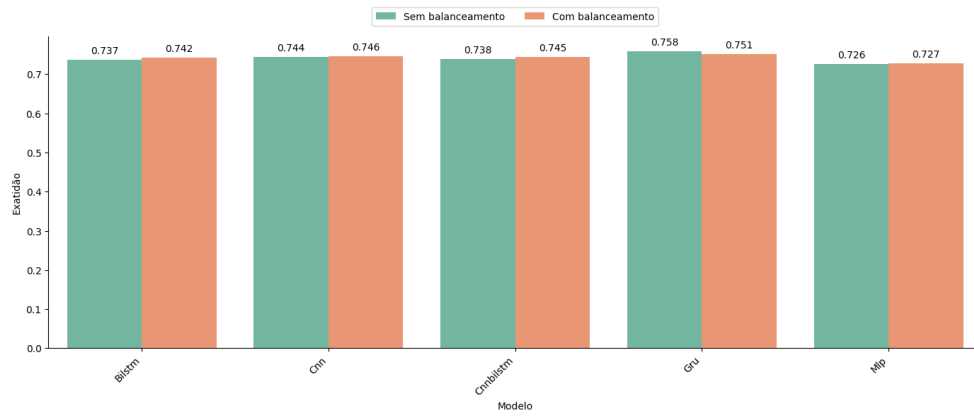


Figura 34 - Exatidão máxima por modelo baseado em redes neuronais, com e sem aplicação de técnicas de balanceamento. Fonte: Elaboração própria.

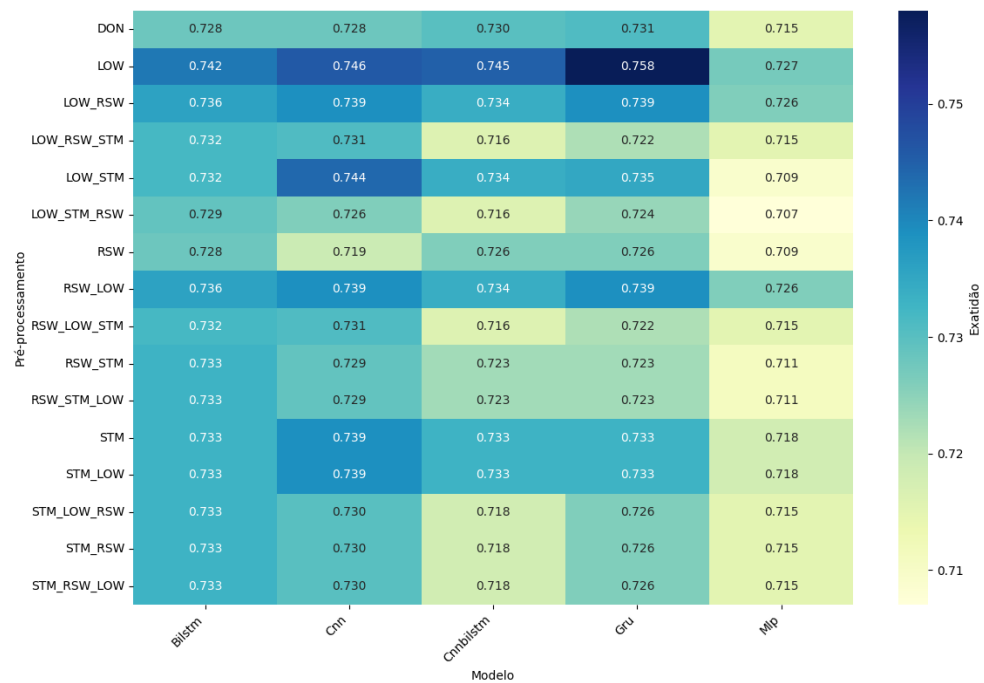


Figura 35 - Exatidão máxima por modelo baseado em redes neuronais e técnica de pré-processamento, com e sem balanceamento. Fonte: Elaboração própria.

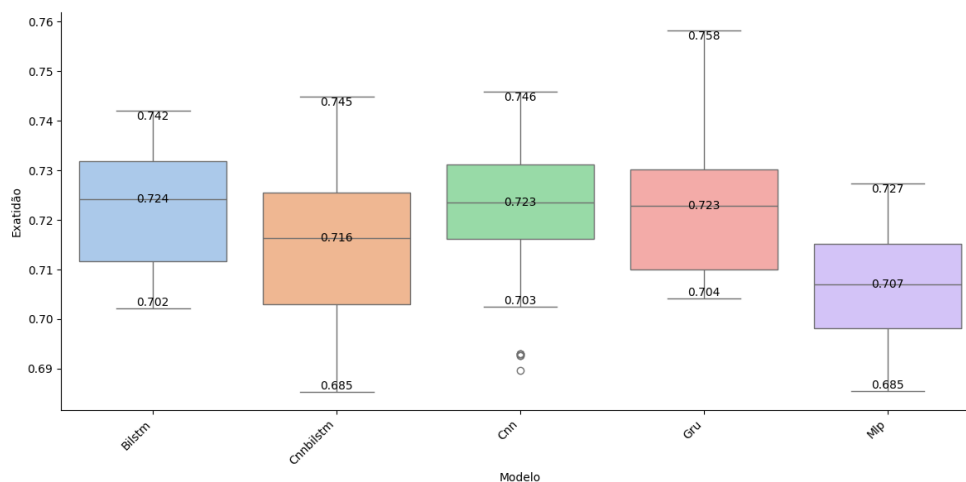


Figura 36 - Distribuição da exatidão dos modelos baseados em redes neuronais, nas diferentes execuções, com e sem balanceamento. Fonte: Elaboração própria.

Os resultados obtidos demonstram que o *GRU* obteve a melhor exatidão máxima (**0.758**) no cenário sem balanceamento, seguido por *CNN* (**0.746**), *CNN-BiLSTM* (**0.745**) e *BiLSTM* (**0.742**) no cenário com balanceamento. O *MLP* apresentou o pior resultado (**0.727**).

4.1.3.1 Análise de sensibilidade à técnica de balanceamento

A Tabela 45 apresenta, para cada modelo baseado em redes neuronais, os valores de exatidão máxima obtidos nos cenários com e sem aplicação de técnicas de balanceamento, bem como a variação absoluta entre ambos. A Figura 37 representa essas variações sob a forma de gráfico de barras.

Tabela 45 - Comparação da exatidão máxima dos modelos baseados em redes neuronais, com e sem balanceamento dos dados, incluindo a variação observada. Fonte: Elaboração própria.

Modelo	Sem balanceamento	Com balanceamento	Varição
<i>CNN-BiLSTM</i>	0.738	0.745	0.007
<i>GRU</i>	0.758	0.751	0.007
<i>BiLSTM</i>	0.737	0.742	0.005
<i>CNN</i>	0.744	0.746	0.002
<i>MLP</i>	0.726	0.727	0.001

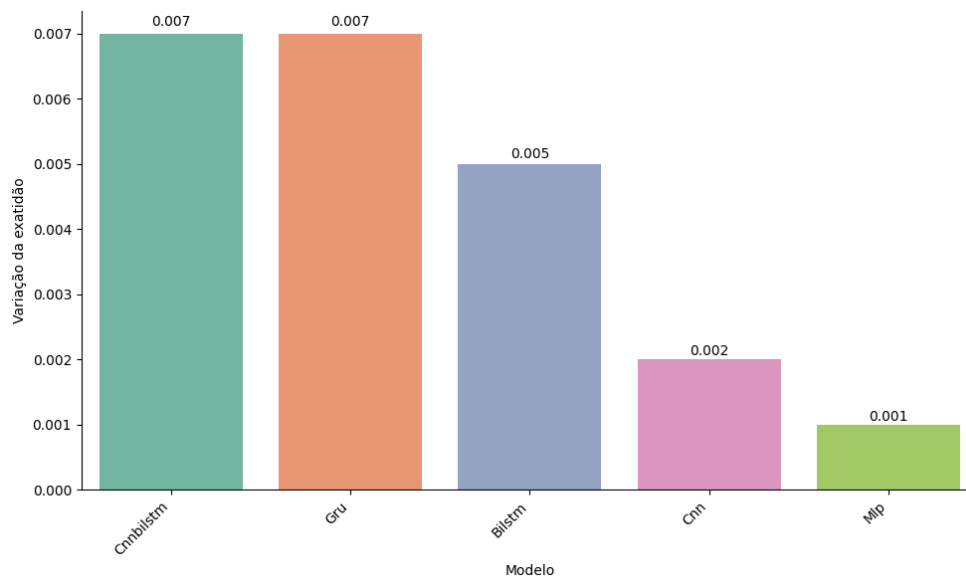


Figura 37 - Variação da exatidão máxima dos modelos baseados em redes neuronais, com e sem balanceamento dos dados. Fonte: Elaboração própria.

Os resultados obtidos demonstram que a variação de exatidão entre cenários com e sem balanceamento foi reduzida. Ao contrário do observado noutras abordagens, alguns modelos, nomeadamente *BiLSTM*, *CNN*, *CNN-BiLSTM* e *MLP*, apresentaram um desempenho ligeiramente superior no cenário com balanceamento. Apenas o *GRU* registou uma pequena redução de desempenho.

4.1.3.2 Análise de sensibilidade à técnica de pré-processamento

A Tabela 46 e Tabela 47 apresentam, para cada modelo baseado em redes neuronais, a técnica de pré-processamento com a qual foi obtida a exatidão máxima, a técnica associada ao valor mínimo, os respetivos valores e a variação absoluta entre ambos, nos cenários sem e com aplicação de balanceamento, respetivamente. A Figura 38 e Figura 39 representam graficamente essas variações.

Tabela 46- Comparação da exatidão máxima dos modelos baseados em redes neuronais, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (sem balanceamento dos dados). Fonte: Elaboração própria.

Modelo	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>GRU</i>	LOW	0.758	RSW	0.712	0.046
<i>CNN</i>	LOW_STM	0.744	DON	0.715	0.029
<i>CNN-BiLSTM</i>	LOW	0.738	LOW_RSW_STM, LOW_STM_RSW, RSW_LOW_STM,	0.716	0.022

			STM_LOW_RSW, STM_RSW, STM_RSW_LOW		
<i>MLP</i>	LOW_RSW, RSW_LOW	0.726	RSW	0.706	0.020
<i>BiLSTM</i>	LOW	0.737	LOW_STM	0.722	0.015

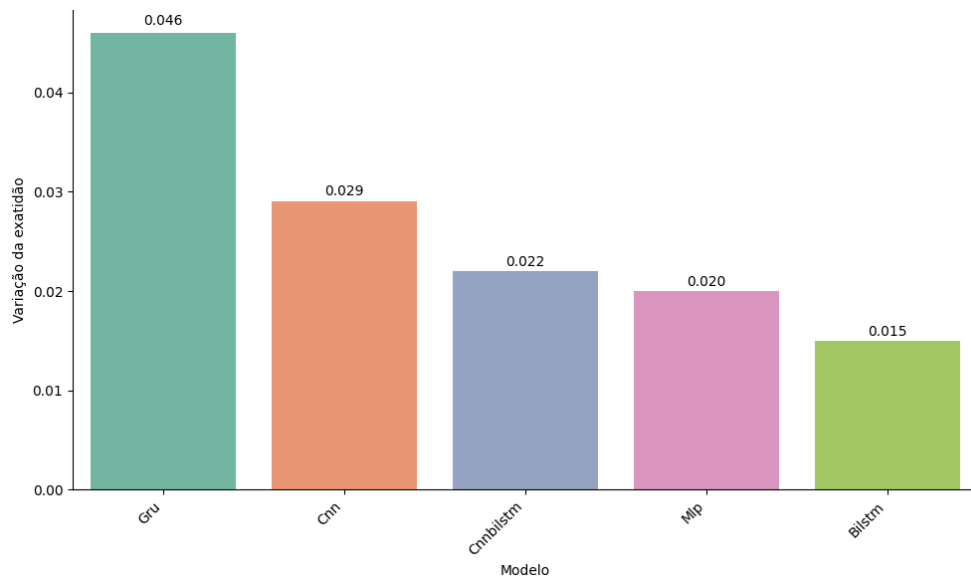


Figura 38 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre combinações de técnicas de pré-processamento, sem balanceamento dos dados. Fonte: Elaboração própria.

Tabela 47 - Comparação da exatidão máxima dos modelos baseados em redes neuronais, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (com balanceamento dos dados). Fonte: Elaboração própria.

Modelo	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>CNN</i>	LOW	0.746	RSW	0.710	0.036
<i>GRU</i>	LOW	0.751	LOW_RSW_STM, RSW_LOW_STM	0.715	0.036
<i>CNN-BiLSTM</i>	LOW	0.745	LOW_STM_RSW, RSW_STM, RSW_STM_LOW	0.712	0.033
<i>BiLSTM</i>	LOW	0.742	RSW	0.711	0.031
<i>MLP</i>	LOW	0.727	LOW_STM_RSW	0.705	0.022

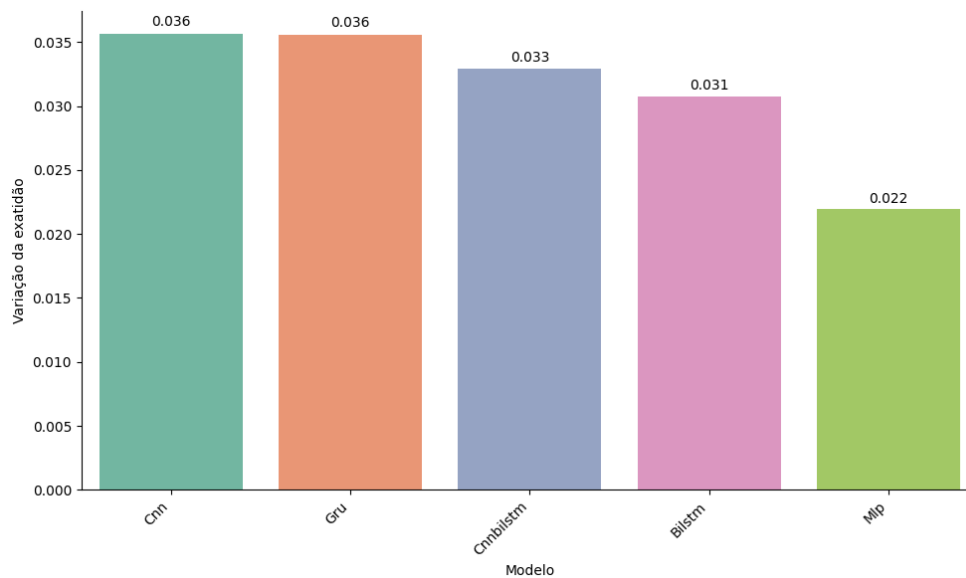


Figura 39 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre combinações de técnicas de pré-processamento, com balanceamento dos dados. Fonte: Elaboração própria.

Os resultados obtidos demonstram que a variação da exatidão por técnica de pré-processamento foi mais marcada do que no caso do balanceamento. Em ambos os cenários, a conversão para minúsculas esteve associada aos melhores resultados, enquanto a remoção de palavras com pouco significado semântico e combinações mais complexas surgiram frequentemente entre os piores. Os modelos *CNN* e *GRU* foram os que registaram maior variação de desempenho, revelando maior sensibilidade à escolha da técnica aplicada.

4.1.3.3 Análise de sensibilidade à técnica de representação textual

A Tabela 48 e a Figura 40 apresentam a variação da exatidão dos modelos baseados em redes neuronais consoante a representação textual utilizada, no cenário sem balanceamento dos dados. De forma equivalente, a Tabela 49 e a Figura 42 mostram os mesmos resultados com aplicação de técnicas de balanceamento. A Figura 41 e Figura 43 complementam as análises anteriores ao ilustrar, de forma detalhada, a exatidão máxima alcançada por cada modelo baseado em redes neuronais para cada uma das representações textuais testadas.

Tabela 48 - Exatidão dos modelos baseados em redes neuronais para cada representação textual, incluindo a melhor e a pior exatidão máxima registradas e a respetiva variação (sem balanceamento dos dados). Fonte: Elaboração própria.

Modelo	FastText	Word2Vec	Melhor	Pior	Variação
<i>CNN</i>	0.744	0.737	0.744	0.737	0.007
<i>CNN-BiLSTM</i>	0.733	0.738	0.738	0.733	0.005

<i>GRU</i>	0.758	0.754	0.758	0.754	0.004
<i>BiLSTM</i>	0.737	0.736	0.737	0.736	0.001
<i>MLP</i>	0.726	0.725	0.726	0.725	0.001

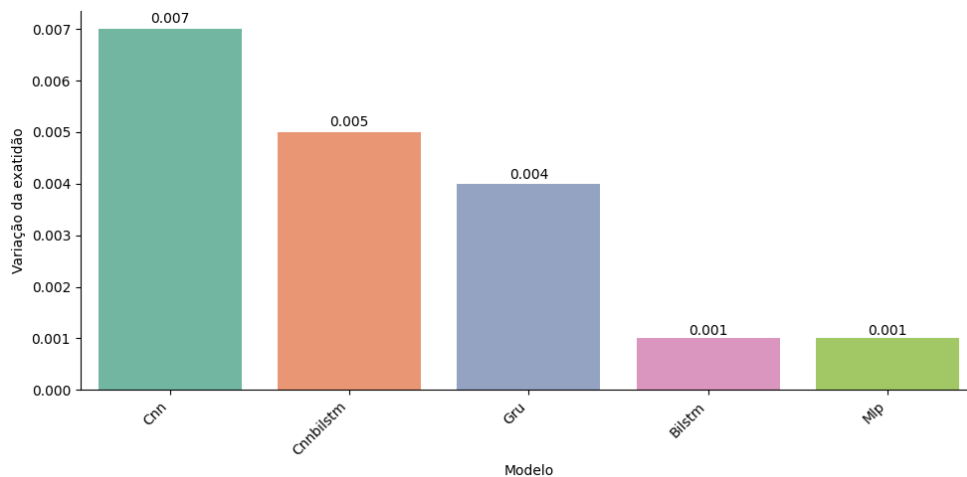


Figura 40 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre diferentes representações textuais, sem balanceamento dos dados. Fonte: Elaboração própria.

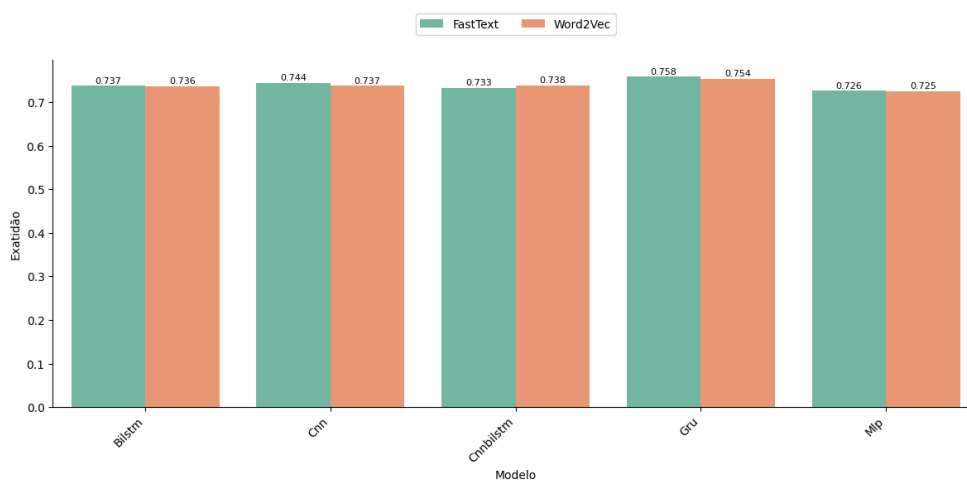


Figura 41 - Exatidão máxima dos modelos baseados em redes neuronais para cada representação textual (sem balanceamento dos dados). Fonte: Elaboração própria.

Tabela 49 - Exatidão dos modelos baseados em redes neuronais para cada representação textual, incluindo a melhor e a pior exatidão máxima registadas e a respetiva variação (com balanceamento dos dados). Fonte: Elaboração própria.

Modelo	FastText	Word2Vec	Melhor	Pior	Variação
<i>CNN-BiLSTM</i>	0.745	0.741	0.745	0.741	0.004
<i>BiLSTM</i>	0.739	0.742	0.742	0.739	0.003

<i>GRU</i>	0.751	0.748	0.751	0.748	0.003
<i>CNN</i>	0.746	0.744	0.746	0.744	0.002
<i>MLP</i>	0.727	0.725	0.727	0.725	0.002

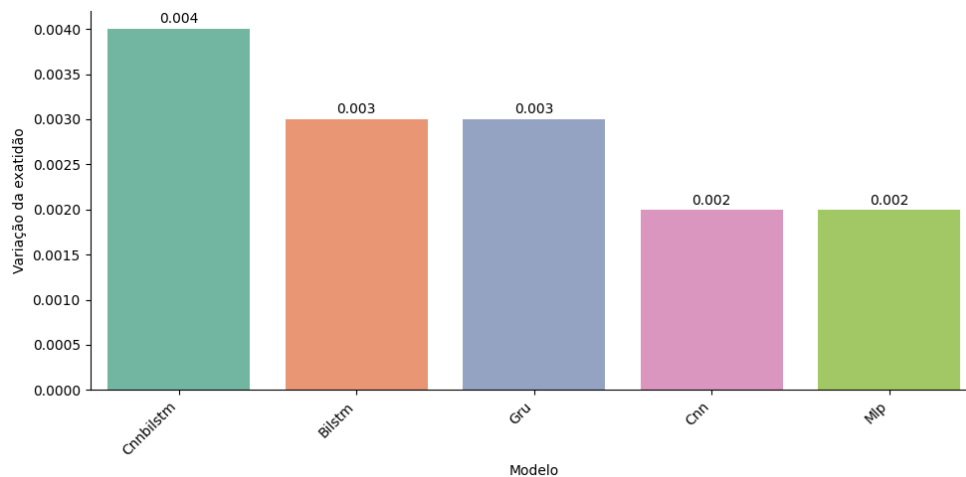


Figura 42 - Variação da exatidão máxima dos modelos baseados em redes neuronais, entre diferentes representações textuais, com balanceamento dos dados. Fonte: Elaboração própria.

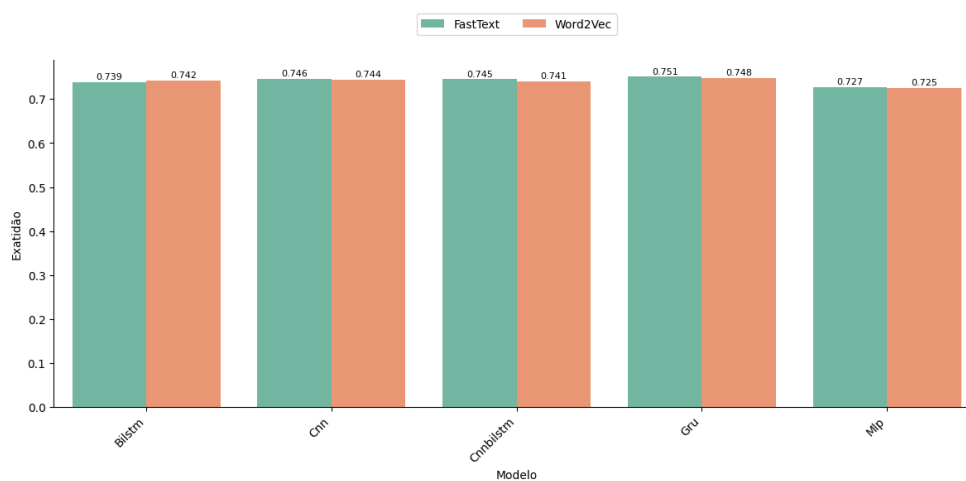


Figura 43 - Exatidão máxima dos modelos baseados em redes neuronais para cada representação textual (com balanceamento dos dados). Fonte: Elaboração própria.

Os resultados obtidos demonstram que, no geral, os modelos apresentam variações muito pequenas de exatidão entre diferentes representações textuais, revelando maior robustez face a esta escolha do que os modelos de aprendizagem computacional tradicional. Sem balanceamento, a *CNN* foi o modelo mais sensível, com uma variação de apenas **0.007**, enquanto os restantes tiveram diferenças marginais. Com balanceamento, a estabilidade foi ainda maior, com variações inferiores a **0.004**. Não se verificou uma superioridade

consistente entre as representações textuais testadas: o *FastText* teve ligeira vantagem sem balanceamento, e o *Word2Vec* mostrou-se ligeiramente melhor com balanceamento.

4.1.3.4 Melhores combinações

A Tabela 50 sintetiza as melhores combinações de pré-processamento, balanceamento e representação textual obtidas para cada modelo baseado em redes neurais.

Tabela 50 - Comparação das melhores combinações de pré-processamento, balanceamento, representação textual e desempenho para cada modelo baseado em redes neurais. Fonte: Elaboração própria.

Modelo	Melhor estratégia pré-processamento	Melhor estratégia balanceamento	Melhor estratégia repres. textual	Melhor exatidão média
GRU	LOW	NOB	FastText	0.758
<i>CNN</i>	LOW	USM	FastText	0.746
<i>CNN-BiLSTM</i>	LOW	USM	FastText	0.745
<i>BiLSTM</i>	LOW	USM	Word2Vec	0.742
<i>MLP</i>	LOW	USM	FastText	0.727

4.1.3.5 Análise ao desempenho computacional (utilização da GPU)

A Tabela 51 apresenta os valores médios de utilização de memória *GPU* (em MiB) por modelo, calculados com base em ambos os conjuntos de dados (com e sem balanceamento).

Tabela 51 - Utilização média de GPU (em MiB) durante o treino dos modelos baseados em redes neurais. Fonte: Elaboração própria.

Modelo	Utilização média GPU (MiB)
GRU	118.7
<i>CNN-BiLSTM</i>	103.2
<i>BiLSTM</i>	77.3
<i>CNN</i>	66.8
<i>MLP</i>	17.8

Os resultados obtidos demonstram que os modelos baseados em redes neurais apresentam diferenças significativas na utilização da *GPU*. O modelo *GRU* foi o mais exigente em termos de memória, seguido do *CNN-BiLSTM* e do *BiLSTM*. Por outro lado, o *MLP* revelou-se o mais leve, com uma utilização média substancialmente inferior.

4.1.4 Modelos baseados em transformadores

A Figura 44 mostra, para cada modelo, a melhor exatidão alcançada considerando a técnica de balanceamento aplicada. A Figura 45 representa os mesmos resultados sob a forma de mapa de calor, facilitando a visualização comparativa da distribuição dos desempenhos por modelo e técnica de pré-processamento. A Figura 46 evidencia a distribuição da exatidão média dos modelos ao longo de diferentes execuções.

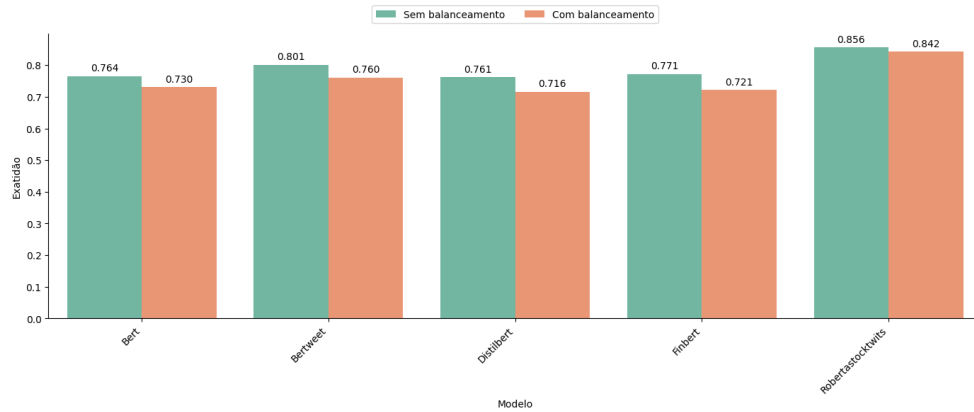


Figura 44 - Exatidão máxima por modelo baseado em transformadores, com e sem aplicação de técnicas de balanceamento. Fonte: Elaboração própria.

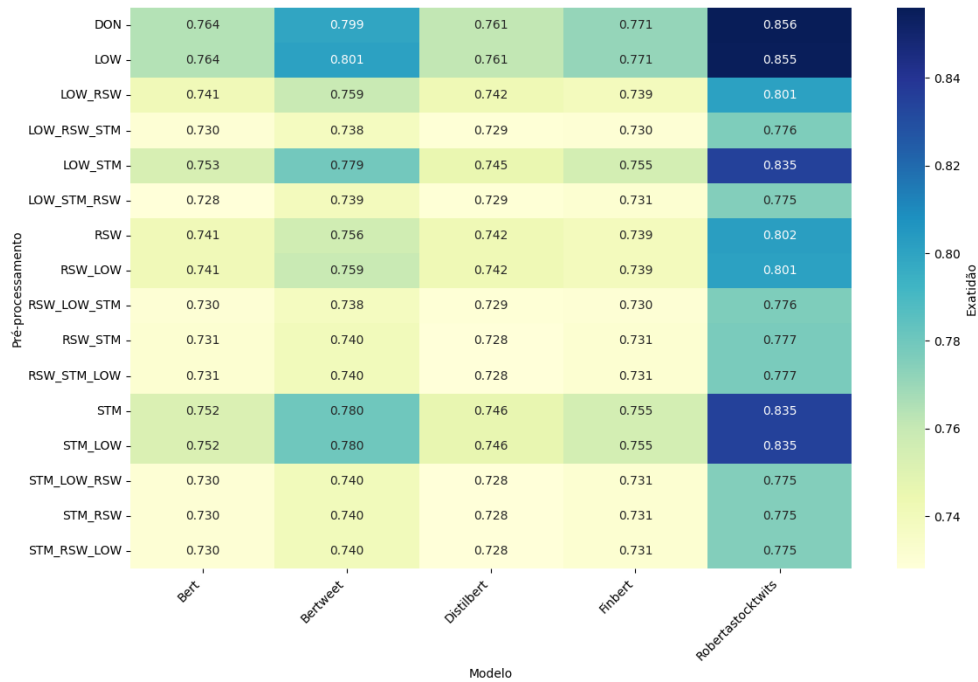


Figura 45 - Exatidão máxima por modelo baseado em transformadores e técnica de pré-processamento, com e sem balanceamento. Fonte: Elaboração própria.

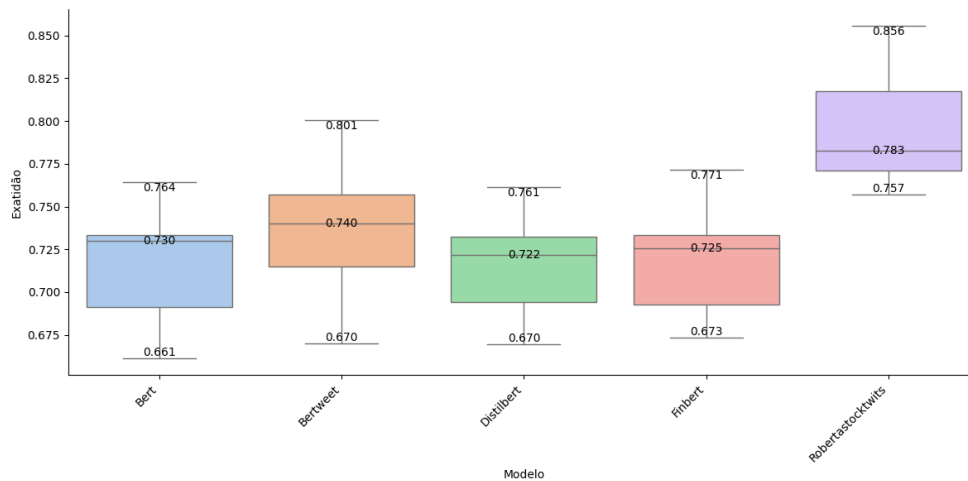


Figura 46 - Distribuição da exatidão dos modelos baseados em transformadores, nas diferentes execuções, com e sem balanceamento. Fonte: Elaboração própria.

Os resultados obtidos demonstram que o *RoBERTaStockTwits* obteve o melhor resultado global de toda a avaliação, com uma exatidão máxima de **0.856**, seguido do *BERTweet* (**0.801**) e do *FinBERT* (**0.771**). Este desempenho confirma a vantagem de modelos especializados e treinados num domínio específico para a classificação de sentimento financeiro. A Figura 46 mostra que o *RoBERTaStockTwits*, além de alcançar o valor mais alto, manteve uma mediana elevada (**0.783**) e baixa dispersão, o que demonstra consistência. Já o *BERTweet*, apesar do bom desempenho máximo, apresentou maior variabilidade, com uma amplitude interquartil mais larga e valores mínimos de **0.670**.

4.1.4.1 Análise de sensibilidade à técnica de balanceamento

A Tabela 52 apresenta os valores de exatidão máxima obtidos pelos modelos baseados em transformadores nos cenários com e sem aplicação de técnicas de balanceamento, bem como a variação absoluta entre ambos. A Figura 47 ilustra essas variações de forma visual.

Tabela 52 - Comparação da exatidão máxima dos modelos baseados em transformadores, com e sem balanceamento dos dados, incluindo a variação observada. Fonte: Elaboração própria.

Modelo	Sem balanceamento	Com balanceamento	Varição
<i>FinBERT</i>	0.771	0.721	0.050
<i>DistilBERT</i>	0.761	0.716	0.045
<i>BERTweet</i>	0.801	0.760	0.041
<i>BERT</i>	0.764	0.730	0.034
<i>RoBERTaStockTwits</i>	0.856	0.842	0.014

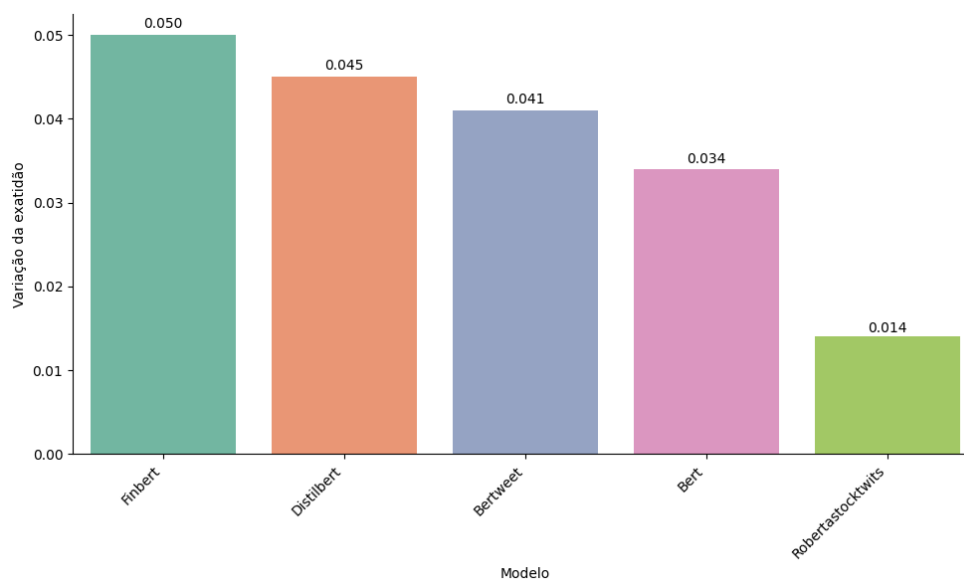


Figura 47 - Variação da exatidão máxima dos modelos baseados em transformadores, com e sem balanceamento dos dados. Fonte: Elaboração própria.

Os resultados obtidos demonstram que todos os modelos registaram uma diminuição de desempenho após a aplicação de balanceamento, sendo a maior variação no modelo *FinBERT* (**0.050**), seguido por *DistilBERT* (**0.045**) e *BERTweet* (**0.041**). O modelo *RoBERTaStockTwits* revelou-se o menos sensível, com uma variação de apenas **0.014**.

4.1.4.2 Análise de sensibilidade à técnica de pré-processamento

A Tabela 53 e Tabela 54 apresentam, para cada modelo baseado em transformadores, a melhor e a pior técnica de pré-processamento identificadas nos cenários sem e com aplicação de técnicas de balanceamento, respectivamente. A Figura 48 e Figura 49 representam graficamente essas variações de exatidão.

Tabela 53 - Comparação da exatidão máxima dos modelos baseados em transformadores, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (sem balanceamento dos dados). Fonte: Elaboração própria.

Modelo	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>RoBERTa StockTwits</i>	DON	0.856	LOW_STM_RSW, STM_LOW_RSW, STM_RSW, STM_RSW_LOW	0.775	0.081
<i>BERTweet</i>	LOW	0.801	LOW_RSW_STM, RSW_LOW_STM	0.738	0.063
<i>FinBERT</i>	DON, LOW	0.771	LOW_RSW_STM, RSW_LOW_STM	0.730	0.041
<i>BERT</i>	DON, LOW	0.764	LOW_STM_RSW	0.728	0.036

<i>DistilBERT</i>	DON, LOW	0.761	RSW_STM, RSW_STM_LOW, STM_LOW_RSW, STM_RSW, STM_RSW_LOW	0.728	0.033
-------------------	----------	-------	---	-------	-------

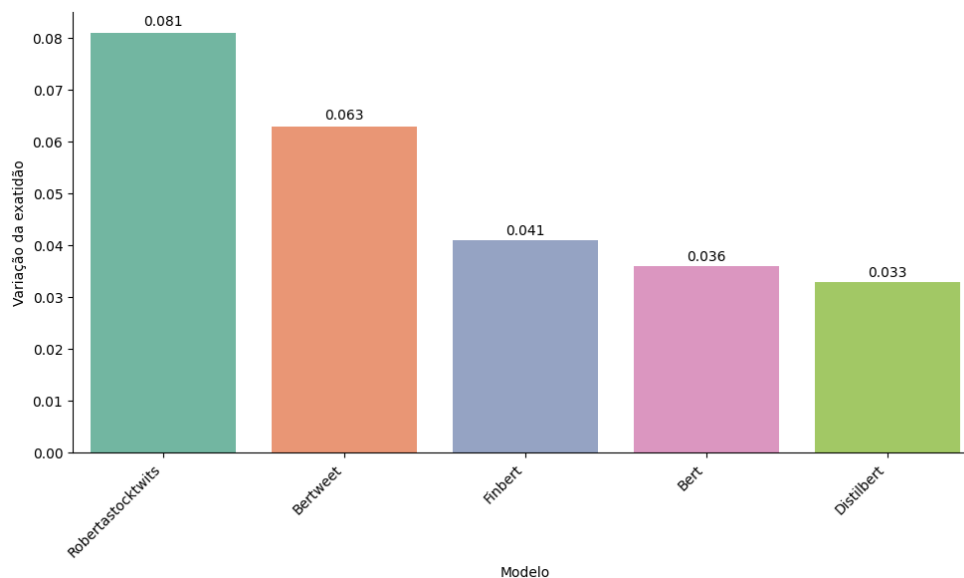


Figura 48 - Variação da exatidão máxima dos modelos baseados em transformadores, entre combinações de técnicas de pré-processamento, sem balanceamento dos dados. Fonte: Elaboração própria.

Tabela 54 - Comparação da exatidão máxima dos modelos baseados em transformadores, entre a melhor e a pior combinação de técnicas de pré-processamento, incluindo a variação observada (com balanceamento dos dados).

Modelo	Melhor(es) técnica(s)	Exatidão	Pior(es) técnica(s)	Exatidão	Variação
<i>BERTweet</i>	DON	0.760	STM_LOW_RSW, STM_RSW, STM_RSW_LOW	0.670	0.090
<i>RoBERTa StockTwits</i>	DON	0.842	LOW_STM_RSW, STM_LOW_RSW, STM_RSW, STM_RSW_LOW	0.757	0.085
<i>BERT</i>	DON, LOW	0.730	STM_LOW_RSW, STM_RSW, STM_RSW_LOW	0.661	0.069
<i>FinBERT</i>	DON, LOW	0.721	RSW_STM, RSW_STM_LOW	0.673	0.048
<i>DistilBERT</i>	DON, LOW	0.716	LOW_STM_RSW	0.670	0.046

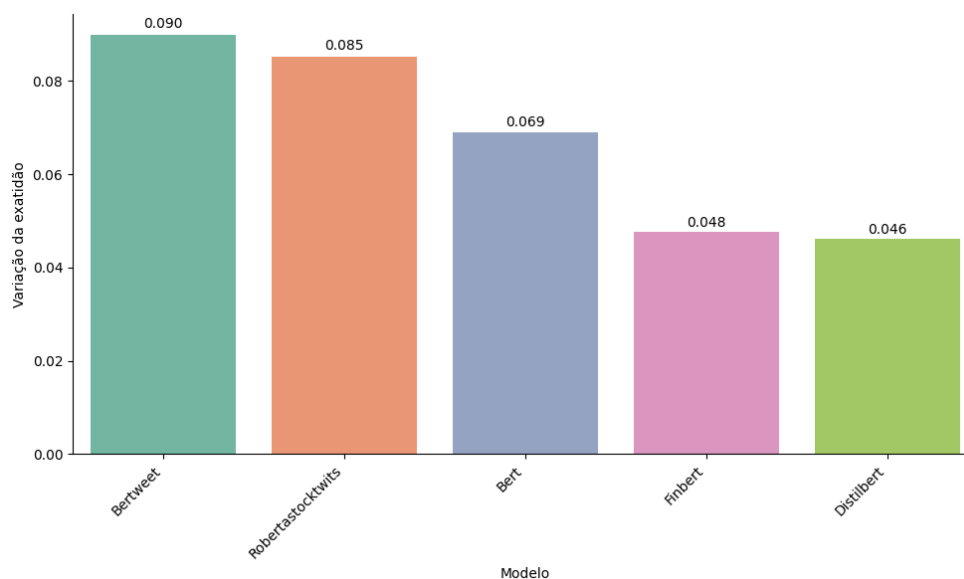


Figura 49 - Variação da exatidão máxima dos modelos baseados em transformadores, entre combinações de técnicas de pré-processamento, com balanceamento dos dados. Fonte: Elaboração própria.

Os resultados obtidos demonstram que os modelos baseados em transformadores são bastante sensíveis às técnicas de pré-processamento, independentemente de os dados estarem balanceados ou não. Sem balanceamento, o *RoBERTaStockTwits* teve a maior variação (**0.081**). Com balanceamento, foi o *BERTweet* a registrar a maior oscilação (**0.090**). De forma geral, técnicas simples geraram melhores resultados, enquanto combinações mais complexas tendem a prejudicar o desempenho.

4.1.4.3 Melhores combinações

A Tabela 55 apresenta as melhores combinações de pré-processamento e balanceamento identificadas para cada modelo baseado em transformadores, bem como a respectiva exatidão média.

Tabela 55 - Comparação das melhores combinações de pré-processamento, balanceamento e desempenho para cada modelo baseado em transformadores. Fonte: Elaboração própria.

Modelo	Melhor estratégia pré-processamento	Melhor estratégia balanceamento	Melhor exatidão média
<i>RoBERTaStockTwits</i>	DON	NOB	0.856
<i>BERTweet</i>	LOW	NOB	0.801
<i>FinBERT</i>	DON, LOW	NOB	0.771
<i>BERT</i>	DON, LOW	NOB	0.764
<i>DistilBERT</i>	DON, LOW	NOB	0.761

4.1.4.4 Análise ao desempenho computacional (utilização da GPU)

A Tabela 56 apresenta os valores médios de utilização de memória GPU (em MiB) por modelo, calculados com base em ambos os conjuntos de dados (com e sem balanceamento).

Tabela 56 - Utilização média de GPU (em MiB) durante o treino dos modelos baseados em transformadores. Fonte: Elaboração própria.

Modelo	Utilização média GPU (MiB)
<i>BERTweet</i>	4397.5
<i>RoBERTaStockTwits</i>	4247.4
<i>BERT</i>	4010.9
<i>FinBERT</i>	4010.1
<i>DistilBERT</i>	2220.9

Os resultados obtidos demonstram que os modelos baseados em transformadores apresentam um consumo de GPU substancialmente superior, com destaque para o *BERTweet* e o *RoBERTaStockTwits*. Tal como ilustrado na Tabela 30, estes modelos possuem um elevado número de parâmetros (135M e 125M, respetivamente). Esta complexidade explica o maior custo computacional associado e reforça a importância de ponderar os recursos disponíveis no processo de seleção do modelo.

4.2 Análise de padrões temporais entre o sentimento e o comportamento do mercado

A Tabela 57 e Tabela 58 apresentam as distâncias médias calculadas com o algoritmo *Dynamic Time Warping* (DTW) para diferentes valores de desfasamento (*lag*) entre o índice de sentimento agregado e os preços de fecho ajustado das ações, considerando janelas móveis de quatro dias. Esta análise foi conduzida sobre o período compreendido entre 1 de janeiro e 31 de dezembro de 2024.

Para cada ação, foram testados três valores de desfasamento temporal: sem atraso ($\tau = 0$), um dia antes ($\tau = 1$) e dois dias antes ($\tau = 2$). A Tabela 57 refere-se ao sentimento agregado dos comentários do *StockTwits*, enquanto a Tabela 58 apresenta os resultados

para o sentimento proveniente de notícias financeiras. O desfasamento ótimo corresponde ao valor de τ que minimiza a distância DTW média ao longo das janelas analisadas.

Tabela 57 - Distância DTW média para o sentimento dos comentários e respetivos desfasamentos ótimos por ação.
Fonte: Elaboração própria.

Ação	Lag $\tau = 0$	Lag $\tau = 1$	Lag $\tau = 2$	Lag τ ótimo
AAPL	0.540093	0.539046	0.538709	2
AMZN	0.648716	0.648678	0.650758	1
META	0.410402	0.409198	0.411169	1
MSFT	0.473569	0.473242	0.471721	2
NFLX	0.630243	0.632725	0.634936	0
GOOG	0.564478	0.563369	0.566095	1

Tabela 58 - Distância DTW média para o sentimento das notícias e respetivos desfasamentos ótimos por ação.
Fonte: Elaboração própria.

Ação	Lag $\tau = 0$	Lag $\tau = 1$	Lag $\tau = 2$	Lag τ ótimo
AAPL	0.909357	0.908610	0.907403	2
AMZN	0.948200	0.945879	0.945598	2
META	0.774333	0.775836	0.778893	0
MSFT	0.745222	0.743265	0.741567	2
NFLX	1.085823	1.088415	1.090788	0
GOOG	0.823361	0.825259	0.827083	0

A Tabela 59 apresenta os resultados do teste de alinhamento direcional entre o sentimento dos comentários do *StockTwits* e a direção do preço de fecho das ações, considerando o desfasamento ótimo previamente identificado para cada ativo. De forma análoga, a Tabela 60 reporta os resultados obtidos quando se considera o sentimento extraído de notícias financeiras.

Para cada ação, são apresentados os seguintes indicadores:

- **Lag (τ):** representa o desfasamento temporal aplicado ao índice de sentimento, correspondendo ao número de dias anteriores ao preço de fecho em que o sentimento é considerado. Este valor foi previamente determinado como ótimo com base na menor distância *DTW*.

- **Tamanho da janela (w):** indica o número de dias consecutivos considerados em cada janela deslizante. Em cada uma destas janelas é verificada a semelhança na direção (subida ou descida) entre o sentimento e o preço.
- **Taxa de alinhamentos (A):** corresponde à proporção de janelas em que a direção do sentimento coincidiu exatamente com a direção do preço em todos os dias da janela, sendo calculada por: $A = \frac{N_{\text{alinhadas}}}{N_{\text{total}}}$.
- **Número de janelas alinhadas ($N_{\text{alinhadas}}$):** total de janelas em que se verificou alinhamento perfeito entre a direção do sentimento e a direção do preço.
- **Total de janelas válidas (N_{total}):** número total de janelas deslizantes consideradas para a ação em causa, após aplicação do desfasamento.
- **Média das permutações (\bar{N}_{perm}):** valor médio de janelas alinhadas obtido ao repetir o processo com 1000 permutações aleatórias da série de sentimento, preservando os preços.
- **Máximo das permutações ($N_{\text{perm}}^{\text{máx}}$):** maior número de janelas alinhadas observado em qualquer uma das 1000 permutações aleatórias.
- **Valor-p (p):** proporção de permutações em que o número de alinhamentos foi igual ou superior ao observado.

Tabela 59 - Resultados do teste de alinhamento direcional com permutação para cada ação (comentários StockTwits). Fonte: Elaboração própria.

Ação	τ	w	A	$N_{\text{alinhadas}}$	N_{total}	\bar{N}_{perm}	$N_{\text{perm}}^{\text{máx}}$	p
AAPL	2	4	0.067	24	360	16.600	35	0.079
AMZN	1	4	0.044	16	361	15.870	39	0.508
META	1	4	0.066	24	361	17.190	37	0.105
MSFT	2	4	0.044	16	360	15.860	34	0.515
NFLX	0	4	0.072	26	361	16.400	34	0.035
GOOG	1	4	0.097	35	361	16.670	39	0.001

No caso dos comentários, destacam-se duas ações com resultados relevantes. Para a ação *GOOG*, foram observadas 35 janelas alinhadas ($\frac{35}{361} \approx 0.097$), comparando com uma média de apenas 16,670 nas permutações e um máximo de 39, o que resultou num $p = 0.001$, indicando que, entre todas as 1000 permutações realizadas, raramente se obteve um número de janelas alinhadas igual ou superior ao observado. Também a ação *NFLX*

apresentou 26 janelas alinhadas ($\frac{26}{361} \approx 0.072$), acima da média de 16,400 e próximo do máximo de 34, entre todas as 1000 permutações realizadas ($p = 0.035$). Estes resultados sugerem que o sentimento expresso nos comentários dos utilizadores poderia antecipar ou acompanhar a direção dos preços, em especial nestas duas ações.

*Tabela 60 - Resultados do teste de alinhamento direcional com permutação para cada ação (notícias financeiras).
Fonte: Elaboração própria.*

Ação	τ	w	A	$N_{\text{alinhadas}}$	N_{total}	\bar{N}_{perm}	$N_{\text{perm}}^{\text{máx}}$	p
AAPL	2	4	0.025	9	360	14.540	29	0.912
AMZN	2	4	0.042	15	360	12.310	27	0.274
META	0	4	0.036	13	361	13.200	28	0.536
MSFT	2	4	0.036	13	360	12.540	27	0.475
NFLX	0	4	0.019	7	361	6.930	19	0.524
GOOG	0	4	0.036	13	361	14.400	31	0.634

Os resultados das notícias evidenciam uma tendência geral de alinhamento mais fraca, sem significância estatística em nenhuma ação. Os números de janelas alinhadas foram reduzidos e próximos da média das permutações. Estes valores confirmam que, para as notícias, não foi identificada evidência estatisticamente relevante de alinhamento direcional.

5.1 Síntese dos resultados

As conclusões apresentadas tiveram como objetivo dar resposta às duas questões de investigação definidas no início do trabalho.

Questão de Investigação 1: Qual é a combinação mais eficaz de modelos e técnicas de processamento de linguagem natural para otimizar o desempenho da análise de sentimento no contexto do mercado financeiro de ações?

Nas abordagens baseadas em léxicos, o *FinVADER*, conseguiu a maior exatidão dentro desta categoria, justificada pelo facto de ter sido concebido para o domínio financeiro. Na aprendizagem computacional tradicional, o modelo com melhor desempenho foi o *XGBoost*, justificado pelo facto de ser uma técnica de conjunto, capaz de combinar múltiplas árvores de decisão para melhorar a generalização. No que diz respeito aos modelos baseados em redes neuronais, a arquitetura *GRU* destacou-se apresentando maior exatidão, face a arquiteturas mais complexas. Os modelos baseados em transformadores obtiveram os melhores resultados globais em termos de exatidão, com destaque para o *RoBERTaStockTwits*, atingindo uma exatidão máxima de 84.2% e 85.6%, nos cenários com e sem balanceamento de dados, respetivamente, e sem aplicação de técnicas de pré-processamento. A diferença de desempenho entre este modelo e os restantes transformadores, incluindo modelos genéricos e modelos treinados especificamente em dados financeiros (*FinBERT*) e em redes sociais (*BERTweet*), foi considerável, demonstrando a importância do processo de ajuste fino sobre uma grande quantidade de dados do mesmo domínio.

No que respeita às técnicas de pré-processamento, constatou-se que as transformações mais simples, como a conversão de texto para minúsculas, ou mesmo a própria ausência de qualquer técnica de pré-processamento, de uma forma geral, produziram resultados superiores às combinações mais complexas. Isto foi especialmente evidente nos modelos baseados em transformadores, em que, na maior parte dos casos, o impacto de pré-processamentos com maior complexidade resultou em degradação do desempenho.

Verificou-se que as representações textuais esparsas, baseadas em frequências de palavras, como *Bag-of-Words* e *TF-IDF*, fizeram com que a maioria dos modelos de aprendizagem computacional tradicional conseguissem obter melhor desempenho. Já para os modelos baseados em redes neuronais, entre as representações densas, o *FastText* demonstrou um desempenho ligeiramente superior.

Quanto à técnica de balanceamento, a redução da amostra da classe majoritária apresentou diferentes resultados em cada categoria de modelos. Nos modelos de aprendizagem computacional tradicional, em regra, observou-se impacto considerável do balanceamento, com os melhores resultados a surgirem em dados não balanceados. A exceção foi a *Decision Tree* que não apresentou variações significativas entre os dois cenários. Nos modelos baseados em transformadores, o impacto foi também evidente, onde os melhores resultados foram obtidos com dados não balanceados. A única exceção foi o *RoBERTaStockTwits*, no qual o efeito do balanceamento foi menos pronunciado, não resultando, ainda assim, em melhorias. Já nos modelos baseados em redes neurais, observaram-se comportamentos mistos. Alguns modelos atingiram melhores resultados em dados balanceados, enquanto noutros casos ocorreu o inverso. De forma geral, o efeito do balanceamento nesta categoria revelou-se muito pouco pronunciado.

A análise do desempenho computacional de utilização de memória da *GPU* evidenciou diferenças entre as categorias de modelos testados e que conseguiram tirar partido deste recurso. Os modelos baseados em transformadores revelaram um elevado consumo de memória, com valores médios de utilização de *GPU* superiores a 4GiB. Os modelos baseados em redes neurais apresentaram um consumo de *GPU* significativamente inferior.

Questão de Investigação 2: De que forma os sentimentos gerados a partir dos comentários e das notícias se correlacionam com a variação diária do valor das ações?

A análise temporal realizada com *Dynamic Time Warping* e os testes de significância estatística revelaram resultados distintos consoante sejam comentários ou notícias. No caso dos comentários, foram identificados alinhamentos estatisticamente significativos com a direção dos preços em algumas ações, nomeadamente na *GOOG* ($p = 0.001$) e na *NFLX* ($p = 0.035$). Estes resultados sugerem que o sentimento extraído de redes sociais poderá ter a capacidade de antecipar ou acompanhar, de forma mais direta, as variações dos preços de fecho das ações. Por outro lado, o sentimento de notícias financeiras apresentou uma relação muito fraca e menos consistente com a variação dos preços.

À luz da Teoria do Mercado Eficiente, os resultados obtidos poderão não ter dimensão e consistência suficiente para a contrariar de uma forma clara e evidente. No entanto, os indícios observados apontam para a possibilidade de existência de ineficiências de curto

prazo, particularmente no que se refere ao impacto dos comentários nas redes sociais, sugerindo que o mercado poderá não reagir de forma totalmente instantânea à informação pública. Estas evidências poderão vir a ser exploradas em trabalhos futuros.

5.2 Limitações do estudo

Este estudo apresenta algumas limitações que importa reconhecer, tanto ao nível metodológico como no que respeita aos recursos disponíveis, que devem ser tidas em conta na interpretação dos resultados.

A capacidade computacional disponível para o treino dos modelos foi um fator limitativo. A utilização de hardware com poucos recursos, incluindo uma *GPU* com memória reduzida, condicionou a possibilidade de aumentar a complexidade dos modelos a treinar e o volume de dados passível de ser processado. No caso dos modelos baseados em redes neuronais e transformadores, o ajuste fino foi realizado sobre um conjunto de dados reduzido. Este processo deveria ter sido efetuado sobre um maior número de exemplos e ao longo de mais épocas.

Importa ainda referir a limitação associada às fontes de dados no que diz respeito às notícias financeiras. A análise baseou-se numa única fonte de informação, que poderá não refletir de forma adequada a diversidade, o volume e a atualidade das notícias relevantes para o mercado.

Por fim, a escolha dos modelos considerados poderia ter sido mais abrangente. No campo dos modelos baseados em transformadores, a opção de incluir arquiteturas com decodificadores poderia trazer resultados interessantes sobre a capacidade destes modelos.

5.3 Trabalho futuro

Uma possibilidade de trabalho futuro, e dando sequência a esta investigação, na qual se identificou a existência de padrões temporais, poderá passar pelo desenvolvimento de modelos preditivos (regressão ou classificação) com capacidade de quantificar a relação entre os índices de sentimento e as variações diárias do preço de fecho das ações, com o objetivo de identificar eventuais ineficiências na incorporação de informação pública nos

preços de fecho das ações, e desta forma, avaliar a possibilidade de obtenção de lucros consistentes, o que poderia contrariar a Teoria do Mercado Eficiente.

Este tipo de abordagem poderá exigir engenharia de atributos, procurando identificar e combinar variáveis que representem não apenas as características do sentimento (polaridade, intensidade, volume de mensagens e número de seguidores dos autores), mas também fatores de mercado relevantes (volatilidade ao longo do dia, volume de negociação, ou distribuição de resultados).

REFERÊNCIAS BIBLIOGRÁFICAS

- Abdelwahab, O., Bahgat, M., Lowrance, C. J., & Elmaghraby, A. (2015). Effect of training set size on SVM and Naïve Bayes for Twitter sentiment analysis. *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 46–51. <https://doi.org/10.1109/ISSPIT.2015.7394379>
- Aftab, F., Bazai, S. U., Marjan, S., Baloch, L., Aslam, S., Amphawan, A., & Neo, T. K. (2023). A Comprehensive Survey on Sentiment Analysis Techniques. *International Journal of Technology*, 14(6), 1288–1298. <https://doi.org/10.14716/IJTECH.V14I6.6632>
- Agarwal, K., & Tewari, D. (2025). Exploring emotional influences on investment choices: Stock Selection and Trading. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5114355>
- Aguilar-Moreno, J. A., Palos-Sanchez, P. R., & Pozo-Barajas, R. (2024). Sentiment analysis to support business decision-making. A bibliometric study. *AIMS Mathematics*, 9(2), 4337–4375. <https://doi.org/10.3934/math.2024215>
- Ahmad, H. O., & Umar, S. U. (2023). Sentiment Analysis of Financial Textual data Using Machine Learning and Deep Learning Models. *Informatica*, 47(5), 153–158. <https://doi.org/10.31449/inf.v47i5.4673>
- Airlangga, G. (2024). Spam Detection in YouTube Comments Using Deep Learning Models: A Comparative Study of MLP, CNN, LSTM, BiLSTM, GRU, and Attention Mechanisms. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1533–1538. <https://doi.org/10.57152/malcom.v4i4.1671>
- Almeida, F., & Xexéo, G. (2023). *Word Embeddings: A Survey*. <http://arxiv.org/abs/1901.09069>
- Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., & Muneer, A. (2023). LSTM Inefficiency in Long-Term Dependencies Regression Problems. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 30(3), 16–31. <https://doi.org/10.37934/araset.30.3.1631>
- Al-Shabi, M. A. (2020). Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining. *IJCSNS International Journal of Computer Science and Network Security*, 20, 11–18.

- Alshahrani, M., Zhu, F., Sameh, A., Zheng, L., & Mumtaz, S. (2018). Evaluating the Influence of Twitter on the Saudi Arabian Stock Market Indicators. *Em Advances in Intelligent Systems and Computing* (Vol. 753, pp. 113–132). Springer Verlag. https://doi.org/10.1007/978-3-319-78753-4_10
- Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data 2023 10:1, 10(1)*, 1–82. <https://doi.org/10.1186/S40537-023-00727-2>
- Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., & Manicardi, S. (2016, Dezembro). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. *Proceedings of the 2nd International Workshop on Knowledge Discovery on the WEB*.
- Ankenbrand, M. J., Shainberg, L., Hock, M., Lohr, D., & Schreiber, L. M. (2021). Sensitivity analysis for interpretation of machine learning based segmentation models in cardiac MRI. *BMC Medical Imaging, 21(1)*, 27. <https://doi.org/10.1186/s12880-021-00551-1>
- Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *CoRR, abs/1908.10063*. <http://arxiv.org/abs/1908.10063>
- Ardakani, A., Haan, A., Tan, S., Popovici, D. T., Cheung, A., Iancu, C., & Sen, K. (2024). SlimFit: Memory-Efficient Fine-Tuning of Transformer-based Models Using Training Dynamics. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 1, 6218–6236. <https://doi.org/10.18653/v1/2024.naacl-long.345>
- Areshey, A., & Mathkour, H. (2024). Exploring transformer models for sentiment classification: A comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet. *Expert Systems, 41(11)*. <https://doi.org/10.1111/exsy.13701>

- Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review*, 56(9), 10345–10425. <https://doi.org/10.1007/s10462-023-10419-1>
- Baker, B. (2025, Janeiro 28). *What is the Nasdaq Composite?* Bankrate. <https://www.bankrate.com/investing/what-is-nasdaq-composite>
- Baker, M. P., & Wurgler, J. A. (2003). Investor Sentiment and the Cross-section of Stock Returns. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.464843>
- Batra, R., & Daudpota, S. M. (2018). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018-January*, 1–5. <https://doi.org/10.1109/ICOMET.2018.8346382>
- Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies (ICOT)*, 1–6. <https://doi.org/10.1109/ICOT.2018.8705796>
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281–305.
- Bharal, R., & V Vamsi Krishna, O. (2021). Social Media Sentiment Analysis Using CNN-BiLSTM. *International Journal of Science and Research (IJSR)*, 10(9), 656–661. <https://doi.org/10.21275/SR21913110537>
- Bhargavi, A. D. (2025). Comparative Study of Static and Contextual Text Vectorization for Sentiment Analysis. *International Journal for Research in Applied Science and Engineering Technology*, 13(7), 484–488. <https://doi.org/10.22214/ijraset.2025.73045>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. <https://doi.org/10.1016/J.KNOSYS.2021.107134>
- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A. L., Deng, D., & Lindauer, M. (2021). Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges. *Wiley*

- Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2).
<https://doi.org/10.1002/widm.1484>
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, 135–146. https://doi.org/10.1162/tacl_a_00051
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, 56(11), 12505–12560. <https://doi.org/10.1007/s10462-023-10442-2>
- Bozanta, A., Angco, S., Cevik, M., & Basar, A. (2021, Dezembro). Sentiment Analysis of StockTwits Using Transformer Models. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1253–1258. <https://doi.org/10.1109/ICMLA52953.2021.00204>
- Branco, P., Torgo, L., & Ribeiro, R. P. (2017). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys*, 49(2). <https://doi.org/10.1145/2907070>;WGROUP:STRING:ACM
- Çano, E., & Morisio, M. (2019). *Word Embeddings for Sentiment Analysis: A Comprehensive Empirical Survey*. <http://arxiv.org/abs/1902.00753>
- Cao, F., Zhang, G., Yao, J., Ni, S., & Zhang, Y. (2022). *Sentiment Inferencing model for stock related comments*. Hugging Face. <https://huggingface.co/zhayunduo/roberta-base-stocktwits-finetuned>
- Cao, L., & Tay, F. E. H. (2001). Financial Forecasting Using Support Vector Machines. *Neural Computing & Applications*, 10(2), 184–192. <https://doi.org/10.1007/s005210170010>
- Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling approaches to handle class imbalance: a review from a data perspective. *Journal of Big Data* 2025 12:1, 12(1), 1–58. <https://doi.org/10.1186/S40537-025-01119-4>

- Casola, S., Lauriola, I., & Lavelli, A. (2022). Pre-trained transformers: an empirical comparison. *Machine Learning with Applications*, 9, 100334. <https://doi.org/10.1016/J.MLWA.2022.100334>
- Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, 44, 65–77. <https://doi.org/10.1016/j.inffus.2017.12.006>
- Chinnalagu, A., & Durairaj, A. K. (2022). Comparative Analysis of BERT-base Transformers and Deep Learning Sentiment Prediction Models. *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, 874–879. <https://doi.org/10.1109/SMART55829.2022.10047651>
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1724–1734. <https://doi.org/10.3115/V1/D14-1179>
- Danyal, M. M., Khan, S. S., Khan, M., Ghaffar, M. B., Khan, B., & Arshad, M. (2023). Sentiment Analysis Based on Performance of Linear Support Vector Machine and Multinomial Naïve Bayes Using Movie Reviews with Baseline Techniques. *Journal on Big Data*, 5, 1–18. <https://doi.org/10.32604/JBD.2023.041319>
- Darwish, M., Hassanien, E. E., & Eissa, A. H. B. (2025). Stock Market Forecasting: From Traditional Predictive Models to Large Language Models. *Computational Economics*, 1–45. <https://doi.org/10.1007/s10614-025-11024-w>
- de Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. *Applied Soft Computing*, 133, 109924. <https://doi.org/10.1016/j.asoc.2022.109924>
- Delgadillo, J., Kinyua, J., & Mutigwe, C. (2024). FinSoSent: Advancing Financial Market Sentiment Analysis through Pretrained Large Language Models. *Big Data and Cognitive Computing*, 8(8), 87. <https://doi.org/10.3390/bdcc8080087>
- Denecke, K., & Reichenpfader, D. (2023). Sentiment analysis of clinical narratives: A scoping review. *Journal of Biomedical Informatics*, 140, 104336. <https://doi.org/10.1016/J.JBI.2023.104336>

- Deng, M. (2007). Death of the Efficient Market Hypothesis. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.1006716>
- Deveikyte, J., Geman, H., Piccari, C., & Proveti, A. (2022). A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence*, 5. <https://doi.org/10.3389/frai.2022.836809>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1, 4171–4186. <https://arxiv.org/abs/1810.04805v2>
- Di Leo, G., & Sardanelli, F. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European Radiology Experimental*, 4(1), 1–8. <https://doi.org/10.1186/S41747-020-0145-Y/METRICS>
- Dimopoulos, T., & Bakas, N. (2019). Sensitivity Analysis of Machine Learning Models for the Mass Appraisal of Real Estate. Case Study of Residential Units in Nicosia, Cyprus. *Remote Sensing 2019, Vol. 11, Page 3047, 11(24)*, 3047. <https://doi.org/10.3390/RS11243047>
- Divernois, M.-A., & Filipović, D. (2024). StockTwits classified sentiment and stock returns. *Digital Finance*, 6(2), 249–281. <https://doi.org/10.1007/s42521-023-00102-z>
- Drossu, R., & Obradovic, Z. (1997). INFFC data analysis: lower bounds and testbed design recommendations. *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFER)*, 71–74. <https://doi.org/10.1109/CIFER.1997.618915>
- Du, K., Xing, F., Mao, R., & Cambria, E. (2024). Financial Sentiment Analysis: Techniques and Applications. *ACM Computing Surveys*, 56(9), 1–42. <https://doi.org/10.1145/3649451>
- Esuli, A., & Sebastiani, F. (2006, Maio). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of the Fifth International*

- Conference on Language Resources and Evaluation (LREC'06)*.
http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383. <https://doi.org/10.2307/2325486>
- Gao, X., Xie, D., Zhang, Y., Wang, Z., He, C., Yin, H., & Zhang, W. (2025). *A Comprehensive Survey on Imbalanced Data Learning*. <http://arxiv.org/abs/2502.08960>
- Ghoshal, S., & Roberts, S. (2016). Extracting predictive information from heterogeneous data streams using Gaussian Processes. *Algorithmic Finance*, 5(1–2), 21–30. <https://doi.org/10.3233/AF-160055>
- Giachanou, A., & Crestani, F. (2017). Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Computing Surveys*, 49(2), 1–41. <https://doi.org/10.1145/2938640>
- Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P., & Pandey, N. (2021). Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science*, 7, 1–21. <https://doi.org/10.7717/PEERJ-CS.340/SUPP-1>
- González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011). Identifying Sarcasm in Twitter: A Closer Look. Em D. Lin, Y. Matsumoto, & R. Mihalcea (Eds.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 581–586). Association for Computational Linguistics. <https://aclanthology.org/P11-2102>
- Good, P. (2005). Permutation, Parametric and Bootstrap Tests of Hypotheses. Em *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer-Verlag. <https://doi.org/10.1007/b138696>
- Goodell, J. W., Kumar, S., Rao, P., & Verma, S. (2023). Emotions and stock market anomalies: A systematic review. *Journal of Behavioral and Experimental Finance*, 37, 100722. <https://doi.org/10.1016/j.jbef.2022.100722>

- Graves, A., & Schmidhuber, J. (2005). Frameworkise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5–6), 602–610. <https://doi.org/10.1016/J.NEUNET.2005.06.042>
- Gupta, R. (2024). Bidirectional encoders to state-of-the-art: a review of BERT and its transformative impact on natural language processing. *Информатика. Экономика. Управление - Informatics. Economics. Management*, 3(1), 0311–0320. <https://doi.org/10.47813/2782-5280-2024-3-1-0311-0320>
- Gutiérrez-Fandiño, A., Kolm, P. N., i Alonso, M. N., & Armengol-Estapé, J. (2022). FinEAS: Financial Embedding Analysis of Sentiment. *The Journal of Financial Data Science*, 4(3), 45–53. <https://doi.org/10.3905/jfds.2022.1.095>
- Haque, M. R., Akter Lima, S., & Mishu, S. Z. (2019). Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews. *3rd International Conference on Electrical, Computer and Telecommunication Engineering, ICECTE 2019*, 161–164. <https://doi.org/10.1109/ICECTE48615.2019.9303573>
- Hasib, K. Md., Iqbal, Md. S., Shah, F. M., Al Mahmud, J., Popel, M. H., Showrov, Md. I. H., Ahmed, S., & Rahman, O. (2020). A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem. *Journal of Computer Science*, 16(11), 1546–1557. <https://doi.org/10.3844/jcssp.2020.1546.1557>
- Hayes, A. (2025, Agosto 8). *Nasdaq Explained: History, Trading System, and Financial Insights*. Investopedia. <https://www.investopedia.com/terms/n/nasdaq.asp>
- Hillard, D., Ostendorf, M., & Shriberg, E. (2003). Detection of agreement vs. disagreement in meetings. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology companion volume of the Proceedings of HLT-NAACL 2003-short papers - NAACL '03*, 34–36. <https://doi.org/10.3115/1073483.1073495>
- Holt, C. A., & Sullivan, S. P. (2023). Permutation tests for experimental data. *Experimental Economics*, 26(4), 775–812. <https://doi.org/10.1007/s10683-023-09799-6>

- Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, 40(2), 806–841. <https://doi.org/10.1111/1911-3846.12832>
- Huang, F., Li, X., Yuan, C., Zhang, S., Zhang, J., & Qiao, S. (2022). Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9), 4332–4345. <https://doi.org/10.1109/TNNLS.2021.3056664>
- Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://doi.org/10.1609/ICWSM.V8I1.14550>
- Ilgun, H. E., & Kilic, E. (2021). Sentiment Analysis using Transformers and Machine Learning Models. *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 42–45. <https://doi.org/10.1109/UBMK52708.2021.9558931>
- Iooss, B., & Lemaître, P. (2014). A review on global sensitivity analysis methods. *Operations Research/ Computer Science Interfaces Series*, 59, 101–122. https://doi.org/10.1007/978-1-4899-7547-8_5
- Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. *Artificial Intelligence Review*, 57(3), 1–79. <https://doi.org/10.1007/S10462-023-10651-9/FIGURES/13>
- Jaggi, M., Mandal, P., Narang, S., Naseem, U., & Khushi, M. (2021). Text Mining of Stocktwits Data for Predicting Stock Prices. *Applied System Innovation*, 4(1), 1–22. <https://doi.org/10.3390/asi4010013>
- John, M., & Ferbinteanu, J. (2021). Detecting time lag between a pair of time series using visibility graph algorithm. *Communications in Statistics: Case Studies, Data*

Analysis and Applications, 7(3), 315–343.
<https://doi.org/10.1080/23737484.2021.1882356>

Jurek, A., Mulvenna, M. D., & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, 4(1).
<https://doi.org/10.1186/S13388-015-0024-X>

Kaur, G., & Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data*, 10(1), 1–23. <https://doi.org/10.1186/S40537-022-00680-6/TABLES/5>

Kedas, S., Kumar, A., & Jain, P. K. (2022). Dealing with Class Imbalance in Sentiment Analysis Using Deep Learning and SMOTE. In *Advances in Data Computing, Communication and Security* (pp. 407–416). https://doi.org/10.1007/978-981-16-8403-6_37

Khomsah, S., Ramadhani, R., & Wijaya, S. (2022). The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(3), 352–358.
<https://doi.org/10.29207/resti.v6i3.3711>

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. <https://doi.org/10.3115/v1/D14-1181>

Kolchyna, O., Souza, T. T. P., Treleaven, P., & Aste, T. (2015). *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*. <http://arxiv.org/abs/1507.00955>

Komariah, K. S., Machbub, C., Prihatmanto, A. S., & Sin, B.-K. (2016). A Study on Efficient Market Hypothesis to Predict Exchange Rate Trends Using Sentiment Analysis of Twitter Data. *Journal of Korea Multimedia Society*, 19(7), 1107–1115.
<https://doi.org/10.9717/kmms.2016.19.7.1107>

Konstantinidis, T., Iacovides, G., Xu, M., Constantinides, T. G., & Mandic, D. (2024). *FinLlama: Financial Sentiment Classification for Algorithmic Trading Applications*. <http://arxiv.org/abs/2403.12285>

- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Lee, C.-C., Huang, H.-C., & Tsai, C.-L. (2021). Impact of Investor Sentiment on Stock Returns Using Fine-tune BERT. *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 1–2. <https://doi.org/10.1109/ICCE-TW52618.2021.9602873>
- Lee, J., Youn, H. L., Poon, J., & Han, S. C. (2023). *StockEmotions: Discover Investor Emotions for Financial Sentiment Analysis and Multivariate Time Series*. <http://arxiv.org/abs/2301.09279>
- Li, Q., & Shah, S. (2017). Learning Stock Market Sentiment Lexicon and Sentiment-Oriented Word Vector from StockTwits. *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, 301–310. <https://doi.org/10.18653/v1/K17-1031>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-02145-9>
- Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. *Em Mining Text Data* (Vol. 9781461432234, pp. 415–463). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_13
- Liu, J., Leu, J., & Holst, S. (2023). Stock price movement prediction based on Stocktwits investor sentiment using FinBERT and ensemble SVM. *PeerJ Computer Science*, 9, e1403. <https://doi.org/10.7717/peerj-cs.1403>
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Lu, J. (2025). Text vectorization in sentiment analysis: A comparative study of TF-IDF and Word2Vec from Amazon Fine Food Reviews. *ITM Web of Conferences*, 70, 03001. <https://doi.org/10.1051/itmconf/20257003001>

- Mabrouk, A., Redondo, R. P. D., & Kayed, M. (2020). Deep Learning-Based Sentiment Classification: A Comparative Survey. *IEEE Access*, 8, 85616–85638. <https://doi.org/10.1109/ACCESS.2020.2992013>
- Man, X., Luo, T., & Lin, J. (2019). Financial Sentiment Analysis(FSA): A Survey. *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, 617–622. <https://doi.org/10.1109/ICPHYS.2019.8780312>
- Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
- Manning, L. (2024, Outubro 1). *Financial Industry Regulatory Authority (FINRA) Definition*. Investopedia. <https://www.investopedia.com/terms/f/finra.asp>
- Mäntylä, M. V., Graziotin, D., & Kuuttila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>
- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36(4), 102048. <https://doi.org/10.1016/j.jksuci.2024.102048>
- Marozzi, M. (2004). Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica*, 193–201. <https://doi.org/10.6092/issn.1973-2201/32>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Marutho, D., Muljono, Rustad, S., & Purwanto. (2022). Sentiment Analysis Optimization Using Vader Lexicon on Machine Learning Approach. *2022 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 98–103. <https://doi.org/10.1109/ISITIA56226.2022.9855341>

- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/J.ASEJ.2014.04.011>
- Mehenaoui, Z., Merabti, C., Tadjer, H., & Lafifi, Y. (2024). A Comparative Study On Sentiment Lexicons For Automatic Labeling. *Proceedings of the 13th International Conference on Research in computing at Feminine (RIF 2024)*, 152–158.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. <http://arxiv.org/abs/1301.3781>
- Min, B., Ross, H., Sulem, E., Veyseh, A. P. Ben, Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., & Roth, D. (2021). Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. *ACM Computing Surveys*, 56(2). <https://doi.org/10.1145/3605943>
- Mishra, S. (2021). Drug Review Sentiment Analysis using Boosting Algorithms. *International Journal of Trend in Scientific Research and Development*, Volume-5(Issue-4). www.ijtsrd.com
- Mohammad, F. (2018). *Is preprocessing of text really worth your time for online comment classification?* <http://arxiv.org/abs/1806.02908>
- Mokhtari, M., Seraj, A., Saedi, N., & Karshenas, A. (2023). *The Impact of Twitter Sentiments on Stock Market Trends*. <http://arxiv.org/abs/2302.07244>
- Mokhtari, S., Yen, K. K., & Liu, J. (2021). Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning. *International Journal of Computer Applications*, 183(7), 1–8. <https://doi.org/10.5120/ijca2021921347>
- Mosbach, M., Andriushchenko, M., & Klakow, D. (2021). *On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines*. <http://arxiv.org/abs/2006.04884>
- Nanli, Z., Ping, Z., Weiguo, L., & Meng, C. (2012). Sentiment analysis: A literature review. *2012 International Symposium on Management of Technology (ISMOT)*, 572–576. <https://doi.org/10.1109/ISMOT.2012.6679538>
- Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. *Lecture Notes in Computer Science*

- (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*), 8206 LNCS, 194–201. https://doi.org/10.1007/978-3-642-41278-3_24
- Nguyen, D. Q., Vu, T., & Tuan Nguyen, A. (2020). BERTweet: A pre-trained language model for English Tweets. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 9–14. <https://doi.org/10.18653/v1/2020.emnlp-demos.2>
- Nielsen, F. (2011). A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. *Proceedings of the ESWC2011 Workshop on «Making Sense of Microposts»: Big things come in small packages*, 93–98.
- Niroomand, F., Metghalchi, M., & Hajilee, M. (2020). Efficient market hypothesis: a ruinous implication for Portuguese stock market. *Journal of Economics and Finance*, 44(4), 749–763. <https://doi.org/10.1007/S12197-020-09514-8/TABLES/9>
- Noori, M., & Maktoubian, J. (2019). Business Improvement Approach Based on Sentiment Twitter Analysis: Case Study. *EAI Endorsed Transactions on Cloud Systems*, «5»(14), 158417. <https://doi.org/10.4108/EAI.13-7-2018.158417>
- Nousi, C., & Tjortjis, C. (2021, Setembro 24). A Methodology for Stock Movement Prediction Using Sentiment Analysis on Twitter and StockTwits Data. *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, 1–7. <https://doi.org/10.1109/SEEDA-CECNSM53056.2021.9566242>
- Nunez, Y., Gibson, E. A., Tanner, E. M., Gennings, C., Coull, B. A., Goldsmith, J., & Kioumourtoglou, M. A. (2021). Reflection on modern methods: good practices for applied statistical learning in epidemiology. *International Journal of Epidemiology*, 50(2), 685–693. <https://doi.org/10.1093/IJE/DYAA259>
- Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8154 LNAI, 355–365. https://doi.org/10.1007/978-3-642-40669-0_31

- Pagliaro, A. (2025). Artificial Intelligence vs. Efficient Markets: A Critical Reassessment of Predictive Models in the Big Data Era. *Electronics*, 14(9), 1721. <https://doi.org/10.3390/electronics14091721>
- Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 1345–1350. <https://doi.org/10.1109/SCOPES.2016.7955659>
- Pak, A., Ziyaden, A., Saparov, T., Akhmetov, I., & Gelbukh, A. (2024). Word Embeddings: A Comprehensive Survey. *Computación y Sistemas*, 28(4), 2005–2029. <https://doi.org/10.13053/cys-28-4-5225>
- Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Antioch Review*, 63(2), 271–278. <https://doi.org/10.3115/1218955.1218990>
- Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *ACL-05 - 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 115–124. <https://arxiv.org/abs/cs/0506075v1>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment Classification using Machine Learning Techniques*. 79–86. <https://doi.org/10.3115/1118693.1118704>
- Papaioannou, P., Russo, L., Papaioannou, G., & Siettos, C. (2013). Can social microblogging be used to forecast intraday exchange rates? *NETNOMICS: Economic Research and Electronic Networking*, 14(1–2), 47–68. <https://doi.org/10.1007/s11066-013-9079-3>
- Park, S., & Woo, J. (2019). Gender Classification Using Sentiment Analysis and Deep Learning in a Health Web Forum. *Applied Sciences*, 9(6), 1249. <https://doi.org/10.3390/app9061249>

- Pashaei, E., & Pashaei, E. (2021). Training Feedforward Neural Network Using Enhanced Black Hole Algorithm: A Case Study on COVID-19 Related ACE2 Gene Expression Classification. *Arabian Journal for Science and Engineering*, 46(4), 3807. <https://doi.org/10.1007/S13369-020-05217-8>
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Pipalia, K., Bhadja, R., & Shukla, M. (2020). Comparative analysis of different transformer based architectures used in sentiment analysis. *Proceedings of the 2020 9th International Conference on System Modeling and Advancement in Research Trends, SMART 2020*, 411–415. <https://doi.org/10.1109/SMART50582.2020.9337081>
- Prusa, J., Khoshgoftaar, T. M., Dittman, D. J., & Napolitano, A. (2015). Using Random Undersampling to Alleviate Class Imbalance on Tweet Sentiment Data. *2015 IEEE International Conference on Information Reuse and Integration*, 197–202. <https://doi.org/10.1109/IRI.2015.39>
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10), 1872–1897. <https://doi.org/10.1007/s11431-020-1647-3>
- Radhakrishnan, A., Ruiz Luyten, M., Prasad, N., & Uhler, C. (2023). Transfer Learning with Kernel Methods. *Nature Communications 2023 14:1*, 14(1), 1–12. <https://doi.org/10.1038/s41467-023-41215-8>
- Radhiya, M., Adiwijaya, & Purbolaksono, M. D. (2022). Sentiment Analysis on Food Review from Website Zomato Using K-Nearest Neighbor (KNN) Algorithm. *2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)*, 250–255. <https://doi.org/10.1109/ICoSEIT55604.2022.10030024>
- Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y., & Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(1), 26. <https://doi.org/10.9781/IJIMAI.2016.415>

- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PLOS ONE*, *10*(9), e0138441. <https://doi.org/10.1371/journal.pone.0138441>
- Ray, P., & Chakrabarti, A. (2017). Twitter sentiment analysis for product review using lexicon method. *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, 211–216. <https://doi.org/10.1109/ICDMAI.2017.8073512>
- Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2015). SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1). <https://doi.org/10.1140/epjds/s13688-016-0085-1>
- Rizinski, M., Peshov, H., Mishev, K., Jovanovik, M., & Trajanov, D. (2024). Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex). *IEEE Access*, *12*, 7170–7198. <https://doi.org/10.1109/ACCESS.2024.3349970>
- Rodriguez Inserte, P., Nakhlé, M., Qader, R., Caillaut, G., & Liu, J. (2023). Large Language Model Adaptation for Financial Sentiment Analysis. *Proceedings of the Sixth Workshop on Financial Technology and Natural Language Processing*, 1–10. <https://doi.org/10.18653/v1/2023.finnlp-2.1>
- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, *8*, 842–866. https://doi.org/10.1162/tacl_a_00349
- Royal, J. (2025, Julho 31). *The world's largest stock exchanges: 10 biggest by market capitalization*. Bankrate. <https://www.bankrate.com/investing/worlds-largest-stock-exchanges-biggest-by-market-capitalization>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Saif, H., Fernandez, M., He, Y., & Alani, H. (2013). Evaluation Datasets for Twitter Sentiment Analysis A survey and a new dataset, the STS-Gold. *Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM 2013)*, 9–21.

- Saltelli, A., Puy, A., & Lo Piano, S. (2021). Sensitivity Analysis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3977108>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. <https://arxiv.org/abs/1910.01108v4>
- Sathyanarayanan, S. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 4023–4031. <https://doi.org/10.53555/AJBR.V27I4S.4345>
- Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Information Systems*, 27(2), 1–19. <https://doi.org/10.1145/1462198.1462204>
- Sebastiani, F., Baccianella, S., & Esuli, A. (2010). *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. <https://aclanthology.org/L10-1531/>
- Senin, P. (2008). Dynamic Time Warping Algorithm Review. Em *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*. <https://csdl.ics.hawaii.edu/techreports/2008/08-04/08-04.pdf>
- Sharma, H. D., & Goyal, P. (2023). An Analysis of Sentiment: Methods, Applications, and Challenges. *RAiSE-2023*, 59(1), 68. <https://doi.org/10.3390/engproc2023059068>
- Shen, Y., & Zhang, P. K. (2024). *Financial Sentiment Analysis on News and Reports Using Large Language Models and FinBERT*. <https://doi.org/https://doi.org/10.48550/arXiv.2410.01987>
- Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2025). A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. *Journal on Artificial Intelligence*, 6(1), 301–360. <https://doi.org/10.32604/jai.2024.054314>
- Shivaprasad, T. K., & Shetty, J. (2017). Sentiment analysis of product reviews: A review. *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 298–301. <https://doi.org/10.1109/ICICCT.2017.7975207>

- Siebers, P., Janiesch, C., & Zschech, P. (2022). A Survey of Text Representation Methods and Their Genealogy. *IEEE Access*, *10*, 96492–96513. <https://doi.org/10.1109/ACCESS.2022.3205719>
- Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, *121*, 102342. <https://doi.org/10.1016/j.is.2023.102342>
- Singh, A., & Jain, G. (2021). Sentiment Analysis of News Headlines Using Simple Transformers. *2021 Asian Conference on Innovation in Technology (ASIANCON)*, 1–6. <https://doi.org/10.1109/ASIANCON51346.2021.9544806>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sohangir, S., Petty, N., & Wang, D. (2018). Financial Sentiment Lexicon Analysis. *2018 IEEE 12th International Conference on Semantic Computing (ICSC), 2018-January*, 286–289. <https://doi.org/10.1109/ICSC.2018.00052>
- Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, *5*(1), 3. <https://doi.org/10.1186/s40537-017-0111-6>
- Sreenivas, G., Murthy, K. M., Prit Gopali, K., Eedula, N., & Mamatha, H. R. (2023). Sentiment Analysis of Hotel Reviews - A Comparative Study. *2023 IEEE 8th International Conference for Convergence in Technology, I2CT 2023*. <https://doi.org/10.1109/I2CT57861.2023.10126445>
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049
- Taj, S., Shaikh, B. B., & Fatemah Meghji, A. (2019). Sentiment Analysis of News Articles: A Lexicon based Approach. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1–5. <https://doi.org/10.1109/ICOMET.2019.8673428>

- Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences (Switzerland)*, 13(7). <https://doi.org/10.3390/APP13074550>
- Tao, Z., Xu, Q., Liu, X., & Liu, J. (2023). An integrated approach implementing sliding window and DTW distance for time series forecasting tasks. *Applied Intelligence*, 53(17), 20614–20625. <https://doi.org/10.1007/S10489-023-04590-9/TABLES/11>
- Tarimer, İ., Çoban, A., & Kocaman, A. E. (2019). *Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques*. <http://arxiv.org/abs/1903.11983>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2002-July*, 417–424. <https://arxiv.org/abs/cs/0212032v1>
- Tusar, M. T. H. K., & Islam, M. T. (2021). A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data. *Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021*. <https://doi.org/10.1109/ICECIT54077.2021.9641336>
- Vamossy, D. F. (2021). Investor emotions and earnings announcements. *Journal of Behavioral and Experimental Finance*, 30, 100474. <https://doi.org/10.1016/j.jbef.2021.100474>
- Vamossy, D. F. (2024). *Social Media Emotions and Market Behavior*. <https://doi.org/10.2139/ssrn.4806896>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems, 2017-December*, 5999–6009. <http://arxiv.org/abs/1706.03762>

- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis* (pp. 347–354). <https://aclanthology.org/H05-1044/>
- Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 1–5. <https://doi.org/10.1109/ICIC47613.2019.8985884>
- World Federation of Exchanges. (2025, Maio). *Market Statistics*. <https://focus.world-exchanges.org/issue/may-2025/market-statistics>
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, 50(1), 49–73. <https://doi.org/10.1007/s10462-017-9588-9>
- Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2(3), 249–262. <https://doi.org/10.1007/s41664-018-0068-2>
- Yin, W., Kann, K., Yu, M., SchützeSch, H., & Munich, L. (2017). *Comparative Study of CNN and RNN for Natural Language Processing*. <https://arxiv.org/abs/1702.01923v1>
- Yu, L., & Zhou, N. (2021). *Survey of Imbalanced Data Methodologies*. <http://arxiv.org/abs/2104.02240>
- Zhan, Z. (2025). Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews. *ITM Web of Conferences*, 70, 02013. <https://doi.org/10.1051/itmconf/20257002013>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis : A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), 1–25. <http://arxiv.org/abs/1801.07883>
- Zhang, Y., Cucuringu, M., Shestopaloff, A. Y., & Zohren, S. (2023). *Dynamic Time Warping for Lead-Lag Relationships in Lagged Multi-Factor Models*. <http://arxiv.org/abs/2309.08800>

Zhixiang, Xu, Chen, M., Weinberger, K. Q., & Sha, F. (2013). *An alternative text representation to TF-IDF and Bag-of-Words*. <http://arxiv.org/abs/1301.6770>