



PROJETO E AVALIAÇÃO DE MÉTODOS DE ANÁLISE DE SENTIMENTO PARA APOIO À DECISÃO DE COMPRA E VENDA DE ATIVOS

IGOR AMÂNCIO MACHADO DIAS

igor.dias@ga.ita.br

TRABALHO DE GRADUAÇÃO 2

INSTITUTO TECNOLÓGICO DE AERONÁUTICA

07 DE NOVEMBRO DE 2022

ORIENTADOR

CARLOS HENRIQUE QUARTUCCI FORSTER

carlos.forster@gp.ita.br

SUMÁRIO

OBJETIVO _____

MOTIVAÇÃO _____

PROCESSAMENTO DE
LINGUAGEM NATURAL _____

METODOLOGIA _____

RESULTADOS &
DISCUSSÃO _____

CONCLUSÃO _____

OBJETIVO

MENSURAÇÃO DE SENTIMENTO

Construção e avaliação de modelos de análise de sentimento, usando métodos atualizadas de representação textual e modelos classificatórios;

MODELO ATUALIZADO DE ANÁLISE DE
SENTIMENTO

SENTIMENTO E O MERCADO

Analisar o efeito que o resultado de tais modelos podem contribuir em ganho marginal na predição de tendência de preços

O MERCADO É AFETADO PELO SENTIMENTO DE
PARTE DOS SEUS AGENTES?

MOTIVAÇÃO

Apresentação da motivação , com os últimos trabalhos do assunto e seus resultados

O QUE ME MOTIVOU?

EDWARD THORP

Modelos de Arbitragem Beat the Market: A Scientific Stock Market System (1967)

Posterior lançamento de Beat the Dealer: A Winning Strategy for the Game of Twenty-One (1966)

HIPÓTESE DA EFICIÊNCIA DE MERCADO

Hipóteses:

- 1) O valor de cada ação reflete toda a informação disponível
- 2) Impossível superar o mercado (gerar α) consistente

ANÁLISE DE FATORES

O que justificaria um desempenho superior ?

Busca de fatores que edifiquem o mercado

- 1) ANÁLISE DE 3 FATORES – FAMA/FRENCH 1993
- 2) ANÁLISE DE 5 FATORES – FAMA/FRENCH 2015

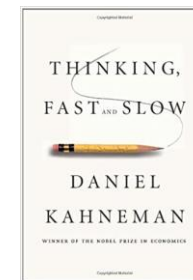
FINANÇAS COMPORTAMENTAIS

O sentimento dos agentes de mercado podem afetar a racionalidade do mercado?

Aversão a perda, viés local, viés de confirmação, etc

Grandes pesquisadores: DANIEL KAHNEMAN, RICHARD THALER e AMOS TVORZKY

Filme “Quebrando a Banca”
– inspirado nos estudos de Edward Thorp



“Pensando, Rápido e Devagar” – *best-seller* de psicologia comportamental

TRABALHOS RELACIONADOS

APRENDIZADO DE MÁQUINA

Predicting stock market behavior using data mining technique and news sentiment analysis
(KHEDR et al., 2017)

- Sentimento c/ Naive-Bayes: 72,73% - 86,21%
- Histórico de preços + Sentimento: 89,80%

	Prediction accuracy		
	K-NN	SVM	Naive Bayes
yahoo inc	75.86%	58.62%	86.21%
Msft inc	69.70%	66.67%	72.73%
Fb inc	72.41%	68.97%	82.76%

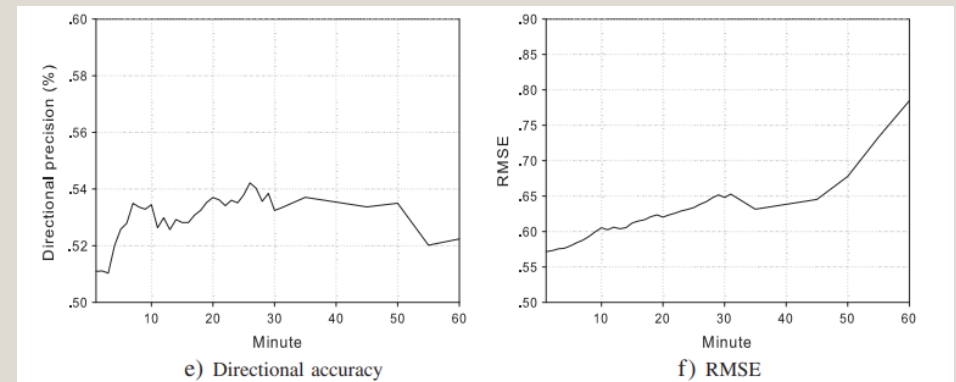
(KHEDR et al., 2017)

Resultado do teste na classificação

Media-Aware Quantitative Trader (MAQT) based on public web information

(LI et al., 2014)

- Uso de dicionários léxicos específico de finanças Loughran and MacDonal
- Maximização de predição por volta de 20 – 30 minutos



(LI et al., 2014)

Avaliação da janela de previsão

TRABALHOS RELACIONADOS

REDES NEURAIS PROFUNDAS

Decision support from financial disclosures with deep neural networks and transfer learning (KRAUS; FEUERRIEGEL, 2017)

- Deep Learning como melhor opção de predição de tendência e seu tamanho

(KRAUS; FEUERRIEGEL, 2017)
Resultado do teste na classificação de retornos nominais

Method	Training set	Test set			Absolute improvement on test set over baseline		
	Accuracy	Accuracy	Balanced accuracy	AUC	Accuracy	Balanced accuracy	AUC
NAÏVE BASELINE							
Majority class	0.549	0.540	0.500	0.500	–	–	–
TRADITIONAL MACHINE LEARNING							
Ridge regression	0.534	0.534	0.528	0.539	–0.006	0.028	0.039
Lasso	0.549	0.540	0.500	0.500	0.000	0.000	0.000
Elastic net	0.549	0.540	0.500	0.500	0.000	0.000	0.000
Random forest	0.557	0.562	0.547	0.552	0.022	0.047	0.052
SVM	0.552	0.545	0.522	0.556	0.005	0.022	0.056
AdaBoost	0.555	0.552	0.538	0.555	0.012	0.038	0.055
Gradient boosting	0.553	0.554	0.532	0.556	0.014	0.032	0.056
DEEP LEARNING							
RNN	0.588	0.545	0.530	0.529	0.005	0.030	0.029
LSTM	0.601	0.577	0.562	0.563	0.037	0.062	0.063
LSTM with word embeddings	0.597	0.576	0.563	0.568	0.036	0.063	0.068
TRANSFER LEARNING							
RNN with pre-training	0.596	0.548	0.533	0.530	0.008	0.033	0.033
LSTM with pre-training	0.576	0.578	0.564	0.577	0.038	0.064	0.077
LSTM with pre-training and word embeddings	0.581	0.580	0.571	0.568	0.040	0.071	0.068

TRABALHOS RELACIONADOS

REDES NEURAIS PROFUNDAS

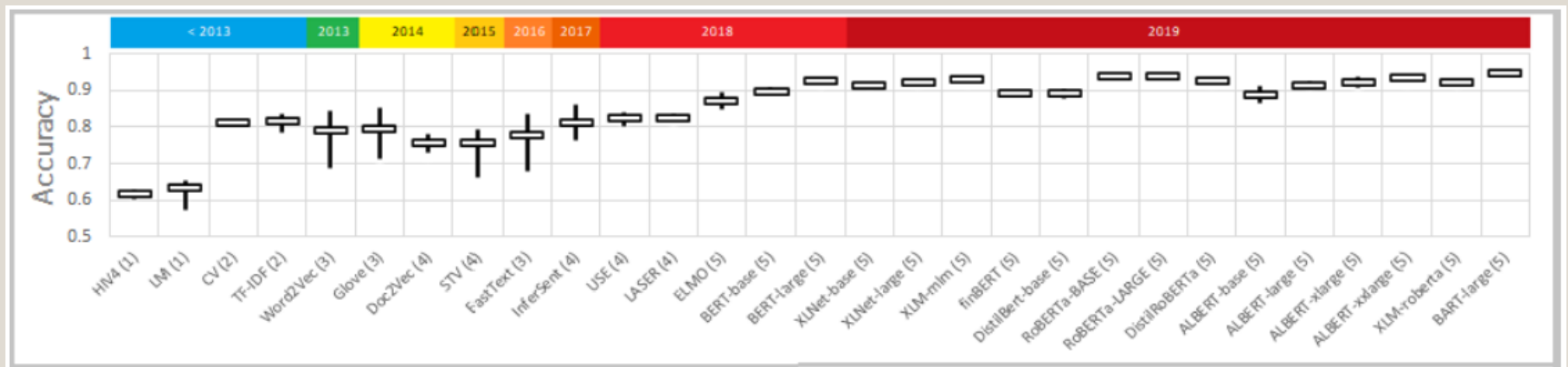
Evaluation of sentiment analysis in finance: From lexicons to transformers

(MISHEV et al., 2020)

- Focando na predição de sentimento
- Avaliando diversos modelos classificatórios, com transformers alcançando 90% de acurácia

(MISHEV et al., 2020)

Acurácia de cada modelo, em ordem cronológica de sua publicação



OUTROS TRABALHOS

APRENDIZADO DE MÁQUINA

SHYNKEVICH, Y.; MCGINNITY, T.; COLEMAN, S.; BELATRECHE, A. **Stock price prediction based on stock-specific and sub-industry-specific news articles**. In: 2015 International Joint Conference on Neural Networks (IJCNN). [S.l.: s.n.], 2015. p. 1–8

BING, L.; CHAN, K. C.; OU, C. **Public sentiment analysis in twitter data for prediction of a company's stock price movements**. In: 2014 IEEE 11th International Conference on e-Business Engineering. [S.l.: s.n.], 2014. p. 232–239.

RAJEEV PADMANAYANA, H. D. A. **Stockguru: smart way to predict stock price using machine learning**. International Journal of Innovative Research in Computer Science Technology (IJIRCST), v. 9, n. 4, p. 48–52, 2021. ISSN 2347 - 5552. Disponível [aqui](#)

SPRENGER, T. O.; TUMASJAN, A.; SANDNER, P. G.; WELPE, I. M. **Tweets and trades: the information content of stock microblogs**. European Financial Management, v. 20, n. 5, p. 926–957, 2014. Disponível [aqui](#)

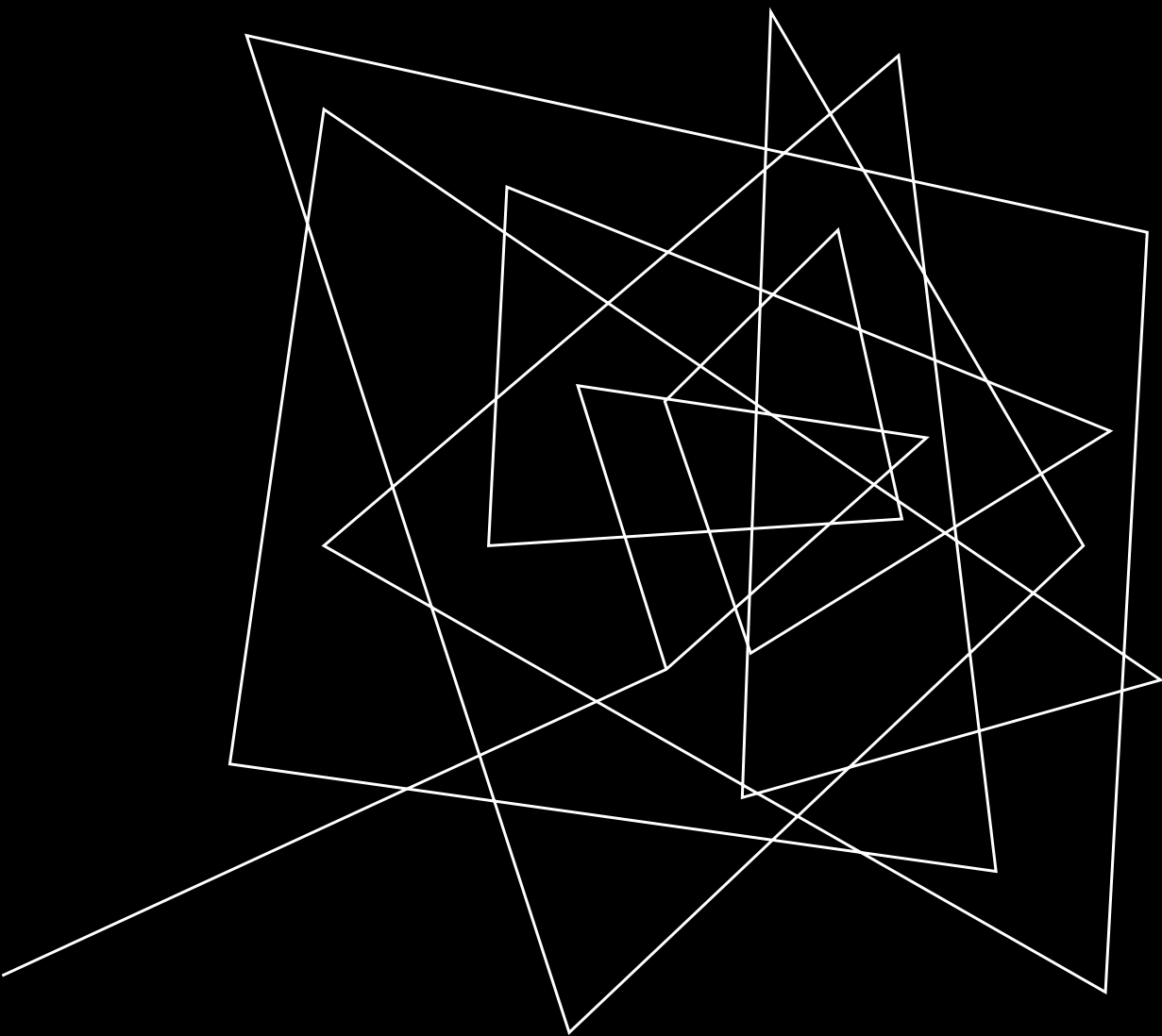
SANFORD, A. Does perception matter in asset pricing? modeling volatility jumps and returns using twitter-based sentiment indices. Journal of Behavioral Finance, 4 2019. Disponível [aqui](#)

REDES NEURAIS PROFUNDAS

SOWINSKA, K.; MADHYASTHA, P. **A Tweet-based Dataset for Company-Level Stock Return Prediction**. 2020.

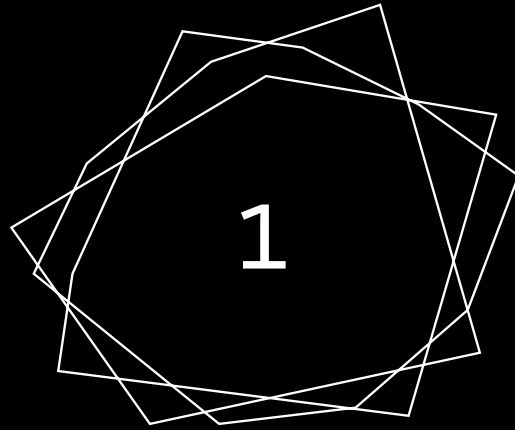
DING, X.; ZHANG, Y.; LIU, T.; DUAN, J. **Using structured events to predict stock price movement: An empirical investigation**. In: . [S.l.: s.n.], 2014. p. 1415–1425.

FEUERRIEGEL, S.; PRENDINGER, H. **News-based trading strategies**. Decision Support Systems, v. 90, p. 65–74, 2016. ISSN 0167-9236. Disponível [aqui](#)



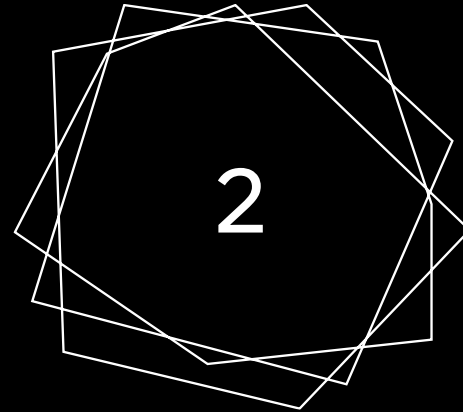
PROCESSAMENTO DE LINGUAGEM NATURAL

PIPELINE



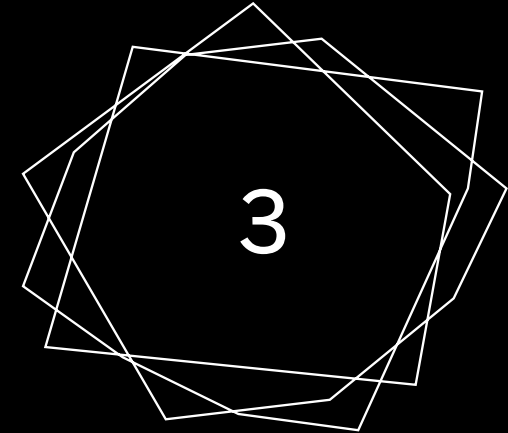
LIMPEZA DOS DADOS

Retirada de ruídos textuais



PRE-PROCESSAMENTO

Representação textual



MODELOS CLASSIFICATÓRIOS

Classificação de sentimento

LIMPEZA DE DADOS

PADRÃO

TOKENIZAÇÃO

Separação de partes do texto em unidades menores – tokens. Baseando-se em palavras, caracteres e subpalavras

REMOÇÃO DE STOP WORDS

Remoção de ruídos textuais, retirando preposições, pronomes, artigos etc que não trazem grandes ganhos informacionais

LEMMATIZATION

Padronização de palavras, levando em consideração o contexto das palavras. Exemplo: “Melhor” → “Bom”

TWITTER

Remoção de links internos e externos

Transformação de “emojis”

Remoção de referências à outros usuários



PRÉ-PROCESSAMENTO

ANÁLISE ESTATÍSTICA

TF-IDF

CODIFICADORES

GLOVE

TD-IDF

TF – TERM FREQUENCY

CONSIDERA O NÚMERO DE OCORRÊNCIA DO TERMO NA SEQUÊNCIA

IDF – INVERSE DOCUMENT FREQUENCY

PONDERAÇÃO SE O TERMO APARECE MAIS NO DOCUMENTO DO QUE NA SEQUÊNCIA

$$tfidf(t, s, S) = tf(t, s) * idf(t, S)$$

$$tf(t, s) = \log(1 + freq(t, s))$$

$$idf(t, S) = \log\left(\frac{N}{count(s \in S : t \in s)}\right)$$

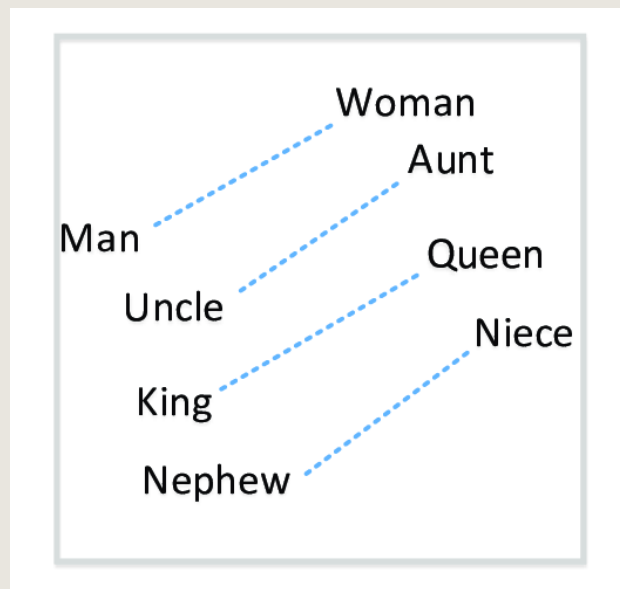
CODIFICADORES

GloVe

Parte da ideia do Word2Vec

Matriz global de fatorização + Skip Gram

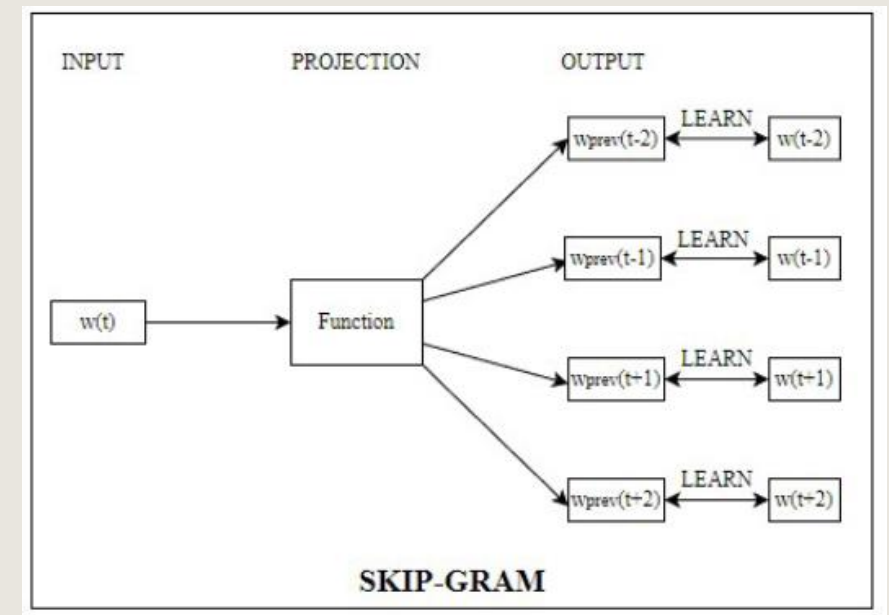
Disponível [aqui](#)



Resultado do GloVe

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k ice)/P(k steam)$	8.9	8.5×10^{-2}	1.36	0.96

Matriz de Ocorrência. Aqui o que importa a frequência relativa a outro [$P(solid, ice) > P(solid, steam)$]



Arquitetura do Skip-Gram. A partir de uma palavra, prever o contexto.



MODELOS CLASSIFICATÓRIOS

NAIVE BAYES

Aprendizado de Máquina tradicional

SUPPORT VECTOR CLASSIFIER

Aprendizado de Máquina tradicional

GRU

Redes Neurais Profundas

BERT

Transformers

NAYVE BAYES

Assume interindependência entre os atributos

A hipótese que possuir a maior probabilidade à posteriori é a hipótese (classe) escolhida

The diagram illustrates Bayes' Theorem with the following components:

- Likelihood of the Evidence given that the Hypothesis is True**: Labeled in yellow, with an arrow pointing to $P(E|H)$ in the numerator.
- Prior Probability of the Hypothesis**: Labeled in red, with an arrow pointing to $P(H)$ in the numerator.
- Posterior Probability of the Hypothesis given that the Evidence is True**: Labeled in blue, with an arrow pointing to $P(H|E)$ in the denominator.
- Prior Probability that the evidence is True**: Labeled in green, with an arrow pointing to $P(E)$ in the denominator.

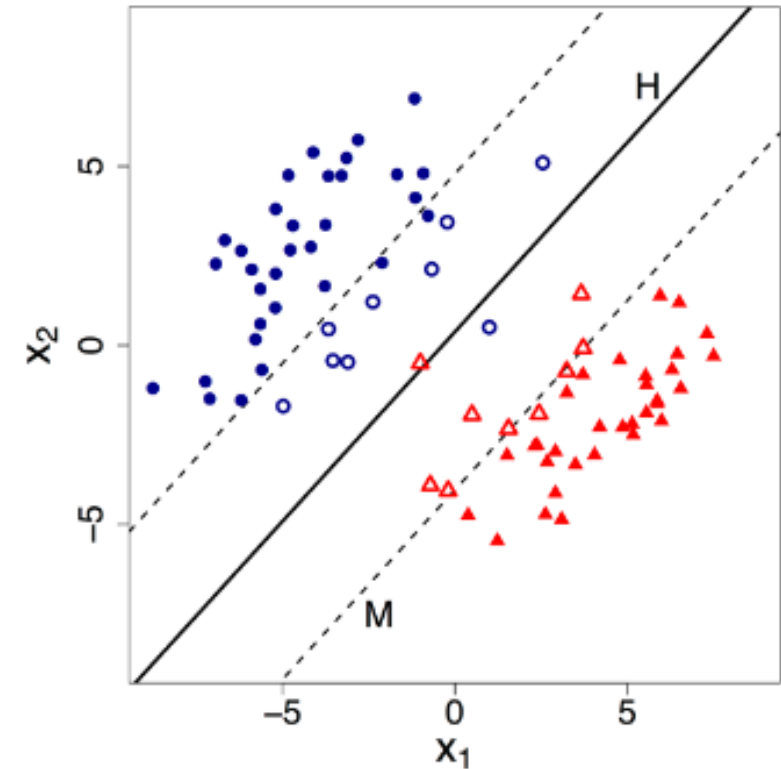
$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

SUPPORT VECTOR CLASSIFIER

Generalização da técnica de Maximal Margin Classifier

Permite-se uma certa permissividade na passagem de margem

$$\begin{aligned} & \underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_0, \epsilon_1, \dots, \epsilon_n, M}{max} \quad M \\ & \text{sujeito} \quad \sum_{j=1}^p \beta_j^2 = 1 \\ & \quad y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \\ & \quad \epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C \end{aligned}$$



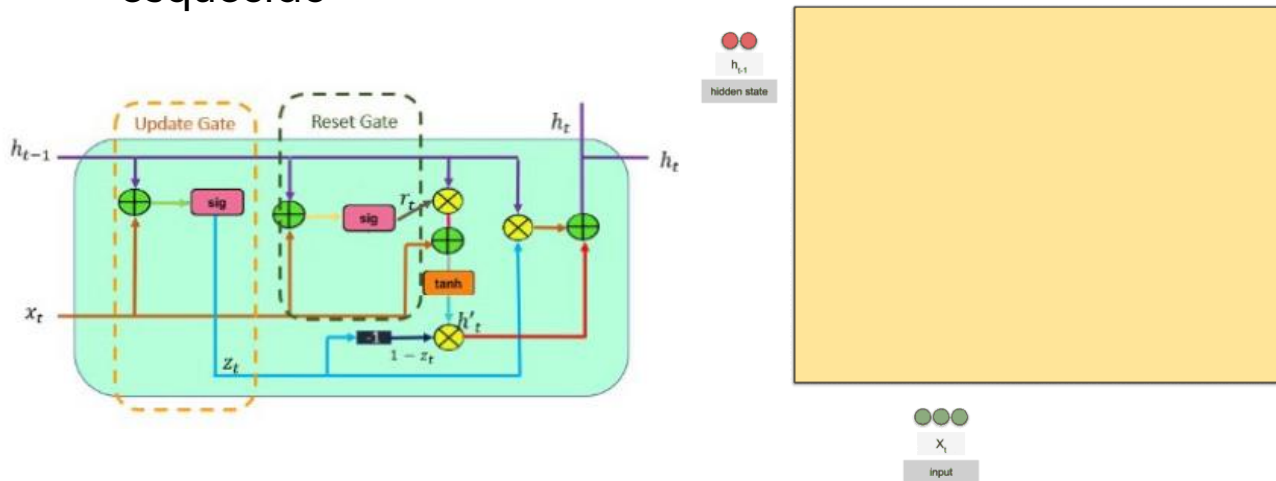
SVM com *soft margins*

GRU

Atacar o problema de RNN de perda de informação

Update Gate – quais informações do passado vão para frente

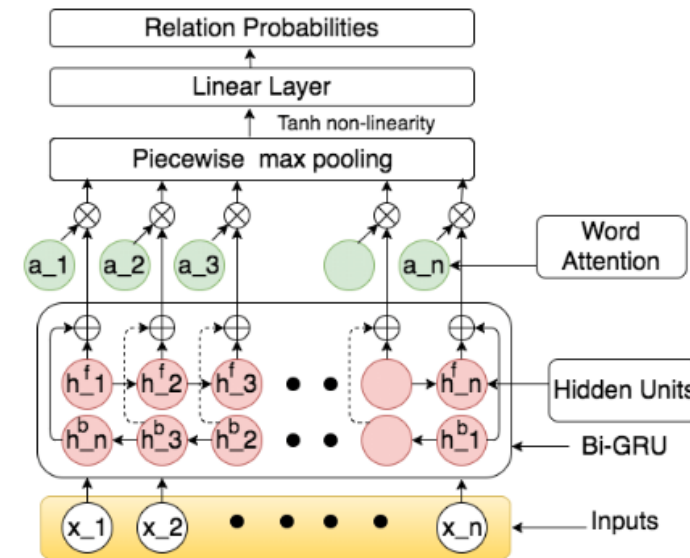
Reset Gate – o que do passado pode ser esquecido



Arquitetura unitária do GRU

GRU com:

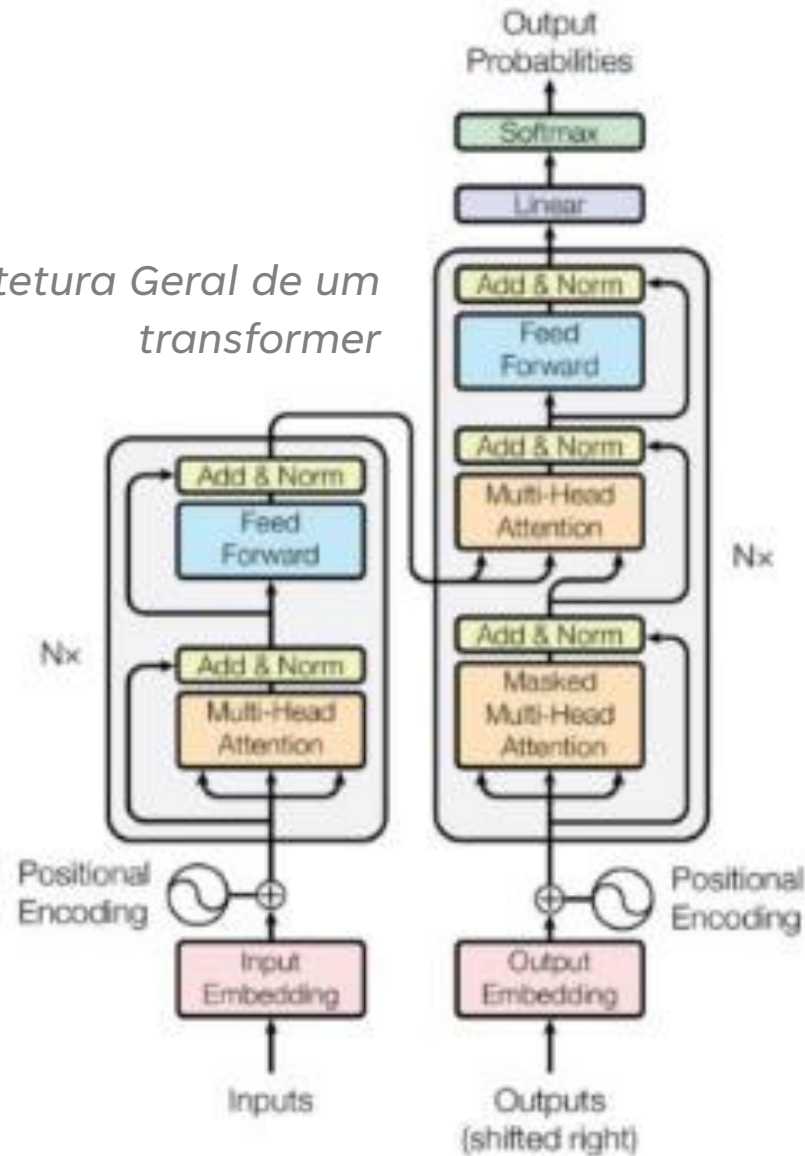
- Bidireção: coleta dados dos atributos nas duas direções
- Attention: foco em determinada entrada



GRU com bidireção e Attention

TRANSFORMER

Arquitetura Geral de um transformer



Arquitetura que transforma uma sequência em outra, por meio de codificador e decodificador sugerido em: “Attention is all you need” (VASWANI et al., 2017)

Em oposição a outros modelos direcionais, os Transformers leem a entrada de palavras completa de uma vez

Positional Encoding (por não usar RNN) + mecanismo focado no Attention, para conseguir aprender melhor relações de longa distância

Necessário um *fine-tuning* com aprendizado supervisionado para melhor adequação ao uso final

TRANSFORMER

MULTI-HEAD ATTENTION

Multi-Head Attention

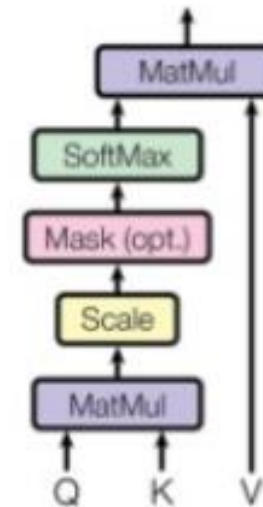
$$Attention(Q, K, V) = a \cdot V$$

$$a = softmax(\frac{Q \cdot K^T}{\sqrt{d_k}})$$

$$MultiHead(Q, K, V) = [head_1, head_2, \dots, head_h] \cdot W^0$$
$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

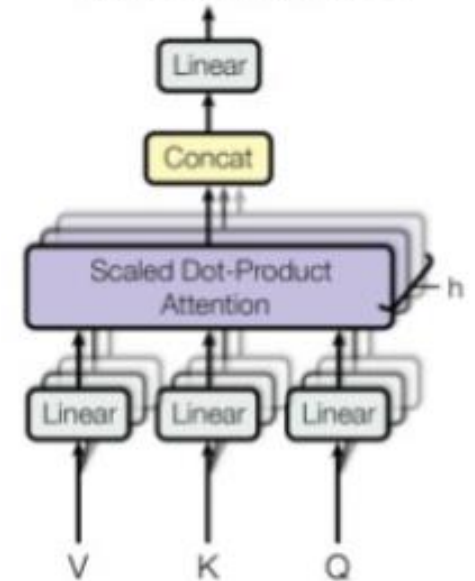
com Q representação vetorial da palavra e K todas as representações da sequência. Uso de pesos W para ajuste para cada head.

Scaled Dot-Product Attention



*Célula do
Attention*

Multi-Head Attention



*Paralelismo
do Attention*

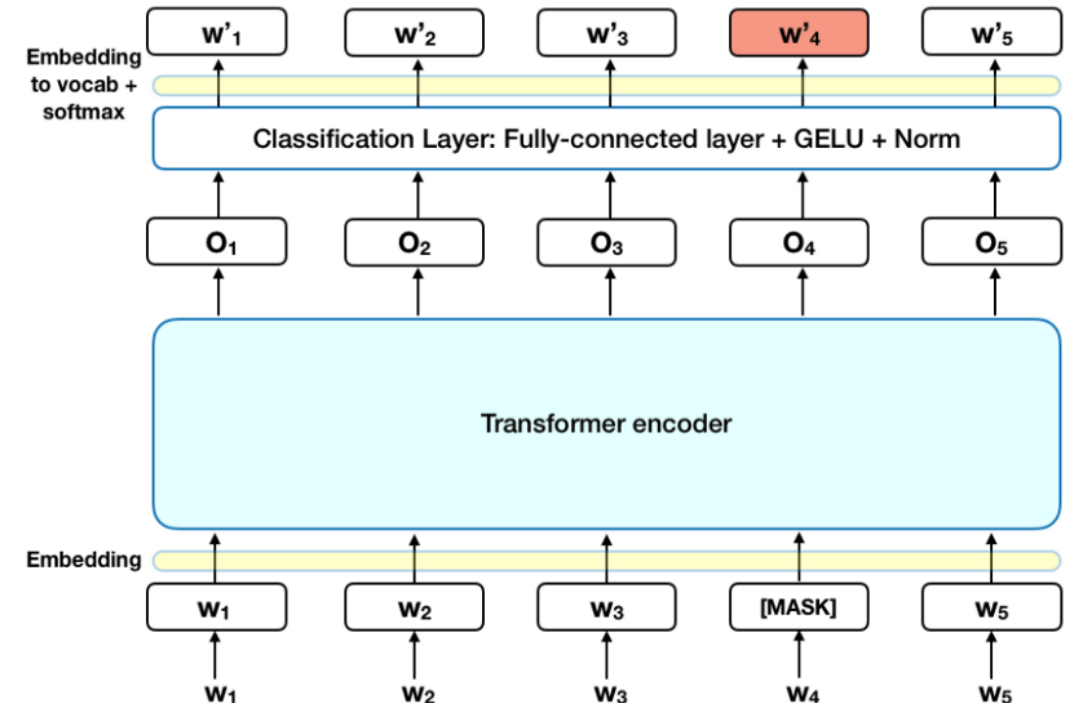
TRANSFORMER

BERT

Uso de estratégias não direcionais, para evitar limites de entendimento de contexto

Método Masked LM

- Uso da codificação do Transformer
- 15% das palavras substituídas pelo token [MASK]
- Busca prever a palavra escondida por meio das outras palavras

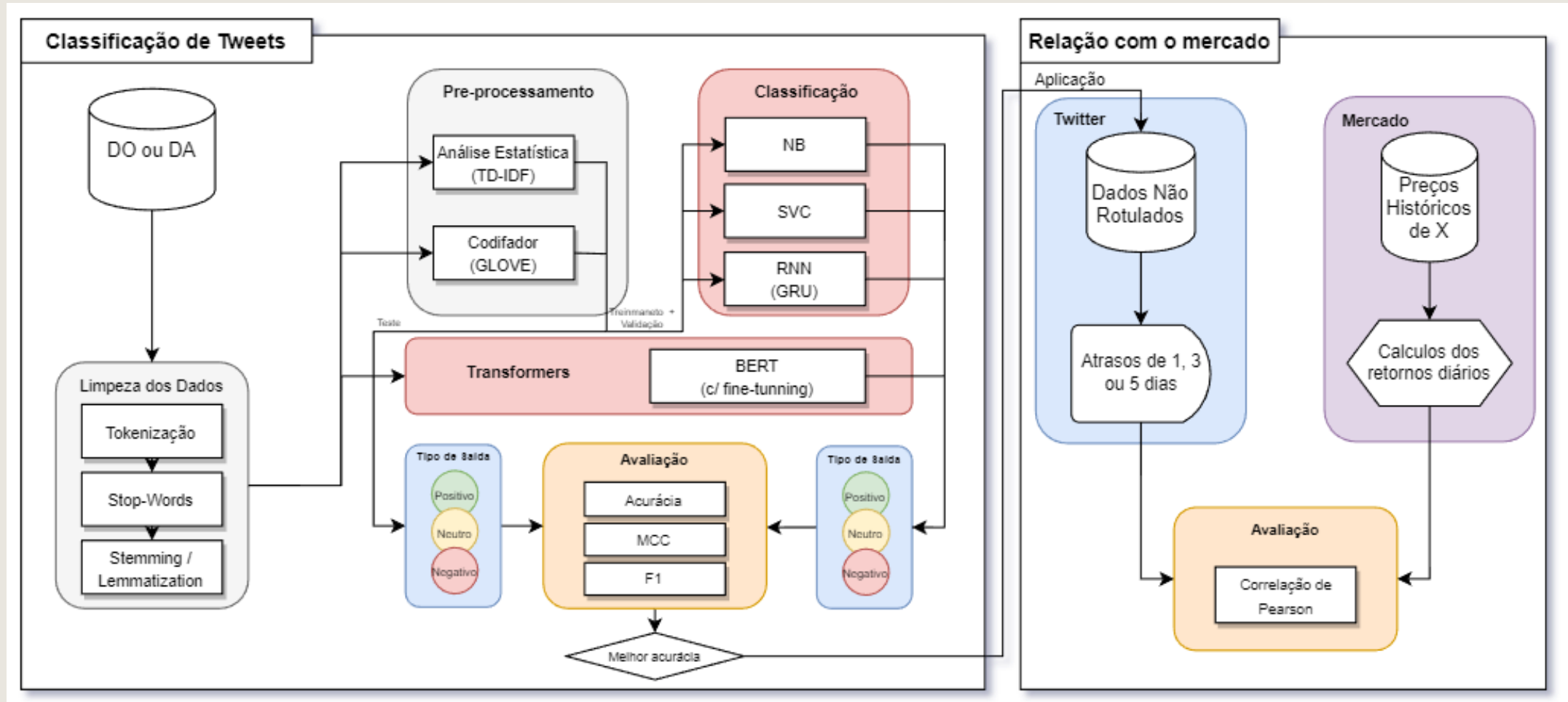


METODOLOGIA

Apresentação dos bancos de dados a serem usados, das métricas de avaliação consideradas e do *pipeline* esperado

PIPELINE

Será constituído em duas partes: análise de sentimento e relação com o mercado



BANCO DE DADOS

Stock Market Tweets Data fornecida pela
IEEEDataPort, fornecido [aqui](#)

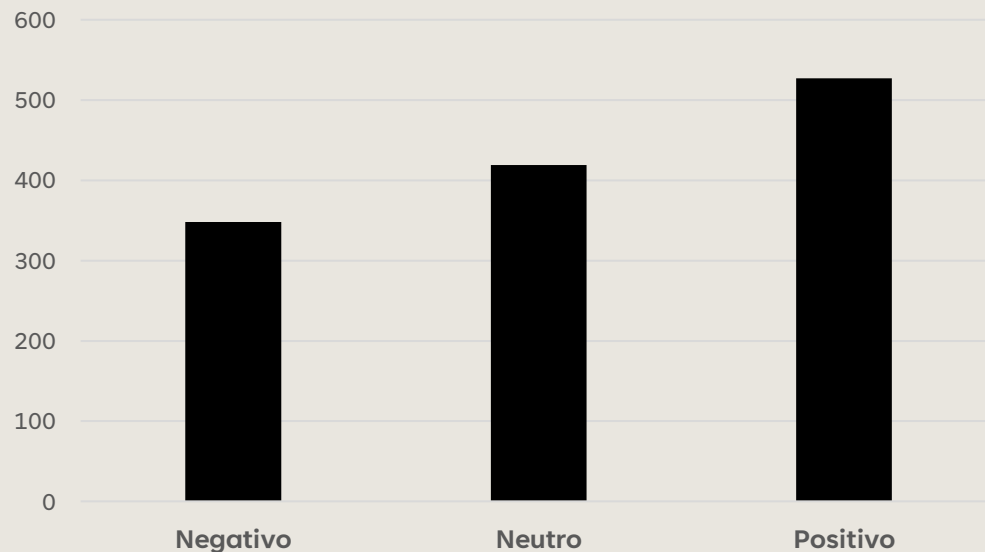
Para treinamento, tem-se mais de 938,672 tweets

Para avaliação, tem-se 1,300 tweets que foram
manualmente classificados e revistos por uma
segunda fonte independente. Para expansão de
treinamento, aplicação de *data augmentation*

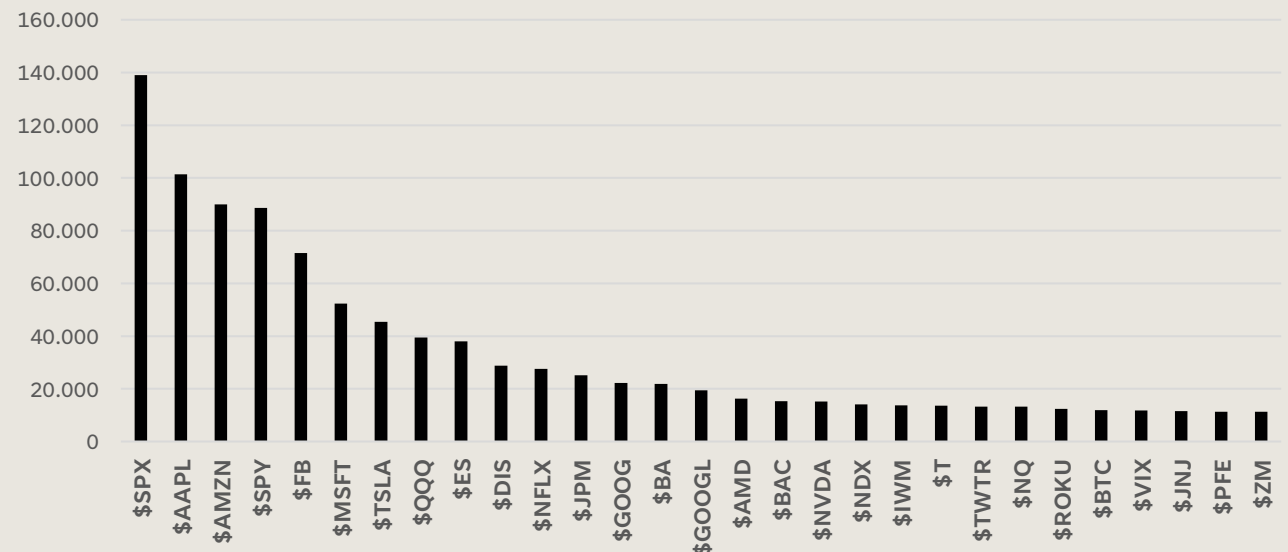
Período de 9 de Abril a 16 de Julho de 2020

Publicações históricas no Twitter

Classificações Textuais



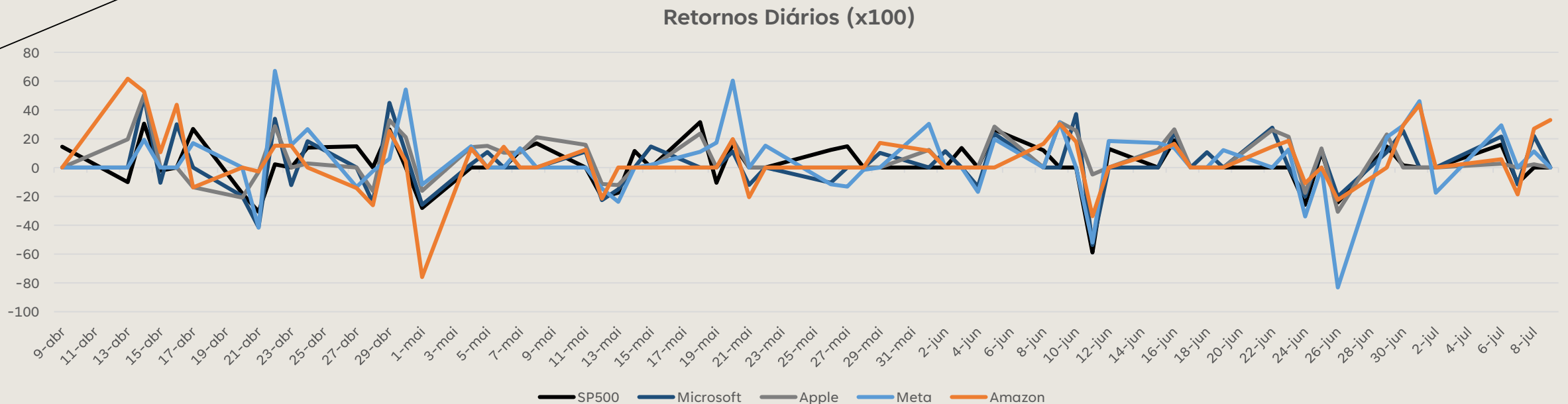
Frequência de Citação



BANCO DE DADOS

Histórico de Preços

Retirada do software Bloomberg Terminal
Dados diários para o mesmo período e para as ações: S&P500, MSFT, AMZN, META E AAPL
Cálculo dos retornos diários



MÉTRICAS

ANÁLISE DE SENTIMENTO

ACURÁCIA: acertos considerando todas as predições

F1: taxa de acerto de verdadeiro entre os verdadeiros (recall) e entre os que foram previstos (precisão)

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

MCC: Importante métrica para classificação binária, retornando valores entre -1 e 1

$$MCC = \frac{VP \cdot VN - FP \cdot FN}{\sqrt{(VP + FP) \cdot (FN + PN) \cdot (FP + VN) \cdot (VP + FN)}}$$

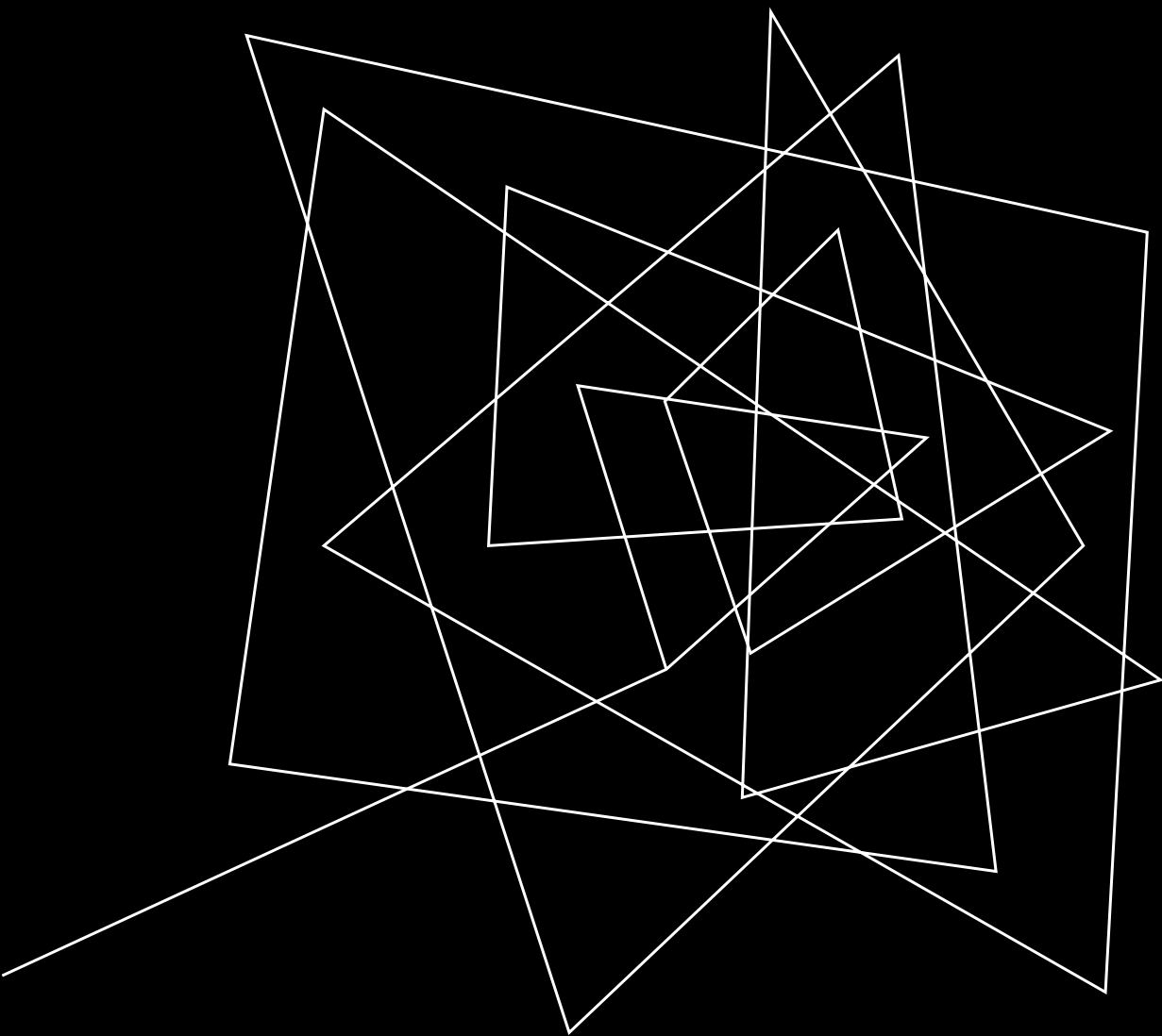
MOVIMENTOS DO MERCADO

CORRELAÇÃO DE PEARSON: medição de tendências lineares

$$\rho(X, Y) = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y}$$

RESULTADOS & DISCUSSÕES

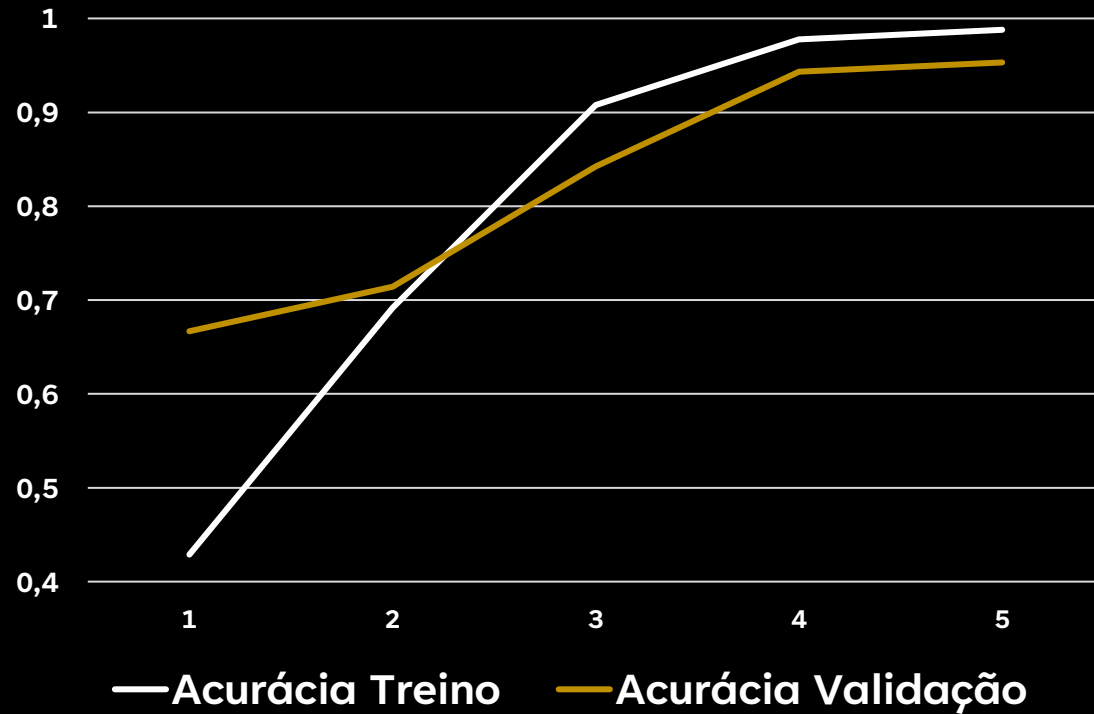
Apresentação dos resultados obtidos
pelos modelos classificadores e as
relações com os movimentos de
mercado



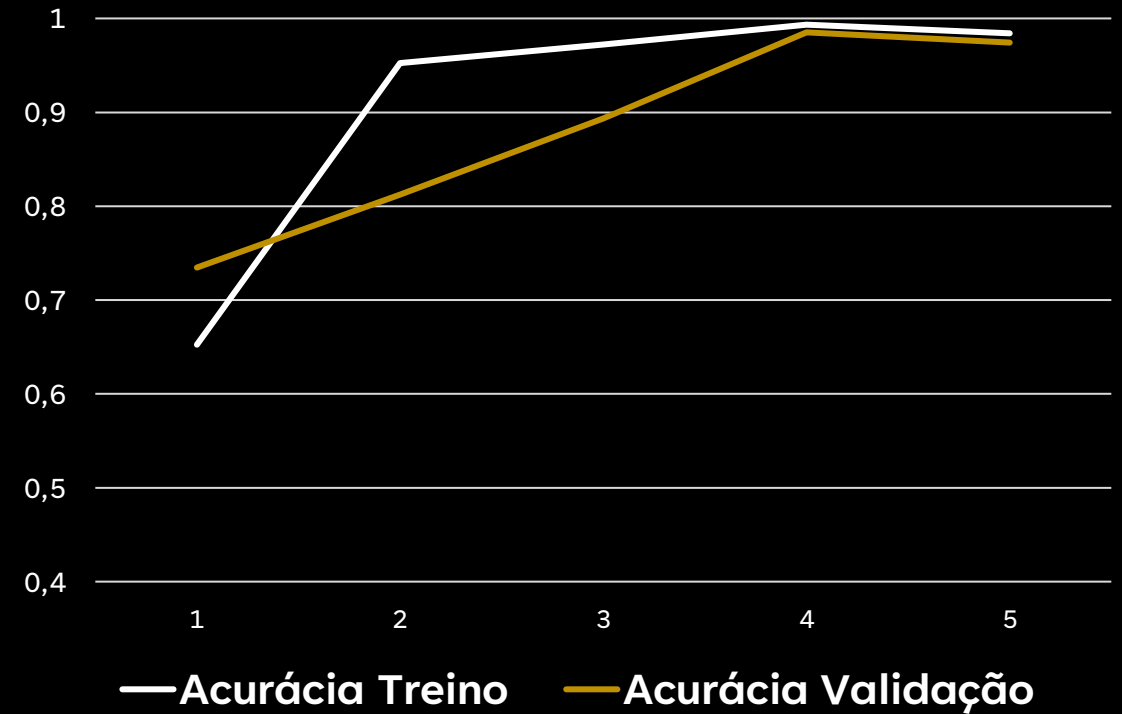
ANÁLISE DE SENTIMENTO

CURVAS DE APRENDIZADOS

BERT com dados de DO



BERT com dados de DA



APLICAÇÃO NO CONJUNTO TESTE

Modelos **mais complexos** (redes neurais e *transformers*) conseguem **alcançar melhores resultados**

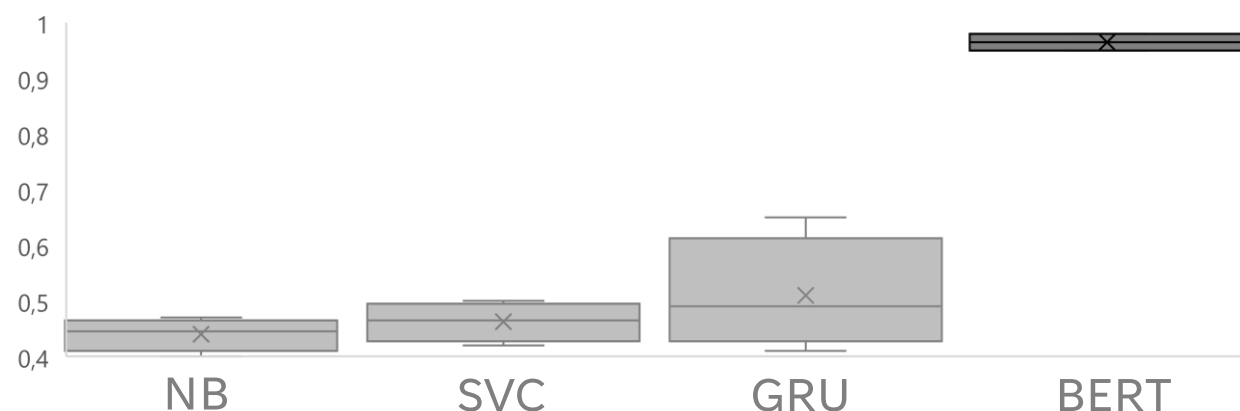
Técnica de ***augmentation*** para treinamento obteve um **melhora de 7% de acurácia na média**

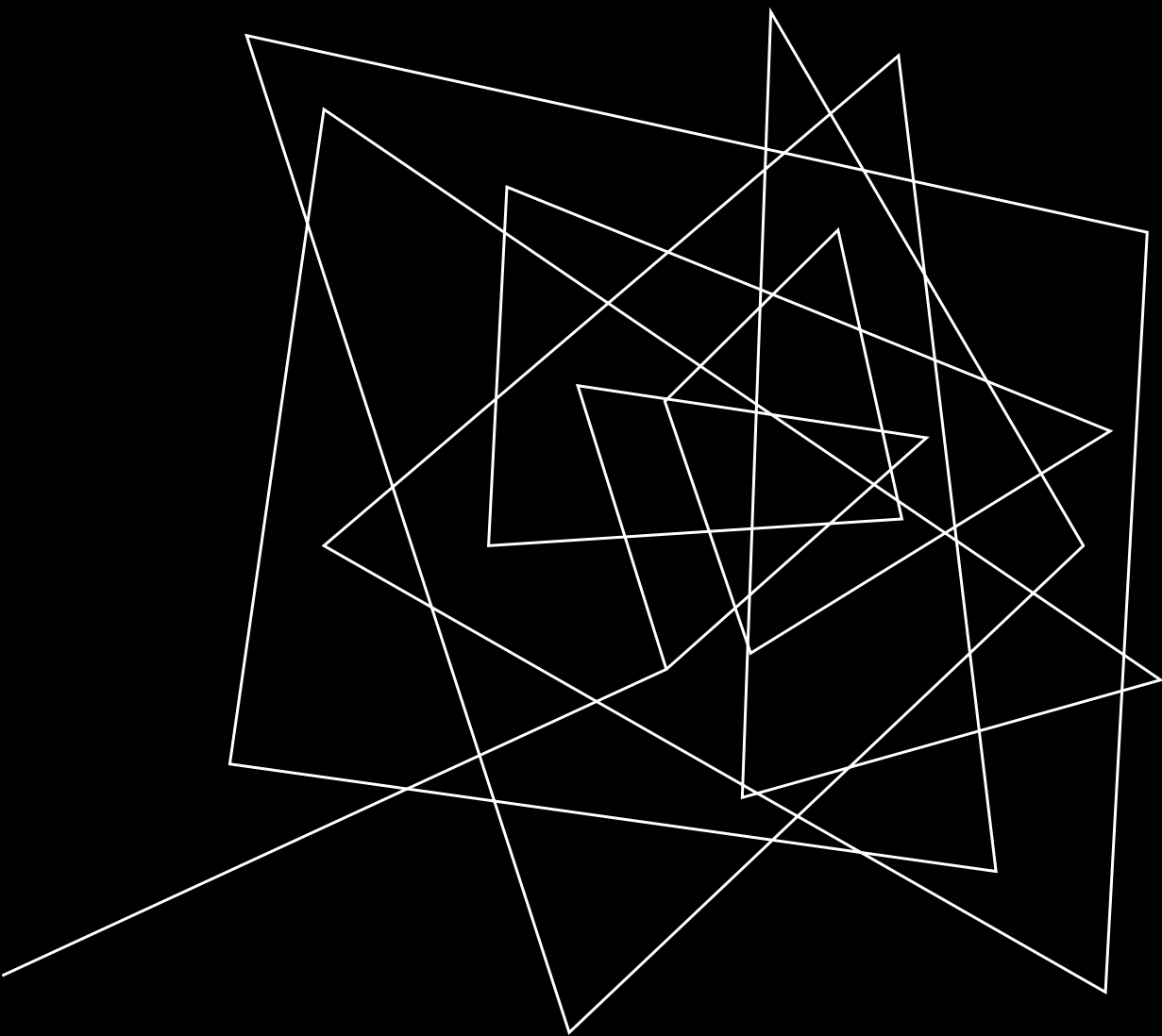
Aplicar **vetorização de tokens** mostrou ser **mais útil para arquiteturas de redes**, 0.02 tradicionais vs 0.11 r.n.

BERT alcançou resultados bem **expressivos e precisos**

TABELA 4.1 – Resultado de análise de sentimento dos modelos

Modelos	Dados Originais			Dados Aumentados		
	Acc	MCC	F1-Score	Acc	MCC	F1-Score
NB-TDIDF	0.40	0.16	0.39	0.47	0.18	0.47
SVC-TDIDF	0.42	0.19	0.39	0.48	0.20	0.41
GRU-TDIDF	0.41	0.13	0.38	0.50	0.21	0.48
NB-GLOVE	0.45	0.19	0.40	0.44	0.16	0.44
SVC-GLOVE	0.45	0.12	0.38	0.50	0.19	0.45
GRU-GLOVE	0.48	0.14	0.45	0.65	0.42	0.62
BERT	0.95	0.93	0.94	0.98	0.97	0.98

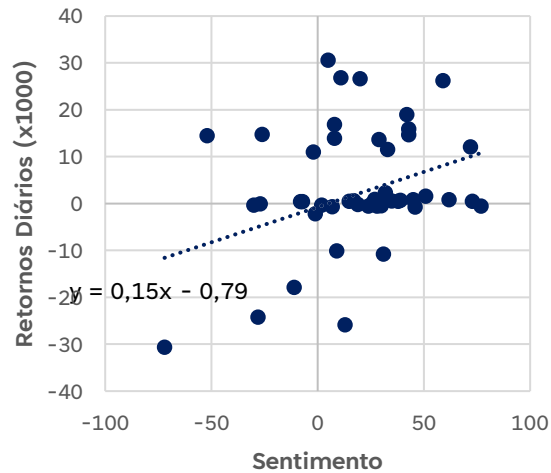




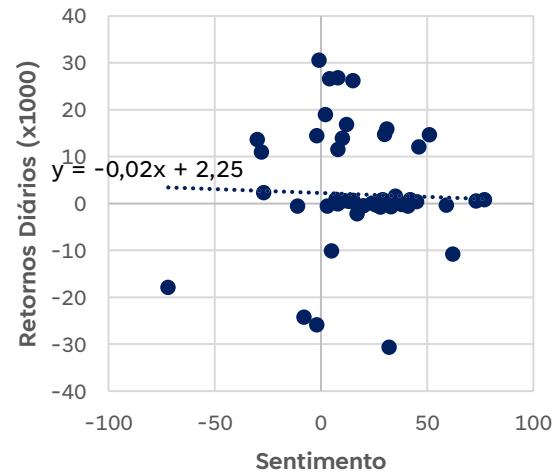
RELAÇÃO COM O
MERCADO

ANALISANDO O ÍNDICE S&P500

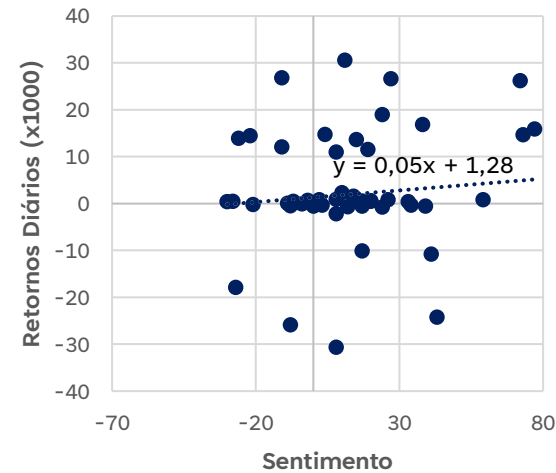
S&P500 x Sentimento dos
Tweets
Lag 0



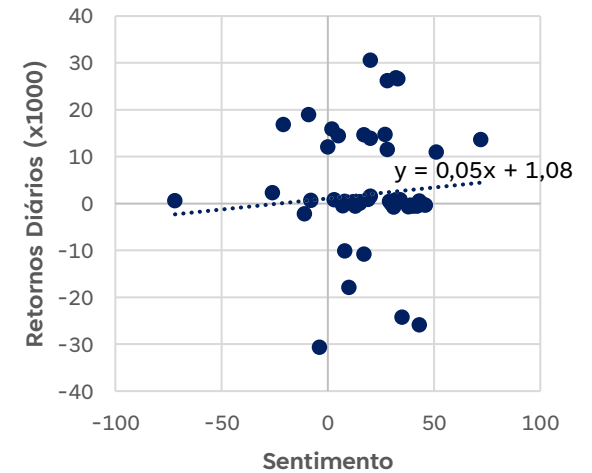
S&P500 x Sentimento dos
Tweets
Lag 1



S&P500 x Sentimento dos
Tweets
Lag 3



S&P500 x Sentimento dos
Tweets
Lag 5



**Por se tratar de um índice,
composição de várias ações, uma
menor relação seria esperado**

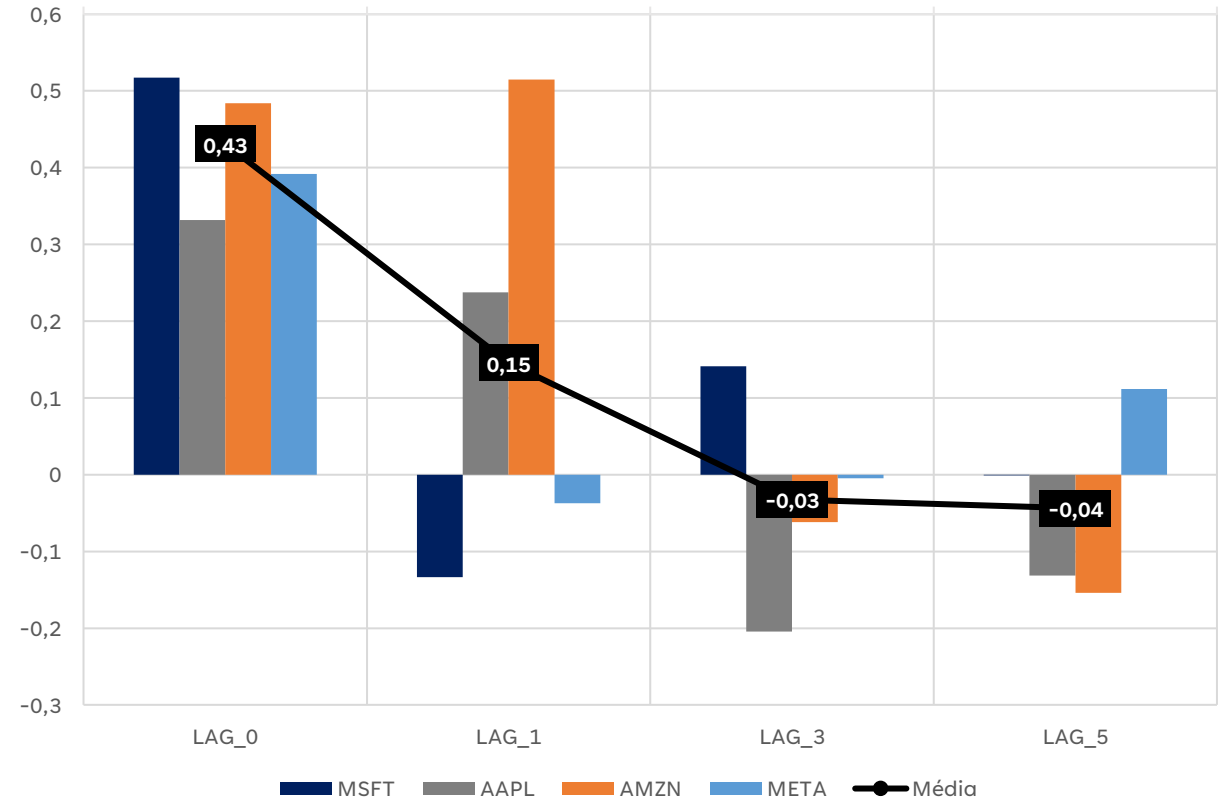
**Relação de sentimento e retornos
caindo com o aumento do número de
atrasos para 5 dias**

MAIORES COMPANHIAS

Relação caindo com o aumento de atrasos, tornando-se imperceptíveis após 3 dias

Possibilidade de uma **relação maior** (capacidade preditiva) para uma **frequência intradiária**

Correlação entre mercado e análise de sentimento



CONCLUSÃO

Principais pontos que podem ser retirados do trabalho,
com indicação para os possíveis próximos passos

ONDE SE CHEGOU?

MENSURAÇÃO DE SENTIMENTO

Aplicação e análise de 7 modelos classificatórios diferentes

Comparação entre os modelos, alcançando resultados de acordo com a literatura

Escolha do modelo BERT, conforme esperado, para classificação de tweets

SENTIMENTO E O MERCADO

Análise da classificação de sentimento dos tweets com os retornos diários

Comportamento da curva conforme esperado

Avaliação por outras fontes de notícias e outros dados alternativos

Análise em contexto intradiário para estudo capacidade preditiva

Aplicação de derivações atualizadas dos transformers e para língua portuguesa

PRÓXIMOS PASSOS



MUITO OBRIGADO!

Igor Amâncio Machado Dias

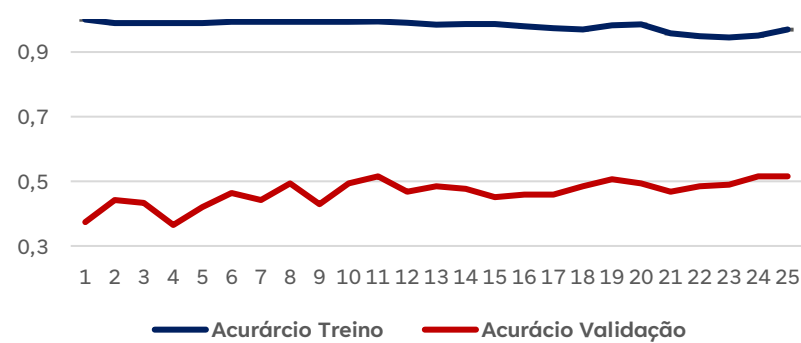
igor.dias@ga.ita.br



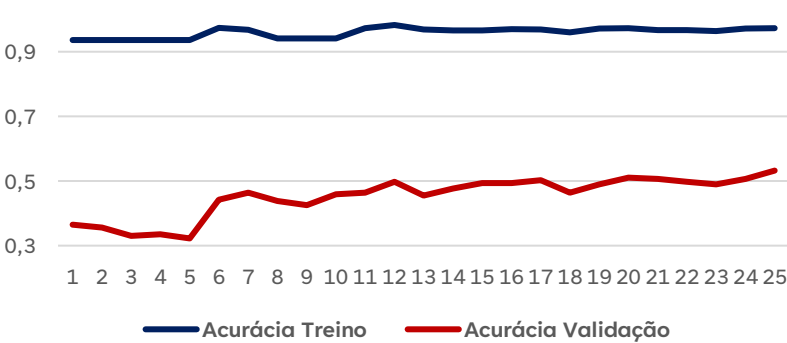
ANEXOS

CURVAS DE APRENDIZADO EM DO

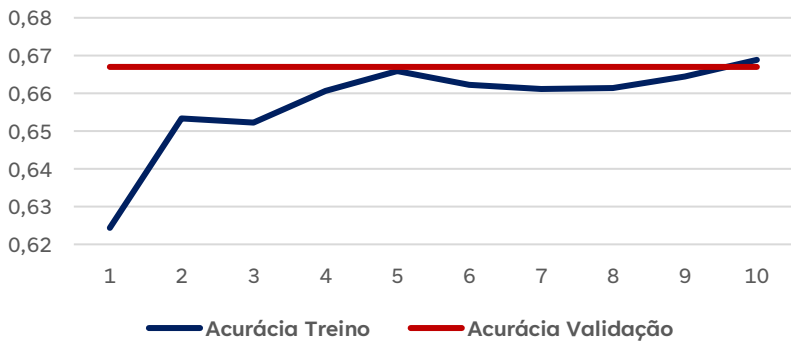
NB
TDIDF



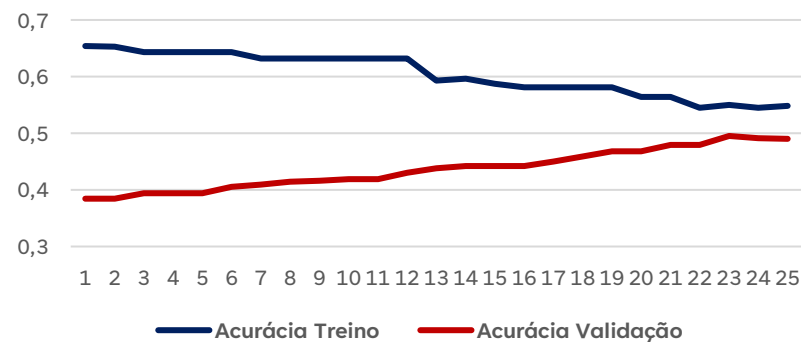
SVC
TDIDF



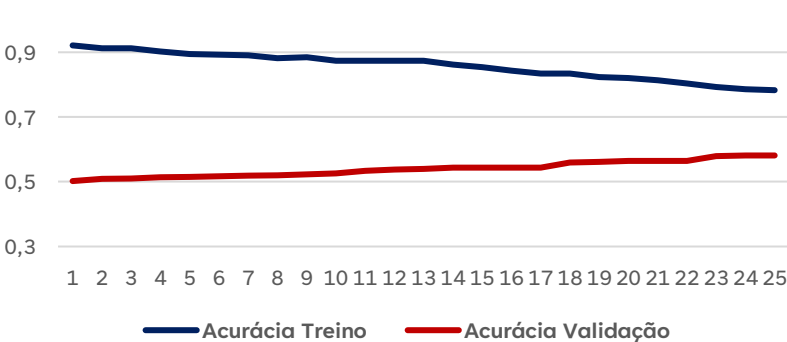
GRU
TDIDF



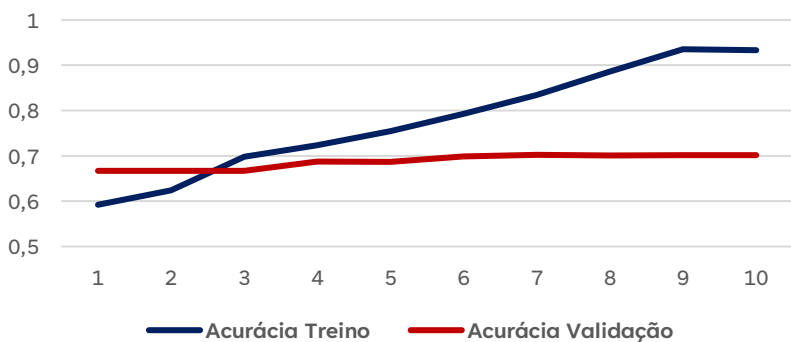
NB
GLOVE



SVC
GLOVE

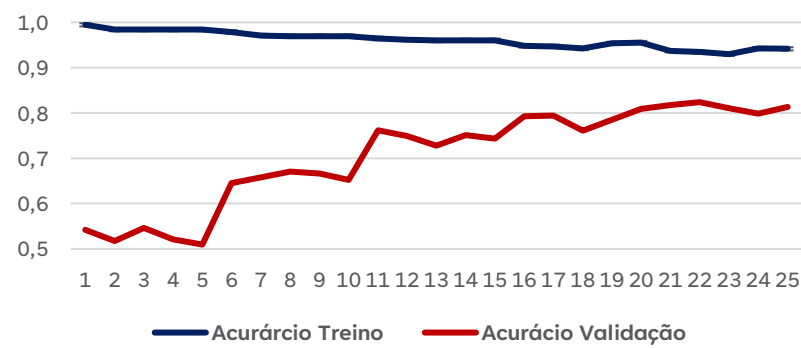


GRU
GLOVE

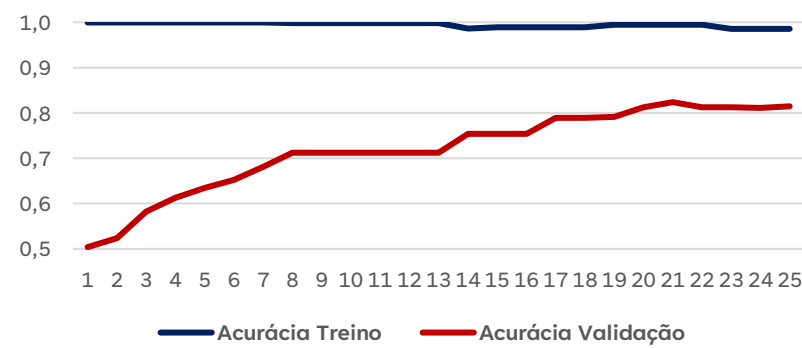


CURVAS DE APRENDIZADO EM DA

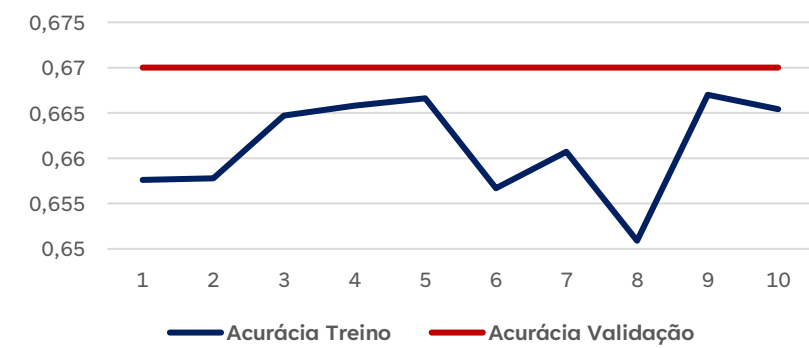
**NB
TDIDF**



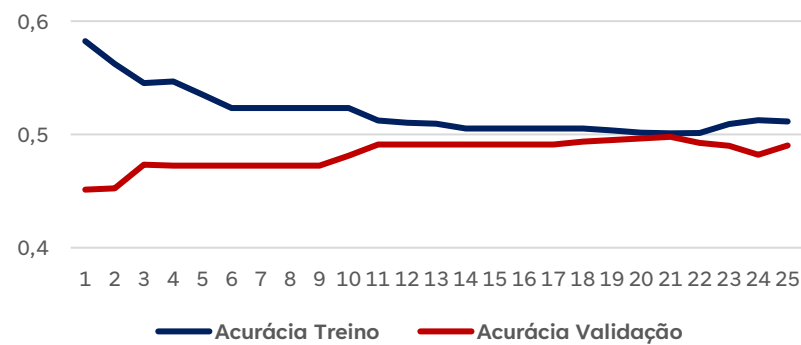
**SVC
TDIDF**



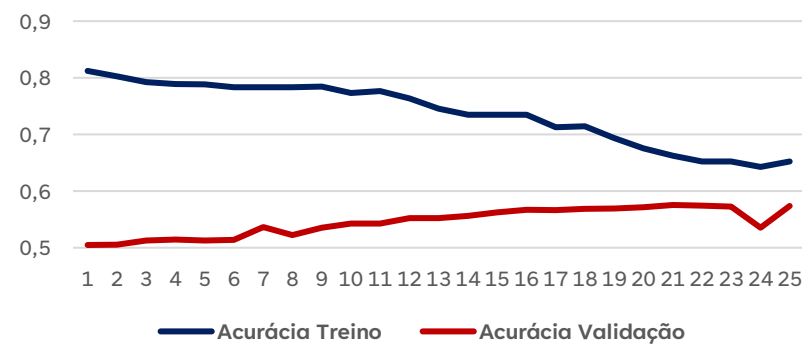
**GRU
TDIDF**



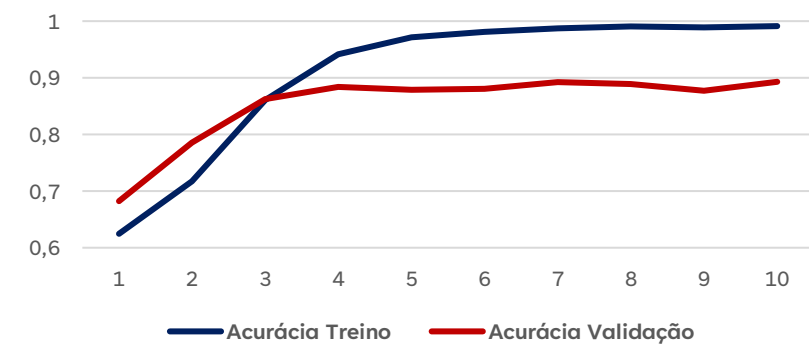
**NB
GLOVE**



**SVC
GLOVE**



**GRU
GLOVE**



REFERÊNCIAS

- BAHDANAU, D.; CHO, K.; BENGIO, Y. **Neural machine translation by jointly learning to align and translate**. In: BENGIO, Y.; LECUN, Y. (Ed.). 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. [s.n.], 2015.
- BAKER, M.; WURGLER, J. **Investor sentiment and the cross section of stock returns**. Journal of Finance, v. 61, n. 4, 2006. ISSN 1645-1680.
- BANKO, M.; CAFARELLA, M.; SODERLAND, S.; BROADHEAD, M.; ETZIONI, O. **Open information extraction from the web**. In: . [S.l.: s.n.], 2007. p. 2670–2676.
- BANZ, R. W. The relationship between return and market value of common stocks. Journal of Financial Economics, v. 9, n. 1, p. 3–18, 1981. ISSN 0304-405X.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. **Learning long-term dependencies with gradient descent is difficult**. IEEE Transactions on Neural Networks, v. 5, n. 2, p. 157–166, 1994.
- BING, L.; CHAN, K. C.; OU, C. **Public sentiment analysis in twitter data for prediction of a company's stock price movements**. In: 2014 IEEE 11th International Conference on e-Business Engineering. [S.l.: s.n.], 2014. p. 232–239.
- BLOOMBERG. Gráfico de histórico de preços para ações - GP. 2022.
- BONDT, W. F. M. D.; THALER, R. Does the stock market overreact? The Journal of Finance, v. 40, n. 3, p. 793–805, 1985.

REFERÊNCIAS

- BRAASCH, A. 82. generic dictionaries for multiple booktitle = Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography, publisher = De Gruyter Mouton, pages = 1186–1194. In: . [s.n.], 2013.
- CHUNG JUNYOUNG; GULCEHRE, C. C. K. e. B. Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In: NIPS 2014 Workshop on Deep Learning, December 2014. [S.l.: s.n.], 2014.
- CONNEAU, A.; KIELA, D.; SCHWENK, H.; BARRAULT, L.; BORDES, A. Supervised learning of universal sentence representations from natural language inference data. In: . [S.l.: s.n.], 2017. p. 670–680.
- CONNEAU, A.; LAMPLE, G. **Cross-lingual language model pretraining**. In: . Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2019.
- DEVLIN, J.; CHANG, M.; LEE, K.; TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805, 2018.
- DING, X.; ZHANG, Y.; LIU, T.; DUAN, J. Using structured events to predict stock price movement: An empirical investigation. In: . [S.l.: s.n.], 2014. p. 1415–1425.
- FACEBOOK, I. fastText: Library for fast text representation and classification. [S.l.], 2016.
- FAMA, E. F. The behavior of stock-market prices. The Journal of Business, University of Chicago Press, v. 38, n. 1, p. 34–105, 1965. ISSN 00219398, 15375374.

REFERÊNCIAS

FAMA, E. F.; FRENCH, K. R. **Common risk factors in the returns on stocks and bonds**. Journal of Financial Economics, v. 33, n. 1, p. 3–56, 1993. ISSN 0304-405X.

FAMA, E. F.; FRENCH, K. R. **The capm is wanted, dead or alive**. The Journal of Finance, [American Finance Association, Wiley], v. 51, n. 5, p. 1947–1958, 1996. ISSN 00221082, 15406261.

FAMA, E. F.; FRENCH, K. R. **A five-factor asset pricing model**. Journal of Financial Economics, v. 116, n. 1, p. 1–22, 2015. ISSN 0304-405X.

FEUERRIEGEL, S.; PRENDINGER, H. **News-based trading strategies**. Decision Support Systems, v. 90, p. 65–74, 2016. ISSN 0167-9236.

FRIEDMAN, J. H. **Greedy function approximation: A gradient boosting machine**. The Annals of Statistics, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189 – 1232, 2001.

GOODFELLOW I.; BENGIO, Y. e. C. A. **Deep Learning**. [S.l.]: MIT Press, 2017. ISBN 9780262035613.

GROB-KLUBMANN, A.; EBNER, M.; KÖNIG, S. **Structure in the tweet haystack: Uncovering the link between text-based sentiment signals and financial markets**. 10 2015.

HARRIS, Z. S. **Distributional structure**. In: . Word, 1954. v. 10, p. 146–162.

KAHNEMAN, D. **Thinking, Fast and Slow**. 1st. ed. [S.l.]: Farrar Straus Giroux, 2013. ISBN 0374533555.

KAHNEMAN, D.; SUNSTEIN, C. **Noise: A Flaw in Human Judgment**. 1st. ed. [S.l.]: Little, Brown Spark, 2021. ISBN 0316451401.

REFERÊNCIAS

KALAMARA, E.; TURRELL, A.; REDL, C.; KAPETANIOS, G.; KAPADIA, S. **Making text count: economic forecasting using newspaper text.** Journal of Applied Econometrics, 2022.

KARALEVICIUS, V.; DEGRANDE, N.; WEERDT, J. D. **Using sentiment analysis to predict interday bitcoin price movements.** The Journal of Risk Finance, v. 19, n. 1, p. 93–105, 2018. ISSN 1526-5943.

KHEDR, A. E.; SALAMA, S. E.; YASEEN, N. **Predicting stock market behavior using data mining technique and news sentiment analysis.** I.J. Intelligent Systems and Applications, v. 9, n. 7, p. 22–30, 2017. ISSN 0304-405X.

KIM, Y. **Convolutional neural networks for sentence classification.** In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014. p. 1746–1751.

KINGMA, D.; BA, J. **Adam: A method for stochastic optimization.** International Conference on Learning Representations, 12 2014.

KRAUS, M.; FEUERRIEGEL, S. **Decision support from financial disclosures with deep neural networks and transfer learning.** Decis. Support Syst., 2017.

LEBARON, B.; ARTHUR, W.; PALMER, R. **Time series properties of an artificial stock market.** Journal of Economic Dynamics and Control, v. 23, n. 9, p. 1487–1516, 1999. ISSN 0165-1889.

REFERÊNCIAS

- LEWIS, M.; LIU, Y.; GOYAL, N.; GHAZVININEJAD, M.; MOHAMED, A.; LEVY, O.; STOYANOV, V.; ZETTLEMOYER, L. **BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension**. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020. p. 7871–7880.
- LI, B.; CHAN, K. C.; OU, C.; RUIFENG, S. **Discovering public sentiment in social media for predicting stock movement of publicly listed companies**. Information Systems, v. 69, p. 81–92, 2017. ISSN 0306-4379.
- LI, Q.; WANG, T.; GONG, Q.; CHEN, Y.; LIN, Z.; SONG, S. kwang. **Media-aware quantitative trading based on public web information**. Decision Support Systems, v. 61, p. 93–105, 2014. ISSN 0167-9236.
- LOUGHRAN, T.; MCDONALD, B. **When is a liability not a liability? textual analysis, dictionaries, and 10-ks**. The Journal of Finance, v. 66, n. 1, p. 35–65, 2011.
- MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. **Efficient estimation of word representations in vector space**. Proceedings of Workshop at ICLR, v. 2013, 01 2013.
- MILLER, G. A. **Wordnet: A lexical database for english**. Commun. ACM, Association for Computing Machinery, New York, NY, USA, v. 38, n. 11, p. 39–41, nov 1995. ISSN 0001-0782.
- MISHEV, K.; GJORGJEVIKJ, A.; VODENSKA, I.; CHITKUSHEV, L. T.; TRAJANOV, D. **Evaluation of sentiment analysis in finance: From lexicons to transformers**. IEEE Access, v. 8, p. 131662–131682, 2020.

REFERÊNCIAS

MPQA. OpinionFinder (version 1.0). 2005.

NIELSEN, F. **A new anew: Evaluation of a word list for sentiment analysis in microblogs.** CoRR, 03 2011.

PATTERSON, S. **The Quants: How a New Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed It.** 1st. ed. [S.l.]: Crown Business, 2011. ISBN 0307453383.

PENNINGTON, J.; SOCHER, R.; MANNING, C. D. **Glove: Global vectors for word representation.** In: Empirical Methods in Natural Language Processing (EMNLP). [s.n.], 2014. p. 1532–1543.

RAJEEV PADMANAYANA, H. D. **A stockguru: smart way to predict stock price using machine learning.** International Journal of Innovative Research in Computer Science Technology (IJIRCST), v. 9, n. 4, p. 48–52, 2021. ISSN 2347 - 5552.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning Representations by Back-propagating Errors.** Nature, v. 323, n. 6088, p. 533–536, 1986.

SANFORD, A. **Does perception matter in asset pricing? modeling volatility jumps and returns using twitter-based sentiment indices.** Journal of Behavioral Finance, 4 2019.

SCHULER, K. **VerbNet: A broad-coverage, comprehensive verb lexicon.** Tese (Doutorado) — University of Pennsylvania, 01 2005.

SCHUMAKER, R. P.; CHEN, H. **Textual analysis of stock market prediction using breaking financial news: The azfin text system.** ACM Trans. Inf. Syst., Association for Computing Machinery, New York, NY, USA, v. 27, n. 2, mar 2009. ISSN 1046-8188.

REFERÊNCIAS

- SHYNKEVICH, Y.; MCGINNITY, T.; COLEMAN, S.; BELATRECHE, A. **Stock price prediction based on stock-specific and sub-industry-specific news articles**. In: 2015 International Joint Conference on Neural Networks (IJCNN). [S.l.: s.n.], 2015. p. 1–8.
- SOWINSKA, K.; MADHYASTHA, P. **A Tweet-based Dataset for Company-Level Stock Return Prediction**. 2020.
- SPRENGER, T. O.; TUMASJAN, A.; SANDNER, P. G.; WELPE, I. M. **Tweets and trades: the information content of stock microblogs**. European Financial Management, v. 20, n. 5, p. 926–957, 2014.
- STONE, P.; DUNPHY, D.; SMITH, M.; OGILVIE, D. **The General Inquirer: A Computer Approach to Content Analysis**. [S.l.: s.n.], 1966.
- TABORDA, B.; ALMEIDA, A. de; DIAS, J. C.; BATISTA, F.; RIBEIRO, R. **Stock Market Tweets Data**. IEEE Dataport, 2021.
- THALER, R. H. **Misbehaving: The Making of Behavioral Economics**. 1st. ed. [S.l.]: W. W. Norton Company, 2016. ISBN 039335279X.
- THORP, E.; KASSOUF, S. **Beat the Market: A Scientific Stock Market System**. 1st. ed. [S.l.]: Random House, 1967. ISBN 0394424395.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. **Attention is all you need**. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017. v. 30.
- YANG, X.; MACDONALD, C.; OUNIS, I. **Using word embeddings in twitter election classification**. Information Retrieval Journal, v. 21, 2018. ISSN 1573-7659.
- ZHANG, X.; ZHANG, Y.; WANG, S.; YAO, Y.; FANG, B.; YU, P. S. **Improving stock market prediction via heterogeneous information fusion**. Knowledge-Based Systems, Elsevier BV, v. 143, p. 236–247, mar 2018.