

NEWS IMPACT: FERRAMENTA COMPUTACIONAL PARA PREVISÃO DOS IMPACTOS DE NOTÍCIAS E FAKE NEWS NO MERCADO DE AÇÕES BRASILEIRO

Emanuel Elias Ferreira¹; Marcos Roberto Ribeiro²;

¹ Bolsista (CNPq, FAPEMIG), Engenharia de Computação, IFMG - Campus Bambuí, contato@emanuelelias.eng.br

² Orientador, IFMG - Campus Bambuí, GPSisCom, marcos.ribeiro@ifmg.edu.br

RESUMO

As informações difundidas na mídia nem sempre possuem conteúdo confiável. Atualmente, o principal canal de informação dos brasileiros são as redes sociais. Com o advento da Web 2.0, o poder de disseminação das notícias foi ampliado, e, por consequência, a influência dos dados expostos também. Em uma sociedade, na qual, os cidadãos possuem fácil acesso à informação, a propagação de notícias falsas se apresenta um grande problema. Popularmente, conhecidas por *fake news*, estas notícias são veiculadas como se fossem verdade, embora não representem de fato, a realidade. O uso mal intencionado dessas informações, pode acarretar uma série de complicações em diversos setores da sociedade. As *fake news*, tem afetado drasticamente a sociedade contemporânea. Consideradas um dos maiores problemas do século XXI. Seus impactos, são de difícil mensuração. Esta dificuldade, ocorre porque este fenômeno, além de possuir impacto global, o mesmo atinge diversas áreas da sociedade ao mesmo tempo, dentre elas o mercado financeiro. Logo, é possível quantificar, esses impactos através dos indicadores do mercado de ações. Este presente projeto de pesquisa, se propõe à auxiliar na previsão dos efeitos desse fenômeno, a fim de contribuir para a compreensão do mesmo, e por consequência minimizar os prejuízos. Além de colaborar com trabalhos correlatos futuros. Visto que, é possível estabelecer uma relação, entre: o alcance, a repercussão de determinadas notícias, e seus respectivos impactos sociais e econômicos. Neste caso, utilizando o mercado financeiro do Brasil, como um indicador plausível, através da combinação de análises de mercado, técnicas de inteligência artificial, e a análise de sentimentos. Vislumbra-se que as principais contribuições com a conclusão do presente trabalho, sejam as seguintes, o desenvolvimento de ferramenta computacional capaz de prever os impactos de *fake news*, confrontando com os de notícias, no mercado de ações brasileiro (B3), e a concepção de bases de notícias, sintetizadas a partir da extração das notícias dos portais mais influentes no Brasil, desta forma, viabilizando diversos trabalhos futuros.

Palavras-chave: Análise de Sentimentos. Inteligência Artificial. Mercado Financeiro.

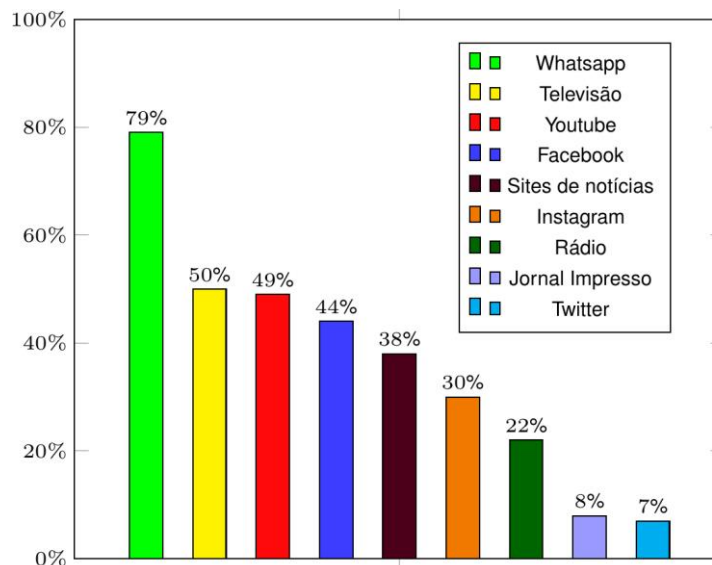
INTRODUÇÃO:

A partir da década de 1990, a Inteligência Artificial se desenvolveu bastante, em especial nas áreas de Processamento de Linguagem Natural, de Aprendizado de Máquina e de Representação de Conhecimento (RUIZ; FOGUEM; GRABOT, 2014; SUN; LUO; CHEN, 2017; GÜNDÜZ *et al.* 2019). Tal desenvolvimento tornou possível a solução de problemas consideravelmente complexos como a detecção de notícias falsas (SILVA, R. M. *et al.* 2020; MORENO; BRESSAN, 2019; MONTEIRO *et al.* 2018; SHU *et al.* 2017), e a predição sobre os impactos de redes sociais (CAROSIA; COELHO; SILVA, A. E. A., 2020; MEDEIROS; BORGES, 2019; MACHADO; PEREIRA, 2018; ARAÚJO, J. G. de; MARINHO, 2018).

As movimentações de impacto significativo no valor de mercado das empresas, dependem de notícias e eventos externos (MALKIEL, 2003; FAKHRY, 2016). Portanto, é interessante que o investidor tenha conhecimento a respeito da repercussão de notícias relacionadas as empresas e segmentos de mercado de seu interesse. Contudo, com o aumento do volume de notícias na Internet pode dificultar o trabalho do investidor na tarefa de analisar as notícias e determinar como as mesmas afetam o mercado (CHAN, 2003; LI *et al.* 2014). Assim, é importante que essas pessoas tenham acesso a ferramentas capazes de auxiliar na análise das notícias e seus impactos no mercado financeiro.

Segundo DataSenado (2019), as notícias falsas recebem maior visibilidade em relação às verdadeiras. Ainda de acordo com a mesma pesquisa, o principal meio de acesso à informação dos brasileiros são as redes sociais, como pode-se observar na Figura 1. E por consequência disto, também são o maior veículo de informações falaciosas. Portanto, é interessante analisar como essas notícias falsas influenciam a sociedade. Outro fator que torna-se pertinente essa análise, ocorre em virtude do Brasil ser o país mais afetado por fake news de acordo com a pesquisa realizada pelo IPSOS (2018).

Figura 1 – Principais Fontes de Informação dos Brasileiros que possuem acesso à internet.



Fonte: Adaptado de DataSenado (2019).

Atualmente, os principais métodos de previsão do mercado financeiro, são baseados na combinação de técnicas de aprendizagem de máquina com as seguintes análises de mercado: fundamentalista, técnica e séries cronológicas (ATSALAKIS; VALAVANIS, 2009). Entretanto, uma das abordagens que demonstrou melhores resultados, no atual estado da arte, utiliza-se a combinação de diferentes técnicas de inteligência artificial, análises de mercado e a análise de sentimentos. O presente trabalho, realiza uma contribuição solucionando alguns desafios propostos em trabalhos correlatos (CAROSIA; COELHO; SILVA, 2019; CAROSIA; COELHO; SILVA, 2020).

O fenômeno de fake news tem afetado diversas áreas da sociedade, dentre elas o mercado de ações. Apesar da existência de trabalhos nesta área, como a previsão sobre as influências de mídias sociais e notícias (CAROSIA; COELHO; SILVA, 2020; MACHADO; PEREIRA, 2018; MEDEIROS; BORGES, 2019; ARAÚJO, J. G. de; MARINHO, 2018), ainda não foi analisado o impacto de fake news no mercado de ações brasileiro.

Levando em consideração a literatura internacional, existem diversos estudos sobre o tema, todavia estes estudos apenas analisam de forma separada a relação entre como as notícias reais ou como as falsas afetam o mercado, mas não os dois tipos ao mesmo tempo. Até onde foi pesquisado, não há estudos que analisem ou comparem simultaneamente como esses tipos de notícias influenciam o mercado financeiro.

Do ponto de vista científico, é interessante desenvolver e utilizar técnicas de processamento de linguagem natural e análise de sentimentos para analisar como as notícias falsas ou verdadeiras podem impactar o mercado de ações. Além disto, estas técnicas podem ser aplicadas em outras esferas importantes como política e segurança digital. Do ponto de vista social e econômico, uma previsão do impacto das notícias pode auxiliar os investidores no mercado de ações a evitar prejuízos que possam afetar a economia e a sociedade.

METODOLOGIA:

A etapa inicial do presente projeto, envolveu a identificação dos principais portais online de notícias do Brasil. Tal identificação foi realizada com base na popularidade e número de acessos dos portais mais influentes do Brasil (NEWMAN et al. 2019; DATASENADO, 2019).

Em seguida, foram utilizadas ferramentas de recuperação de informação para processar as informações das notícias dos portais. E também, do histórico dos índices das ações das empresas pertencentes ao mercado de ações brasileiro, desde os últimos 5 anos até a data de 28 de Abril de 2021, com relação ao estudo quanto os impactos das notícias, os dados utilizados do mercado de ações, foram referentes ao

intervalo de tempo em que as notícias foram coletadas. E por consequência, essas informações foram processadas de acordo com seus segmentos de atuação no mercado financeiro. Após a coleta, esses dados foram aglomerados para a construção das bases de dados.

A critério de estudo comparativo dos impactos de *fake news* e notícias, foram consideradas as notícias presentes na base de dados *fake.br* (MONTEIRO *et al.*, 2018), que incluem as notícias falsas, e as suas respectivas versões verdadeiras, coletadas desde 02 de Maio de 2015 a 30 de Novembro de 2017, visto que esta tarefa de verificação quanto a veracidade das informações contidas nessas notícias não é trivial, por este motivo, para este estudo foi utilizada essa base de *fake news* já corroborada em outros trabalhos científicos.

Após a construção da base de dados, foi realizada a limpeza dessas bases, e o pré-processamento destes dados, por exemplo a remoção de *stopwords*. Em seguida, foi efetuado o processo de análise de sentimento das informações dessas notícias. Como estas notícias estão correlacionadas com os segmentos do mercado financeiro. Os sentimentos extraídos das notícias também está associado a um ou mais segmentos.

De posse dos históricos de sentimentos, e dos históricos dos principais índices das ações, e suas respectivas variações diárias, foram utilizadas técnicas de aprendizado de máquina combinadas com a análise de sentimentos, para a criação de algoritmos capazes de prever o valor de fechamento das ações de determinadas empresas. Por fim, a base de dados juntamente com os algoritmos desenvolvidos serão integrados na forma de um site para facilitar a utilização da ferramenta.

Em relação a metodologia adotada para o desenvolvimento de *software*, foi aplicada o modelo espiral. Essa metodologia é mais flexível, por se tratar de um modelo baseado na prototipação cíclica e evolutiva (BOEHM, 1988; KOSCIANSKI; SANTOS SOARES, 2007). Assim, é possível ajustar possíveis mudanças que aconteçam ao longo do desenvolvimento.

A linguagem de programação padrão empregada durante o projeto foi a linguagem Python (<https://python.org>). No que diz respeito à recuperação de informação, podem ser utilizados o *framework* Scrapy (<https://scrapy.org>), a biblioteca Requests (<https://docs.python-requests.org/>) para a colheita das notícias, e a API Yahoo! Finance (<https://github.com/ranaroussi/yfinance/>) para colheita do histórico dos índices de todas as empresas listadas no mercado de ações brasileiro (B3).

Com relação às tarefas de aprendizado de máquina e manipulação dos dados, foram utilizadas as bibliotecas: Keras (<https://keras.io/>), Scikit-Learn (<https://scikit-learn.org/stable/>), Numpy (<https://numpy.org/>) e Pandas (<https://pandas.pydata.org/>). A técnica de inteligência artificial utilizada neste caso, foi baseada na arquitetura de redes neurais recorrentes LSTM, em virtude, da necessidade de analisar as influências das notícias através do tempo, juntamente, com análise do histórico dos índices das ações.

Para a classificação quanto a subjetividade das informações na base de dados de notícias, a princípio, foram testadas as ferramentas de análise de sentimentos disponíveis em língua portuguesa. Algumas ferramentas interessantes foram desenvolvidas pelo Núcleo Interinstitucional de Linguística Computacional da USP, Subjectivity Classifier for Portuguese (LUANA B. BELISÁRIO LUIZ G. FERREIRA, 2019), Opcluster-PT (VARGAS; PARDO, 2018), LBC (AVANCO; NUNES, 2014).

Além disso, o projeto considerará também o *framework* de análise de sentimentos, desenvolvido pelo Departamento de Ciência da Computação da UFMG, iFeel (MESSIAS *et al.* 2017, 2016; GONÇALVES, P. *et al.* 2016; FABRÍCIO; CHA, 2013; CANUTO; GONÇALVES, M. A.; BENEVENUTO, 2016; ARAÚJO, M.; DINIZ *et al.* 2016; ARAÚJO, M.; GONÇALVES, P.; CHA *et al.* 2014; ARAÚJO, M.; GONÇALVES, P.; BENEVENUTO, 2013). E outras ferramentas que surgiram no decorrer do projeto também foram analisadas.

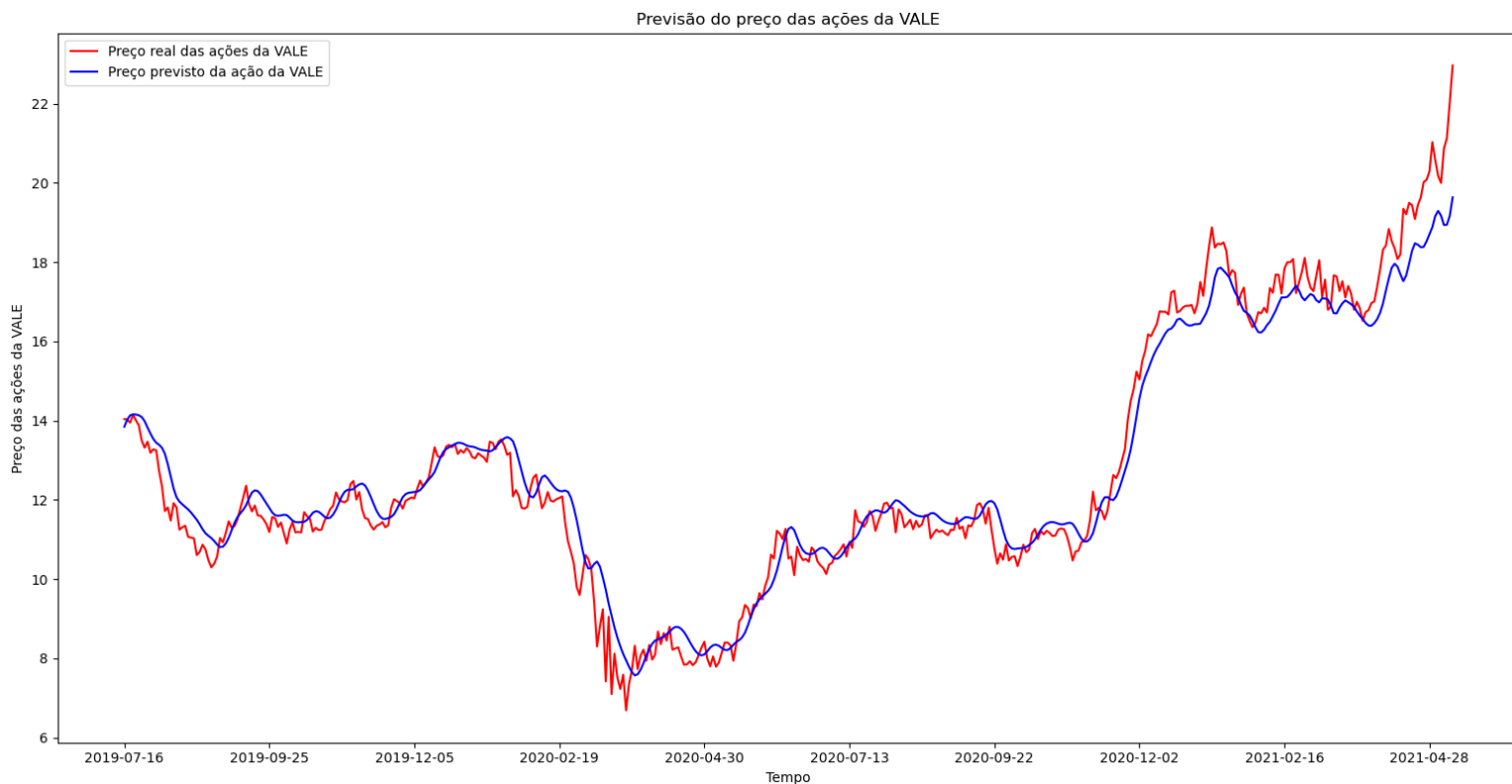
As ferramentas de análise de sentimento levantadas foram analisadas quanto a sua viabilidade e resultados, através de testes de performance. Feito isso, as ferramentas mais adequadas foram utilizadas para realizar a análise de sentimento das bases de dados de notícias e de *fake news*.

Em seguida, o histórico de sentimento das notícias, e dos respectivos índices do mercado de ações, valor de abertura, valor de fechamento, valor mínimo e máximo diário, volume de negócios, serviu como base de treinamento para criação do modelo preditivo. O mesmo foi submetido a uma avaliação de desempenho e acurácia. Isto é possível, utilizando as medidas de precisão comumente utilizadas na área de Aprendizado de Máquina. Dessa maneira, o sistema desenvolvido pode informar também o nível de confiança das previsões.

RESULTADOS E DISCUSSÕES:

Após a execução das tarefas descritas na seção anterior, foi possível obter um modelo preditivo aplicado ao mercado financeiro. *A priori*, como na seção de metodologia foi descrito os resultados obtidos a partir dos procedimentos realizados. Esta seção possuirá maior foco na discussão e análise dos resultados alcançados, através de estudo sobre o gráfico plotado com a execução da ferramenta computacional *News Impact*. De modo a exemplificar a funcionalidade desta ferramenta, a princípio, é necessário escolher uma das empresas presentes no mercado de ações brasileiro (B3), neste caso, de maneira arbitrária a empresa VALE foi selecionada. A ferramenta computacional desenvolvida, demonstrou resultados satisfatórios, como pode-se observar na Figura 2, a seguir.

Figura 2 – Resultado obtido através da ferramenta *News Impact* sobre a previsão das ações da VALE.



Fonte: Os autores (2021).

Através da Figura 2, é possível concluir que esta ferramenta demonstra resultados congruentes com as variações reais do preço das ações da VALE. Como pode ser observado, no período de março de 2020, com a chegada da pandemia no Brasil, o preço das ações da VALE foi afetado negativamente de maneira drástica, assim como diversos outros setores da sociedade, como saúde e educação. Devido ao contexto da pandemia, é importante evidenciar esta informação na análise do gráfico presente na Figura 2.

Também, é pertinente ressaltar, que nestes resultados apresentados, ainda não estão inclusos o histórico dos resultados da análise de sentimentos ao modelo preditivo. Logo, as expectativas são muito promissoras com relação ao desenvolvimento desta ferramenta. Espera-se, concluir esta adição dos atributos que contêm o histórico quantificado das expressões subjetivas presentes nas notícias e *fake news* ao modelo preditivo, de acordo com o cronograma da pesquisa. A fim de realizar o estudo sobre como as notícias reais e falsas afetam a economia e a sociedade. Outra informação importante, é salientar que esta ferramenta é um protótipo, e que pode vir a se desdobrar em diversos outros trabalhos.

CONCLUSÕES:

Em virtude das informações abordadas no decorrer deste trabalho, é possível observar a importância de ferramentas de previsões. Como é possível observar através da Figura 2, no período final do mês de fevereiro

de 2020, antes, do coronavírus atingir o Brasil, a economia do país já estava sendo afetada de maneira indireta. Vislumbra-se que as principais conclusões obtidas a partir dos resultados apresentados da pesquisa sejam as seguintes, até o presente momento, a ferramenta ainda não está completa, e mesmo assim ela demonstra resultados satisfatórios e compatíveis com os preços reais das ações, com pequenas variações dentro do intervalo de confiança das técnicas de aprendizagem de máquina.

Com a conclusão do presente projeto, será possível obter uma ferramenta computacional capaz de prever os impactos de *fake news*, confrontando com os de notícias. Através de uma correlação, com histórico de alguns dos índices de determinadas empresas da B3. Com a adição de atributos de históricos da quantificação de sentimentos presentes nas notícias. Portanto, é possível estabelecer com maior precisão a relação com os índices da bolsa de valores brasileira, viabilizando possíveis trabalhos futuros a respeito do tema. Tornando possível, a previsão das influências de notícias e *fake news* no mercado de ações. Logo, há possibilidade de redução dos prejuízos em diversos setores da economia brasileira, visto que com base nessas informações é possível, antecipar o desdobramento de diversos problemas, permitindo, desta forma, a tomadas de decisões antecipadas.

REFERÊNCIAS BIBLIOGRÁFICAS:

- ARAÚJO, J. G. de; MARINHO, L. B. Using online economic news to predict trends in brazilian stock market sectors. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 24., Salvador. Proceedings. New York: Association for Computing Machinery, 2018. p. 37–44.
- ARAÚJO, M.; DINIZ, J. P. et al. iFeel 2.0: a multilingual benchmarking system for sentence-level sentiment analysis. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, 10., Cologne. Proceedings. ACM, 2016.
- ARAÚJO, M.; GONÇALVES, P.; BENEVENUTO, F. Measuring sentiments in online social networks. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 19. Proceedings. ACM, 2013. p. 97–104.
- ARAÚJO, M.; GONÇALVES, P.; CHA, M. et al. iFeel: a system that compares and combines sentiment analysis methods. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 23. Proceedings. ACM, 2014. p.75–78.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques - "Part II: Soft computing methods." Expert Systems with Applications, v. 36, 3, Part 2, p. 5932–5941, 2009.
- AVANCO, L. V.; NUNES, M. d. G. V. Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese. In: BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEMS. Proceedings. IEEE, 2014. p. 277–281.
- BOEHM, B. W. A spiral model of software development and enhancement. Computer, v. 21, n. 5, p. 61–72, 1988.
- CANUTO, S.; GONÇALVES, M. A.; BENEVENUTO, F. Exploiting New Sentiment-Based Meta-level Features for Effective Sentiment Analysis. In: ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, 9. Proceedings. ACM, 2016.
- CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. Applied Artificial Intelligence, Taylor & Francis, v. 34, n. 1, p. 1–19, 2020.
- CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. d. The Influence of Tweets and News on the Brazilian Stock Market through Sentiment Analysis. In: BRAZILLIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 25., Rio de Janeiro. Proceedings. New York: Association for Computing Machinery, 2019. p. 385–392.
- CHAN, W. S. Stock price reaction to news and no-news: drift and reversal after headlines. Journal of Financial Economics, Elsevier, v. 70, n. 2, p. 223–260, 2003.
- DATASENADO. Redes Sociais, Notícias Falsas e Privacidade de Dados na Internet. 2019. 113 p. Dispo-

nível em: <https://www12.senado.leg.br/institucional/ouvidoria/publicacoes-ouvidoria/redes-sociais-noticias-falsas-e-privacidade-de-dados-na-internet>. Acesso em: 15 fev. 2020.

FABRÍCIO, G. P. A. M. ; B.; CHA, M. Comparing and combining sentiment analysis methods. In: ACM CONFERENCE ON ONLINE SOCIAL NETWORKS. Proceedings. ACM, 2013. p. 27–38.

FAKHRY, B. A literature review of the efficient market hypothesis. *Turkish Economic Review*, v. 3, n. 3, p. 431–442, 2016.

GONÇALVES, P. et al. On the Combination of “Off-The-Shelf” Sentiment Analysis Methods. In: SYMPOSIUM ON APPLIED COMPUTING, 16. Proceedings. ACM, 2016.

GÜNDÜZ, D. et al. Machine Learning in the Air. *IEEE Journal on Selected Areas in Communications*, v. 37, n. 10, p. 2184–2199, 2019.

IPSOS. Fake news, filter bubbles, post-truth and trust. Set. 2018.

KOSCIANSKI, A.; SANTOS SOARES, M. dos. *Qualidade de Software-2a Edição: Aprenda as metodologias e técnicas mais modernas para o desenvolvimento de software*. Novatec Editora, 2007.

LI, X. et al. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, Elsevier, v. 69, p. 14–23, 2014.

LUANA B. BELISÁRIO LUIZ G. FERREIRA, T. A. S. P. *Classificação de Subjetividade para o Português: Métodos Baseados em Aprendizado de Máquina e em Léxico*. Simpósio Internacional de Iniciação Científica e Tecnológica da USP, Universidade de São Paulo, 2019.

MACHADO, E. J.; PEREIRA, A. C. M. Proposal and Implementation of Machine Learning Models for Stock Markets Using Web Data. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 24., Salvador. Proceedings. New York: Association for Computing Machinery, 2018. p. 61–64.

MALKIEL, G. B. The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, New Jersey, v. 17, n. 1, p. 59–82, 2003.

MEDEIROS, M.; BORGES, V. Tweet Sentiment Analysis Regarding the Brazilian Stock Market. In: BRAZILIAN

MESSIAS, J. et al. An evaluation of sentiment analysis for mobile devices. *Social Network Analysis and Mining*, v. 7, n. 1, p. 20, 2017. ISSN 1869-5469.

MESSIAS, J. et al. Towards Sentiment Analysis for Mobile Devices. In: INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING, 16. Proceedings. IEEE, 2016.

MONTEIRO, R. A. et al. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF THE PORTUGUESE LANGUAGE. Proceedings. Cham: Springer International Publishing, 2018. p. 324–334.

MORENO, J.; BRESSAN, G. FACTCK.BR: A New Dataset to Study Fake News. In: BRAZILLIAN SYMPOSIUM.

ON MULTIMEDIA AND THE WEB, 25., Rio de Janeiro. Proceedings. New York: Association for Computing Machinery, 2019. p. 525–527.

NEWMAN, N. et al. Reuters Institute Digital News Report 2019. 2019.

RUIZ, P. P.; FOGUEM, B. K.; GRABOT, B. Generating knowledge in maintenance from Experience Feedback. Knowledge-Based Systems, v. 68, p. 4–20, 2014. Enhancing Experience Reuse and Learning. ISSN 0950-7051.

SHU, K. et al. Fake News Detection on Social Media: A Data Mining Perspective. CoRR, 2017.

SILVA, R. M. et al. Towards automatically filtering fake news in Portuguese. Expert Systems with Applications, v. 146, 2020.

SUN, S.; LUO, C.; CHEN, J. A review of natural language processing techniques for opinion mining systems. Information Fusion, v. 36, p. 10–25, 2017. ISSN 1566-2535.

WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 8., Belém. Anais. Porto Alegre: SBC, 2019, p. 71–82.

VARGAS, F. A.; PARDO, T. A. S. Aspect Clustering Methods for Sentiment Analysis. In: INTERNATIONAL CONFERENCE ON THE COMPUTATIONAL PROCESSING OF PORTUGUESE, 13., Canela. Proceedings. PROPOR, 2018. p. 365–374.