

# Análise de Sentimentos de Textos do Twitter sobre Mercado de Ações Brasileiro

*Sentiment Analysis of Twitter Texts about Brazilian Stock Market*

**Valdi Leão Araújo<sup>1</sup>**

 [orcid.org/0000-0003-0045-2966](https://orcid.org/0000-0003-0045-2966)

**Bruno J. Torres Fernandes<sup>1</sup>**

 [orcid.org/0000-0002-6001-3925](https://orcid.org/0000-0002-6001-3925)

<sup>1</sup>Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. E-mail: [valdileao@gmail.com](mailto:valdileao@gmail.com)

**DOI: 10.25286/repa.v6i5.1750**

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018: ARAÚJO, V. L.; FERNANDES, B. J. T. Análise de Sentimentos de Textos do Twitter sobre Mercado de Ações Brasileiro. Revista de Engenharia e Pesquisa Aplicada, Recife, v.6, n. 5, p. 18-26, Novembro, 2021.

## RESUMO

Utilizada para determinar a atitude, opinião, emoção expressa por alguém sobre um assunto em particular, a análise de sentimentos vem proporcionando diversos estudos sobre a importância da mineração de opinião com o preço no mercado de ações. Pesquisas demonstram que há uma forte relação entre a notícia sobre uma empresa e suas flutuações nos preços das ações. O objetivo desta pesquisa é desenvolver um modelo que identifique e classifique a polaridade na análise de sentimentos utilizando redes neurais LSTM, no qual, foram extraídos de perfis especializados no mercado de ações brasileiro da rede social Twitter, para empresas previamente definida: Petrobras, Vale, JBS, Eletrobras e Oi. Os resultados obtidos permitiram classificar positividade, neutralidade e negatividade dos tweets, como também realizar uma análise e visualizar a frequência em que ocorre.

**PALAVRAS-CHAVE:** Mercado de Ações; LSTM; Análise de sentimentos; Twitter;

## ABSTACT

Used to determine the attitude, opinion, emotion expressed by someone on a particular subject, the sentiment analysis has provided several studies on the importance of mining opinion with the price in the stock market. Research shows that there is a strong relationship between the news about a company and its fluctuations in stock prices. The objective of this research is to develop a model that identifies and classifies the polarity in the analysis of feelings using LSTM neural networks, in which they were extracted from specialized profiles in the Brazilian stock market of the social network Twitter, for companies previously defined: Petrobras, Vale, JBS, Eletrobras and Oi. The results obtained allowed to classify positivity, neutrality and negativity of the tweets, as well as perform an analysis and visualize the frequency in which it occurs.

**KEY-WORDS:** Stock Market; LSTM; Sentiment analysis; Twitter;

## 1 INTRODUÇÃO

Análise de sentimentos é um processo utilizado para determinar a atitude, opinião, emoção expressa por alguém sobre um assunto em particular. Para identificar novas oportunidades, investidores normalmente analisam reviews, avaliação, recomendação e outras formas de opinião sobre o assunto [1].

Os textos extraídos da rede social do Twitter são comumente tratados como problema de classificação (supervisionado), onde os *tweets* são comumente categorizados em três classes - positivo, negativo e neutro - com base nas opiniões expressas em *tweets* especializado no mercado de ações [2].

A importância desta mineração de opinião com o preço no mercado de ações fez com que pesquisadores identificassem uma forte relação entre a notícia sobre uma empresa e suas flutuações nos preços das ações [3].

O objetivo desta pesquisa é desenvolver um método para análise de sentimento, e um dos desafios na criação do modelo é treina-lo. Para tanto, foi realizado um processo de extração dos textos, de perfis especializado no mercado de ações brasileiro da rede social Twitter, para empresas previamente definida: Petrobras, Vale, JBS, Eletrobras e Oi. Em seguida, os textos foram submetidos a técnicas de pré-processamento e uso da aprendizagem profunda, para posteriormente, identificar a polaridade no modelo, permitindo visualizar e analisar os dados plotados com o histórico do preço das ações dos *tweets* classificados.

O artigo está organizado da seguinte forma: A Seção 2 apresenta uma visão geral dos conceitos teóricos que são a base deste artigo, tanto no contexto dos mercados de ações quanto no aprendizado de máquina. Também menciona alguns dos trabalhos relacionados a esse assunto que podem ser encontrados na literatura; A Seção 3 detalha a metodologia e o modelo proposto neste artigo, na Seção 4 os resultados são apresentados e discutidos e, finalmente, a Seção 5 conclui o artigo.

## 2 REFERENCIAL TEÓRICO

Nesta seção abordaremos os tópicos essenciais e os motivos dos direcionamentos aplicados para o desenvolvimento das técnicas utilizadas.

### 2.1 REDES SOCIAIS ONLINE

Uma Rede Social online é o ambiente digital organizado por meio de uma interface virtual própria que se organiza agregando perfis humanos que possuam afinidades, pensamentos e maneiras de expressão semelhantes e interesse sobre um tema comum [4].

Zenha fala que, é um ambiente digital em conexão no qual é possível observar o desenrolar, a evolução e a constante modificação dos embates psicossociais de seus integrantes, embates esses não apenas de ordem tecnológica, mas, sobretudo, humana. A participação ativa das pessoas nas redes sociais por meio da troca generosa de links e da catalisação de conversas apresenta um comportamento indicativo para a conexão, a ligação e a linkagem entre assuntos e pessoas [4].

### 2.2 TWITTER

Criado em 2006 por Jack Dorsey, o Twitter é um sistema de comunicação instantânea na internet disponível para o uso em computadores, tablets e celulares, denominado um blog compacto em função da limitação da inserção do número de caracteres para o registro [4].

Ele é considerado uma mídia social em formato de microblog e permite ao usuário enviar e receber atualizações pessoais de outros contatos ou seguidores. As atualizações são exibidas no perfil do usuário em tempo real e também enviadas aos seguidores que tenham assinado para recebê-las [4].

Como pontua Zenha, inicialmente, o foco do Twitter era os compartilhamentos de ações pessoais, hoje essa troca ampliou-se para discussões no âmbito profissional, questionamentos de assuntos da atualidade, divulgações de marketing, entre outros usos [4].

As razões para utilizar o Twitter são [5]:

- São utilizadas por diferentes pessoas para expressar suas opiniões sobre diferentes tópicos.
- Contem enorme número de textos que cresce a cada dia.
- O público varia de usuários regulares a celebridades, representantes de empresas, políticos e até presidentes de país.
- Público é representado por usuários de diferentes países, permitindo coletar dados de diferentes linguagens.

### 2.3 ANÁLISE DE SENTIMENTOS

A identificação de sentimentos no texto é um importante campo de estudo, com plataformas de mídia social como o Twitter, despertando o interesse de pesquisadores no processamento da linguagem e nas ciências políticas e sociais. A tarefa geralmente envolve detectar se um pedaço de texto expressa um sentimento POSITIVO, NEGATIVO ou NEUTRO. O sentimento pode ser geral ou sobre um tópico específico, como: Uma pessoa, um produto ou um evento [6].

Houve um enorme desenvolvimento no campo de processamento de linguagem natural usando métodos de redes neurais profundas para várias tarefas, incluindo tarefas mais simples, como *POS tagging* e reconhecimento de entidade mencionada, remoção de *stopwords* e *stemming*. A análise de sentimentos é um problema fundamental para fornecer soluções para uma máquina, que entenda as emoções e opiniões no texto [7].

Assim como analisa os pesquisadores V. S. Pagolu e B. Majhi [8], a tarefa de análise de sentimentos é muito específica do campo que está sendo desenvolvido. Há muita pesquisa sobre análise de sentimentos de comentários sobre filmes, artigos de notícias e muitos destes analisadores de sentimentos estão disponíveis como fonte aberta. O principal problema com esses analisadores é que eles são treinados com um corpus diferente. Por exemplo, corpus de filme e corpus de ações não são equivalentes.

Portanto, baseado neste entendimento, neste artigo será desenvolvido o próprio analisador de sentimentos.

### 2.4 WORD EMBEDDING

É uma técnica em que palavras individuais são representadas como vetores com valor real em um espaço vetorial predefinido. Cada palavra é mapeada para um vetor, e os valores do vetor são aprendidos de uma maneira que se assemelha a uma rede neural que, portanto, a técnica é frequentemente agrupada no campo de aprendizado profundo [9].

Brownlee destaca que, *Word Embedding* é uma representação de aprendizagem para texto em que palavras com o mesmo significado têm uma representação semelhante. É essa abordagem para representar palavras e documentos que pode ser considerada uma das principais descobertas de aprendizado profundo em problemas desafiadores

de processamento de linguagem natural [9]. O processo de aprendizado é realizado em conjunto com o modelo de rede neural em alguma tarefa, como a classificação de textos, ou pode ser um processo não supervisionado, usando estatísticas nos textos.

### 2.5 REDE NEURAL – LONG SHORT-TERM MEMORY

LSTM (*Long Short-Term Memory*) é um tipo de rede neural recorrente que se mostrou muito bem-sucedida em vários problemas, devido à sua capacidade de distinguir entre exemplos recentes e anteriores, fornecendo pesos diferentes para cada um, esquecendo a memória que considera irrelevante prever a próxima saída. Desta forma, é capaz de lidar com longas sequências de entrada quando comparado a outras redes neurais recorrentes que são capazes de memorizar apenas sequências curtas [10].

Uma rede recorrente LSTM (RNN), segundo o grupo de pesquisa A. Tholusuri, M. Anumala, B. Malapolu, e G. J. Lakshmi, é a mais eficiente, comparando das outras redes neurais, pois classifica os dados de sequência longa, que utiliza memória de longo prazo e, portanto, pode lidar com as dependências de longo prazo. Os resultados mostraram que o padrão algorítmico das redes de memória de curto prazo supera os outros quanto à exatidão [7].

Uma RNN bidirecional é uma variante comum do RNN que pode oferecer maior desempenho do que um RNN comum em determinadas tarefas. É frequentemente usado no processamento de linguagens naturais. Esta técnica explora esta característica para melhorar o desempenho das RNNs de ordem cronológica. A RNN analisa sua sequência de entrada de ambos os lados, obtendo representações potencialmente mais ricas e capturando padrões que podem ter sido perdidos apenas pela versão de ordem cronológica [10].

## 3 MATERIAIS E MÉTODOS

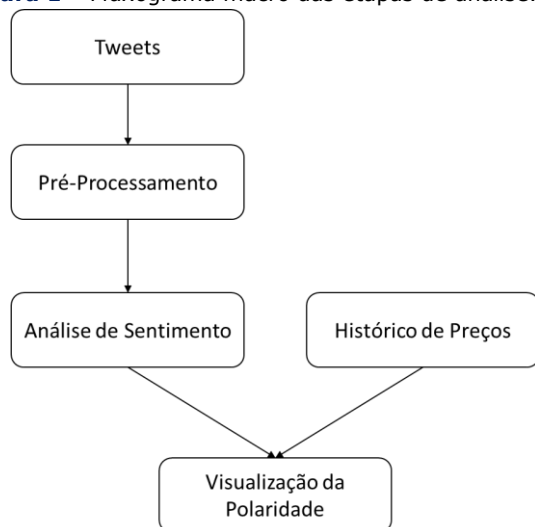
Nesta seção, será apresentada a arquitetura do sistema proposto para realização das etapas de pré-processamento, classificação dos textos e visualização do histórico do preço da ação com as polaridades previstas pelo modelo.

### 3.1 ETAPAS DE ANÁLISE

Nesta subseção do artigo, uma abordagem de combinar a previsão da polaridade de sentimentos com a detecção com o preço histórico para visualização é implementada. É mostrado com maiores detalhes o fluxograma com uma série de etapas do analisador de sentimentos.

A etapa do pré-processamento realiza a limpeza dos *tweets*, que passa para a etapa do analisador de sentimentos. O analisador de sentimentos identifica a polaridade e fornece a porcentagem de sentimentos do dia, que por sua vez, é alimentada para o visualizador de polaridade, juntamente com o preço histórico das ações. A visualização da polaridade nos auxilia analisar a relação da polaridade dos *tweets* com o preço histórico das ações.

**Figura 1** – Fluxograma macro das etapas de análise.



**Fonte:** Autor.

### 3.2 COLETA DOS DADOS

Um total de 6.384 *tweets* durante um período de 20 de dezembro de 2016 a 01 de maio de 2020 das empresas Petrobras, Vale, JBS, Eletrobras e OI foram extraídos do Twitter. Todos os *tweets* foram etiquetados manualmente como: POSITIVO, NEGATIVO e NEUTRO. Os *tweets* foram coletados e filtrados usando palavras-chave como Petrobras, Vale, JBS, Eletrobras e Oi. As opiniões consideradas foram dos públicos especializados como ADVFNBrasil, infomoney, elevenfinancial, empiricus, exame, BloombergBrasil, valoreconomico, Bastter, SunoResearchcom e CPoupadores.

Com base neste princípio, as palavras-chave usadas para filtragem são criadas com muito

cuidado e os *tweets* são extraídos de forma a representar as emoções exatas do público sobre as empresas citadas, durante um período de tempo. Os preços de abertura e fechamento do histórico das ações das empresas é diário e está entre 03 de janeiro de 2020 a 05 de maio de 2020 e foram obtidos do aplicativo MetaTrader5 [11].

### 3.3 PRÉ-PROCESSAMENTO

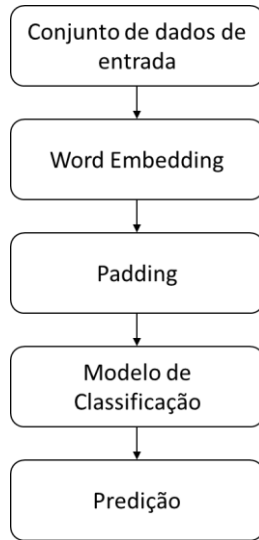
Os *tweets* consistem em muitas siglas, emoticons e dados desnecessários, como caracteres especiais e URLs. Portanto, os *tweets* são pré-processados para representar as palavras chave que identifica a polaridade na qual representa o conjunto de palavras. Para o pré-processamento dos *tweets*, foram empregados estágios de filtragem na seguinte ordem: substituir caracteres maiúsculo por minúsculos, remoção de palavras irrelevantes, remoção de caracteres especiais, URLs e remoção de conteúdos visuais. Na realização detalhada do pré-processamento, cada filtragem foi realizada da seguinte forma:

- Substituir caracteres maiúsculo por minúsculos: Para evitar inconsistências no significado das palavras, já que um caractere maiúsculo difere do mesmo caractere minúsculo, faz-se necessário substituir para um dos. Neste trabalho, foi escolhido minúsculo.
- Remoção de palavras irrelevantes: Foram removidas palavras consideradas irrelevantes no processo de classificação. Artigos, preposições, conectores não foram considerados.
- Remoção de caracteres especiais: Da mesma forma que caractere maiúsculo difere do minúsculo, caractere especial também ocorre este comportamento. Acentuação, pontuação, números não foram considerados.
- Remoção URLs: Links internos ou externos não foram considerados para o desenvolvimento deste trabalho, já que não faz parte do escopo.
- Remoção de conteúdos visuais: Imagens, vídeos e emoji não foram considerados para o desenvolvimento deste trabalho, já que não faz parte do escopo.

### 3.4 ANÁLISE DE SENTIMENTOS

Para aplicar a análise de sentimentos, foi utilizada a biblioteca Keras, rodando sobre a biblioteca TensorFlow para desenvolver a polaridade na rede neural LSTM Bidirecional.

**Figura 2** – Diagrama arquitetural do processo de análise de sentimento.

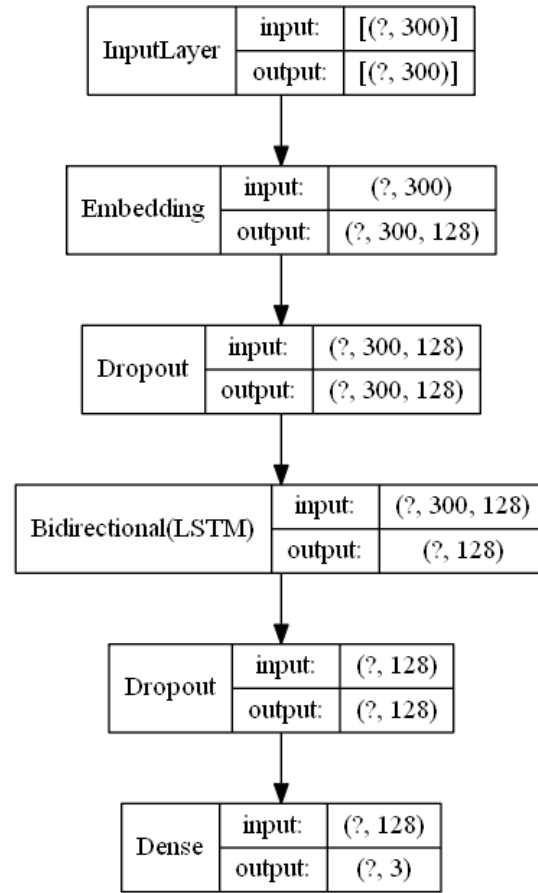


Fonte: Autor.

Os recursos extraídos da coleta de dados dos *tweets* etiquetados manualmente, são representados no formato *One-hot Encoding*. Então, o conjunto de dados é dividido em três conjuntos: treinamento, teste e validação. Neste momento, é aplicado o processo de Tokenização de cada palavra extraída dos textos, no qual é aplicado no conjunto de texto pré-processado que, em termos práticos, é o processo de gerar token com cada palavra do texto, classificando cada token em um número único no vetor. No próximo passo, é necessário aplicar no conjunto de dados o *Padding*, para que sequências mais curtas que outras devem ser preenchidas com zeros e sequências mais longas devem ser truncadas.

Conforme ilustrado na Figura 3, o modelo é treinado com o tipo da rede neural LSTM Bidirecional com uma série de camadas auxiliares. Utilizando a camada do módulo Keras *Embedding*, com a função de aprender as relações contextuais entre as palavras nos dados de treinamento. O *Word Embedding* é realizado com a ajuda dessa camada, que reduz a alta dimensionalidade do texto. Pelo fato de haver poucos dados, foi utilizado Dropout entre as camadas para permitir o modelo generalize melhor. Uma camada LSTM Bidirecional se mostrou efetiva e foi aplicada. O modelo apresenta uma saída de classificação da camada Dense, que retorna um dos três possíveis valores da rede neural no formato *One-Hot Encoding*.

Figura 3 – Modelo de classificação RNN com LSTM Bidirecional.



Fonte: Autor.

Durante o período de 20 de dezembro de 2016 a 28 de dezembro de 2019 os *tweets* das empresas Petrobras, Vale, JBS, Eletrobras e OI foram utilizados como base de dados para treinar, testar e validar o modelo conforme a Quadro 1. Eles foram divididos na proporção da quantidade de *tweets* de cada classe. Os Positivos = 1901, Negativos = 1743 e Neutro = 2400. Os resultados da classificação de sentimentos são discutidos na seção a seguir. O classificador desenvolvido é usado para prever as emoções dos *tweets* selecionados pelas polaridades: Positivo, Negativo e Neutro.

Quadro 1 – Parâmetros gerais do modelo.

PARÂMETRO	VALOR
Base de dados	80% Treino, 10% teste, 10% validação

Número Máximo de Recursos		6044
Tamanho do Embedding		128
Tamanho do LSTM Bidirecional		64
Tamanho do Batch		32
Máximo de Época		2
Otimizador		Adam
Loss		Entropia cruzada categórica
Ativação		Softmax

### 3.5 ANÁLISE DE POLARIDADE

Os *tweets* e o histórico de preços no período de 01 de janeiro de 2020 a 05 de maio de 2020 das empresas Petrobras, Vale, JBS, Eletrobras e OI, foram coletados para visualização.

A pontuação dos sentimentos dos *tweets* foi implementado aplicando a média das polaridades dos sentimentos realizado por dia. Para gerar um resultado claro para visualização, foi empregado para cada polaridade os seguintes valores: Positivo = 3, Negativo = 1 e Neutro = 2.

A visualização do gráfico do preço das ações, como também a polaridade de cada empresa analisada, foram plotados para termos uma dimensão visual de como a relação se distribui. A precisão da análise e a visualização das séries temporais é demonstrada na seção de resultados.

## 4 RESULTADOS

Esta seção fornece uma visão geral das taxas de precisão da Rede Neural LSTM Bidirecional treinada, como também uma análise de relação entre histórico de preço e a polaridade dos *tweets* nos papéis: Petrobras, Vale, JBS, Eletrobras e Oi. Na subseção 1, será demonstrado os resultados obtidos com o classificador da polaridade dos *tweets*. Na subseção 2 veremos mais detalhes dos dados plotados sobre os resultados das polaridades geradas e histórico de preço das ações.

Nesta seção, será apresentada a arquitetura do sistema proposto para realização das etapas de pré-processamento, classificação dos textos e visualização do histórico do preço da ação com as polaridades previstas pelo modelo.

### 4.1 ANÁLISE DE SENTIMENTOS

O modelo teve resultado satisfatório na fase de treinamento, nos dados de treino e teste, obtendo 79% e 75%, respectivamente, na tarefa de classificação na detecção dos *tweets* das empresas no mercado de ações.

Verificando o resultado do loss da Figura 4, percebe-se que existe uma queda entre os dados de treino e teste. Um ponto importante é que não foi utilizado mais épocas na fase de treinamento, pela resposta da acurácia quanto ao tamanho do batch utilizado, alcançando assim o melhor percentual na classificação.

**Figura 4** – Resultado do treino e teste do Modelo RNN com LSTM Bidirecional.

```
Train on 4835 samples, validate on 605 samples
Epoch 1/2
4835/4835 - 204s - loss: 0.9369 - accuracy: 0.5396 - val_loss: 0.7045 - val_accuracy: 0.7058
Epoch 2/2
4835/4835 - 199s - loss: 0.5499 - accuracy: 0.7917 - val_loss: 0.6095 - val_accuracy: 0.7554
```

**Fonte:** Autor.

Com os dados de validação o modelo alcançou intuitivamente 74% de precisão. O modelo LSTM Bidirecional alcançou um desempenho satisfatório em termos de precisão (Precision), Sensibilidade (Recall) e Medida-F (F1-Score).

**Figura 5** – Resultado da previsão do Modelo RNN com LSTM Bidirecional.

```
604/604 - 3s - loss: 0.6411 - accuracy: 0.7434
Score: 0.64
(Evaluate) Validation Accuracy: 0.74
precision recall f1-score support

Negativo 0.70 0.75 0.72 165
Positivo 0.74 0.70 0.72 195
Neutro 0.77 0.78 0.78 244

accuracy 0.74 0.74 0.74 604
macro avg 0.74 0.74 0.74 604
weighted avg 0.74 0.74 0.74 604

Pred Negativo Pred Positivo Pred Neutro
Negativo 123 19 23
Positivo 26 136 33
Neutro 26 28 190
```

**Fonte:** Autor.

Podemos ver na Figura 5, a precisão do modelo LSTM Bidirecional e a classificação correta e incorreta com ajuda dos rótulos das classes negativo, positivo e neutro e sua precisão. Avaliando a classe Positivo, que teve desempenho igual ao negativo analisando pelo F1-Score, porém o neutro teve a melhor precisão entre as três

classes. O desempenho da classe Negativo e Positivo, demonstra ser menor em todos os indicadores, em comparação com a classe Neutro. No entanto, é importante enfatizar, que com a inclusão de mais dados de treinamento dos *tweets*, haverá uma melhoria significativa no desempenho não só da classe Positivo, como também de todo o modelo.

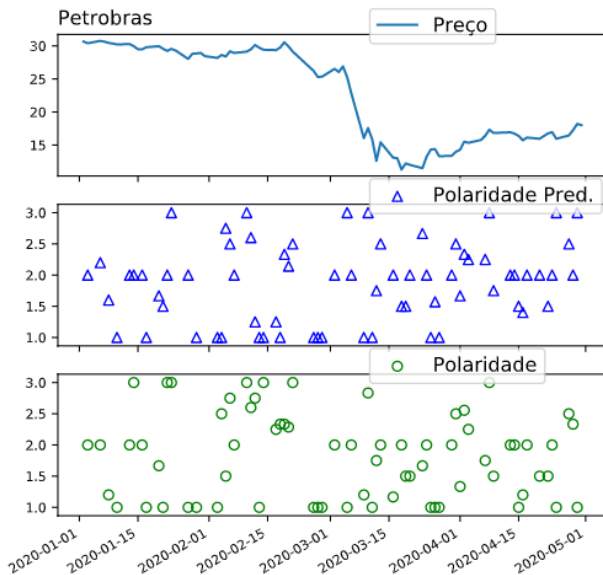
Na matriz de confusão, podemos ver destacado as classes negativo, positivo e neutro, rotulados como correto e incorreto pela coluna com sufixo "Pred". Os números na diagonal são os negativos, positivos e neutros rotulados corretamente, respectivamente.

#### 4.2 VISUALIZAÇÃO DA POLARIDADE

As Figuras 6, 7, 8, 9 e 10 foram plotados os gráficos, que para cada um, contendo o histórico de preço da ação e a média de polaridade dos *tweets* realizados em cada dia.

Na Figura 6, vemos uma polaridade bem distribuída principalmente antes da queda dos preços acentuada. Porém na mudança de tendência depois da data 2020-03-15, é perceptível a polaridade em uma posição intermediária e alguns no campo positivo.

**Figura 6** – Série temporal e predição da análise de sentimento da empresa Petrobras.

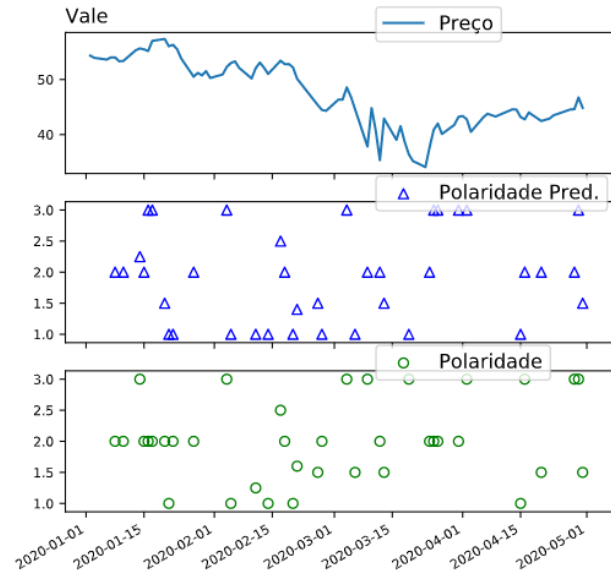


Fonte: Autor.

No gráfico da Figura 7, os valores do preço da ação apresentam-se "comportado" em comparação com a Figura 6, onde a polaridade distribui no centro. A tendência da polaridade se mostra em uma posição intermediária em todo o gráfico com

um pouco acentuado no negativo entre 2020-01-15 a 2020-03-15.

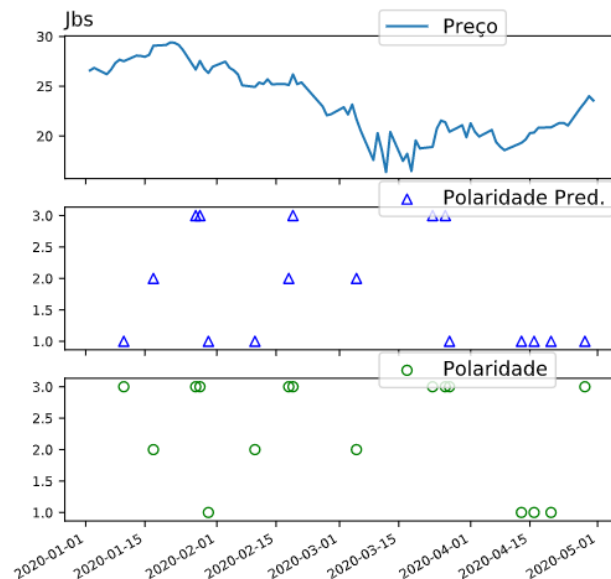
**Figura 7** – Série temporal e predição da análise de sentimento da empresa Vale.



Fonte: Autor.

Na Figura 8, o gráfico com os valores do preço da ação não tem uma queda fortemente acentuada. Percebe-se que a polaridade ocorre em sincronia com o preço das ações que vimos até então. A tendência da polaridade se mostra variável em todo o período do gráfico.

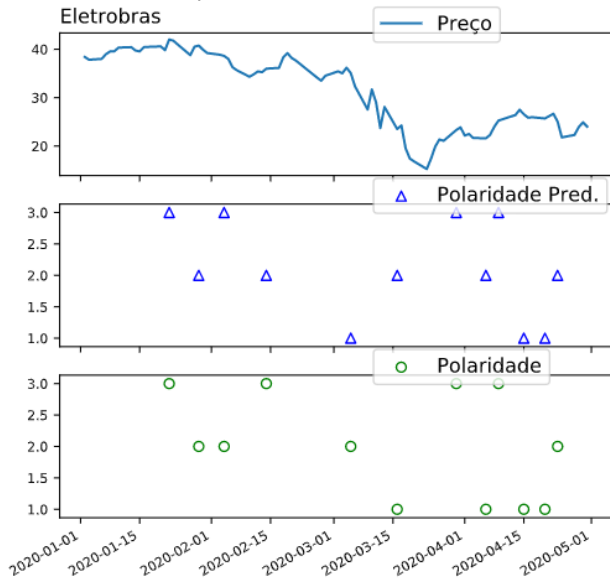
**Figura 8** – Série temporal e predição da análise de sentimento da empresa JBS.



Fonte: Autor.

É perceptível no gráfico da Figura 9, os valores do preço das ações em "sintonia" com a polaridade. Apesar de existir poucos dados, demonstra uma certa harmonização entre o preço da ação e a polaridade em todo o período do gráfico.

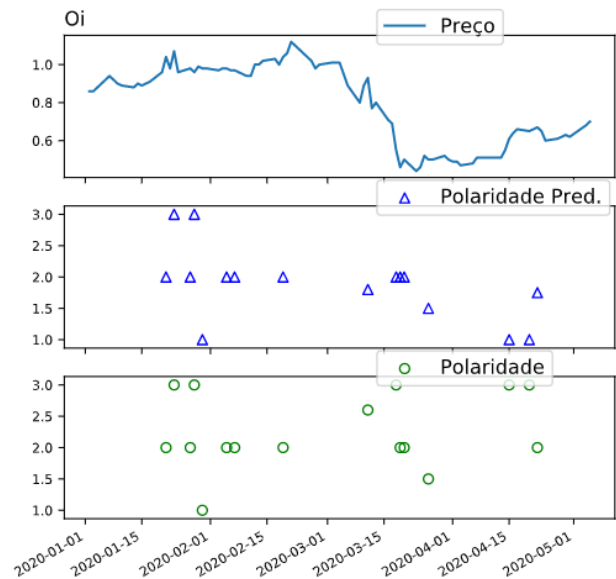
Figura 9 – Série temporal e predição da análise de sentimento da empresa Eletrobras.



Fonte: Autor.

Na Figura 10, temos uma acentuação no valor do preço da ação entre 2020-03-01 a 2020-04-01. Na polaridade extraída, é perceptível a queda de neutro para negativo neste período do gráfico. É visto que após 2020-04-15, a polaridade começa a harmonizar com a tendência do preço das ações visto nos gráficos anteriores.

Figura 10 – Série temporal e predição da análise de sentimento da empresa Oi.



Fonte: Autor.

## 5 CONCLUSÕES

A principal contribuição deste trabalho é análise de sentimentos que pode distinguir entre os tweets positivos, negativos e neutros do mercado de ações brasileiro no Twitter com a precisão de 74%. Embora tivemos um modelo aceitável, caso queira melhorar a precisão do modelo, se faz necessário realizar um treinamento com mais dados. O modelo de fato demonstrou ter uma generalização e uma precisão equilibrada entre as classes, conforme resultado da Figura 5. É perceptível também os Falsos Negativos das classes Positivo e Negativo para a classe Neutro.

A análise visual, mostra uma relação entre a polaridade nos tweets relacionados ao preço das ações com as consequências na tendência. É visível nos gráficos plotados, identificar uma certa relação entre a polaridade com a variação do preço das ações e vice versa. Este se torna claro, com a quantidade de dados na polaridade como Petrobras e Vale. Nas empresas JBS, Eletrobras e Oi, percebemos que existe uma relação em certos pontos da série temporal, mas necessitando de uma maior quantidade de dados para tornar visível a relação.

Em trabalhos futuros, o desenvolvimento de um estudo com técnicas de correlação, para identificar não só a tendência, como a reciprocidade da correlação, como o grupo de pesquisa J. Smailović, M. Grčar, N. Lavrač, e M. Žnidaršič sobre a causalidade de Granger e o estudo do grupo de pesquisa D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, e B. K. Lama sobre correlação de Pearson.



Ambos estudos demonstra a necessidade de entender a causalidade da correlação [12][13]. Outro ponto importante destacado nos *tweets*, é o de notícias relacionadas ao produto ou serviço que tem uma influência no resultado da empresa, enriquecendo ainda mais a avaliação de correlação e a possibilidade de predizer a tendência futura da ação.

A utilização de novas abordagens de técnicas de redes neurais, se torna um fator importante para avaliarmos a melhor técnica no cenário em que se deseja aplicar. A técnica da rede neural Transformer, que tem apresentado resultados importantes para a área de processamento de linguagem natural, ficou conhecida pelo modelo BERT, e se torna importante realizar um trabalho com esta técnica [14].

### REFERÊNCIAS

- [1] Y. Chen, Q. You, J. Yuan, e J. Luo, **Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM**, 2018.
- [2] Y. Yan, H. Yang, e H. Wang, **Two Simple and Effective Ensemble Classifiers for Twitter Sentiment Analysis**, no July, p. 1386–1393, 2017.
- [3] K. Joshi e P. J. R. Prof. Bharathi H. N., **STOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS**, vol. 8, no 3, p. 1–8, 2016.
- [4] L. Zenha, **Redes sociais online: o que são as redes sociais e como se organizam?**, p. 19–42, 2018.
- [5] A. Pak e P. Paroubek, **Twitter as a Corpus for Sentiment Analysis and Opinion Mining**, p. 1320–1326.
- [6] S. Rosenthal, N. Farra, e P. Nakov, **SemEval-2017 Task 4: Sentiment Analysis in Twitter**, 2018, p. 502–518.
- [7] A. Tholusuri, M. Anumala, B. Malapolu, e G. J. Lakshmi, **Sentiment Analysis using LSTM**, no 6, p. 1338–1340, 2019.
- [8] V. S. Pagolu e B. Majhi, **Sentiment Analysis of Twitter Data for Predicting Stock Market Movements**, p. 1345–1350, 2016.
- [9] J. Brownlee, **What Are Word Embeddings for Text?**, 2019. Disponível em: <https://machinelearningmastery.com/what-are-word-embeddings/>. Acesso em: 12 abr. 2020.
- [10] F. Chollet, **Deep Learning with Python**. Editora: Manning Publications, 1ª edição, 2018.
- [11] M. LTD., **MetaTrader5**. [Online]. Disponível em: <https://www.metatrader5.com/>. Acesso em: 23 abr. 2020.
- [12] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, e B. K. Lama, **Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis**, Proc. 2018 IEEE 3rd Int. Conf. Comput. Commun. Secur. ICCCS 2018, p. 128–132, 2018.
- [13] J. Smailović, M. Grčar, N. Lavrač, e M. Žnidaršič, **Stream-based active learning for sentiment analysis in the financial domain**, Inf. Sci. (Ny), vol. 285, no 1, p. 181–203, 2014.
- [14] J. Devlin, M.-W. Chang, K. Lee, e K. Toutanova, **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**, no Mlm, 2018.