

Identificando Sentimentos de Textos Referentes Investimentos com Processamento de Linguagem Natural

Brenda dos Passos Gonçalves Nunes, Isabella Ramos Fonseca, João Pedro França, Renan Henrique da Silva Abbade, Calebe de Paula Bianchini

Faculdade de Computação e Informática Mackenzie (FCI) - Universidade Presbiteriana Mackenzie - Rua da Consolação, 930 – 01302-907 – São Paulo – SP – Brasil

nunesgbrenda@gmail.com, isabella-ramos@hotmail.com,
joaaofranca@hotmail.com, renanabbade1997@gmail.com,
calebe.bianchini@mackenzie.br

Abstract. *This research is a solution based on natural language processing where the application will use the TextBlob API to capture feelings, from real-time databases such as Twitter and news sites specialized in investments analysis regarding the chosen investment in order to directly impact the investor's decision.*

Key-words: *Sentiment analysis, natural language processing, application.*

Resumo. *Este trabalho trata-se de uma solução baseada em processamento de linguagem natural onde o sistema desenvolvido utilizou API TextBlob para capturar sentimentos, de bases em tempo real como Twitter e sites de notícia especializados em investimentos, onde a partir desses dados foi gerado um relatório contendo análises de sentimentos referentes ao investimento escolhido, de modo a impactar diretamente na decisão do investidor.*

Palavras chaves: *Análise de sentimentos, processamento de linguagem natural, aplicação.*

1. Introdução

A análise de sentimentos pode ser realizada por meio de processamento de linguagem natural, linguística computacional e biometria. Seu objetivo é segmentar ou demonstrar as sensações de usuários através de bases de dados sólidas. Esta tecnologia apresenta diversas oportunidades de extrair informações valiosas para diversos negócios. Na pesquisa, foi encontrado um relevante tema - pensando na vida financeira da população - o investimento. E, a partir deste tema, o objetivo é extrair da análise de sentimentos uma solução que será um relevante aliado para o usuário da aplicação, no caso, o investidor.

A partir dessas informações surgiu a ideia de desenvolver uma aplicação que tem como objetivo auxiliar o investidor utilizando linguagem natural (campo da inteligência artificial cujo objetivo é analisar a língua natural para os humanos) e análise de sentimentos (identifica a opinião que foi expressa em um determinado texto), a escolher o melhor tipo de investimento em um determinado momento, através das redes sociais. Essa ferramenta busca consolidar inúmeras notícias por meio de um dashboard para o usuário conferir e tomar decisões por meio das análises geradas pela aplicação.

Uma poderosa fonte de informação vem através da imprensa, como sites de notícias. Eles podem influenciar o processo de tomada de decisões dos investidores, por conta disso, deve-se sempre verificar a veracidade e autenticidade de todas as notas. Por esse motivo a aplicação filtra confiáveis sites e realiza o agrupamento para um melhor aproveitamento dos usuários. A aplicação dá suporte sobre o impacto que essas notícias podem ter e as consolida diversos domínios em um só lugar.

Com isso, o sistema permite filtrar qual ativo de investimento verá e se o mesmo possui uma melhor visualização de conteúdo relacionado. E assim, pode ser tomada uma melhor decisão na compra e venda de seus ativos, se baseando em notícias de mercado em tempo real e opiniões de investidores.

O trabalho foi estruturado da seguinte forma: Na seção 2, foi abordado o referencial teórico em inúmeras subseções (mineração de textos, preparação dos dados, indexação e normalização, cálculo de relevância, seleção dos termos e análise de sentimentos) e informações pertinentes a literaturas em que a pesquisa foi baseada. Na seção 3 foi descrita a metodologia da pesquisa e de construção da aplicação. A seção 4 apresenta os resultados obtidos, na seção 5 a conclusão da pesquisa e, por fim, na seção 6 estão as referências bibliográficas utilizadas.

2. Referencial Teórico

2.1 Linguística computacional

O trabalho de VIEIRA, R. STRUBE, V. (2001) mostra o funcionamento da linguística computacional e todo o desenvolvimento que é feito para relação que existe entre a linguística e a informática. A linguística computacional é determinada por entendimento dos conceitos gramaticais para que a máquina realize as concordâncias corretas. O procedimento de verificação de sintaxe é parecido com um de linguagem de programação, a complexidade vem por conta da separação da língua escrita e da língua falada, da interpretação, do entendimento da semântica, da interpretação da linguagem, da geração e da linguagem em si. A área de linguística computacional tem grande papel para a sociedade pois auxilia, inclui e torna a informação mais acessível.

2.2 Processamento de linguagem natural

Processamento de linguagem natural ou *natural language processing* (PLN ou NLP) é uma tecnologia cujo objetivo é analisar a linguagem natural para humanos a fim de abstrair significados relevantes por meio de algoritmos para compreensão de máquinas. Com o auxílio do PLN, os computadores podem ler textos, reconhecer os trechos mais relevantes, analisar e interpretar discursos, realizar resumos, criar classificações de conteúdo, além de traduzir textos.

O fato de que a linguagem humana pode se expressar de maneiras diferentes torna a interpretação mais complicada. O PLN pode não apenas analisar regras gramaticais, mas também compreender abreviações, corrigir erros de pontuação, acentos regionais, entre outras funções.

2.3 Mineração de textos

Após realizadas as etapas de seleção de textos escritos em linguagem natural para análise, a aplicação passará por um processo de extração de informações úteis, classificadas por relevância de acordo com o algoritmo. O objetivo da mineração de textos é extrair informações de interesse de documentos de texto com grande volume de dados.

Segundo ARANHA, C. (2007) podemos entender que a mineração de texto consiste em várias etapas, sendo as principais:

A coleta de dados, que é responsável por buscar e recuperar os dados a fim de conseguir o material a ser analisado.

O pré-processamento tem o objetivo de fornecer uma primeira formatação do documento e aumentar a qualidade dos dados desse documento. O resultado do pré-processamento é um conjunto de termos independentes.

A indexação funciona como um índice, onde organiza os termos obtidos pela formatação e facilita o acesso ao documento.

A etapa de mineração é responsável por desenvolver os cálculos necessários e decidir quais algoritmos serão usados. As tarefas de mineração que decidem qual informação se quer obter. Entre elas: clusterização; classificação; sumarização; e extração de características.

A etapa de análise da informação é responsável por avaliar e verificar se a aplicação dos algoritmos da etapa de mineração foram efetivos. Existem algumas maneiras de analisar a informação como um todo sendo de forma quantitativa ou qualitativa, os resultados para uma melhor visualização de resultados por meios de gráficos podem ser usados por ferramentas de visualização.

2.4 Análise de Sentimentos

A análise de sentimentos é um recurso de processamento de linguagem natural, que tem por objetivo extrair sentimentos de determinadas frases ou textos, de modo a auxiliar em uma eventual análise referente a algum produto ou empresa. Para a pesquisa, as opiniões dos usuários serão reunidas utilizando por categorizações: positivas, negativas e neutras.

Para LIU, HU, & CHENG (2005) existem termos importantes que são utilizados na análise de sentimentos, sendo eles;

- Objeto: refere-se ao documento a ser analisado.
- Componentes: são as características do documento. Podem ser explícitas e implícitas, onde as explícitas são sentimentos expressos diretamente e implícitas são sentimentos expressos indiretamente.
- Opinião: é o sentimento emitido pelo usuário ou entidade. Também podem ser explícitas e implícitas, onde são expressos diretamente e indiretamente respectivamente.
- Entidade: quem expressa a opinião.
- Polaridade: determina se a expressão da opinião é positiva ou negativa.

2.4.1 Métodos para Análise de Sentimentos no Twitter

Para o trabalho de ARAÚJO M., GONÇALVES P., BENEVENUTO F. (2013), foi usado duas bases de dados onde foi possível utilizar 8 métodos de análise de sentimentos para comparação, sendo eles: *SentiWordNet*, *SASA*, *PANAS-t*, *Emoticons*, *SentiStrength*, *LIWC*, *SenticNet*, e *Happiness Index*. O objetivo das comparações entre os métodos foi focar na polaridade dos conteúdos extraídos da web. Para saber a eficácia do método, foram utilizadas algumas métricas como: a abrangência, a concordância, a taxa de *true positive*, taxa de *true negative*, acurácia, precisão e *F-measure*. Diante dos resultados, não foi possível determinar um método com melhores resultados, entretanto foi notado vários graus de abrangência e acurácia nos métodos. A partir dessas comparações os pesquisadores testaram uma nova forma que chamaram de “método combinado”, onde 7 dos 8 métodos foram combinados, e a partir do resultado desse método foi concluído que combinar os métodos que mais se ajustam a necessidade que o usuário solicita pode se obter resultados bons.

Para o trabalho de RAMOS, P. BENEVENUTO (2016), foi utilizado um conjunto de dados, rotulados como baseline, no qual por meio deste, foi comparada a eficiência das estratégias de análise de sentimento, utilizando o método *SentiStrength*, juntamente com a Microsoft Translator API para a tradução das notícias para o idioma inglês, pois o método utilizado opera com sentenças na língua inglesa somente. Com relação ao desempenho do método *SentiStrength*, foi destacado uma eficácia de 64%, mediante as categorias de sentimento negativa, neutra e positiva, destacando as polaridades extremas como, respectivamente -4 e +4, para análises de sentimento realizadas em notícias dos principais sites nacionais.

A partir da pesquisa de GOMES, H. J. C. (2012) foi testado e avaliado alguns títulos de notícias por meio de endereços de RSS Feeds de economia, no qual a intenção foi analisar a polaridade desses títulos (entre negativo, positivo ou neutro). O software utilizado para fazer as análises de sentimento e mineração das palavras foi o SAS. Foram desenvolvidos modelos de análise de sentimentos, sendo eles o modelo estático, o baseado em regras e o híbrido. Foi concluído que notícias neutras podem ser um problema para serem identificadas na fase de aprendizagem de máquina desses tipos de modelos, assim aumentando a possibilidade de polaridades falsas positivas e falsas negativas. Em relação às notícias negativas os modelos classificam mais corretamente por conta dos meios de comunicação darem mais ênfase em notícias negativas.

2.5 Trabalhos Correlatos

A partir da pesquisa de AUGUSTO G. (2018), foi avaliado a relação entre assertividade na descrição de cartões *Jira*, medindo nível de emoções, métricas de produtividade e analisando sentimentos de colaboradores envolvidos no desenvolvimento do cartão. O trabalho considerou dados apenas de uma aplicação de projeto do *Jira* explorando maneiras diferentes de calcular sentimento e verificação da existência de relações. A partir da análise realizada, entende-se que os métodos de cálculos e verificação de relação utilizados deixaram de fora pontos de estrutura sintática do texto.

De acordo com a pesquisa de SOUZA, R. (2016) - a respeito da ferramenta *Sentiwordnet* utilizada para apresentar os classificadores de polaridade e alterar de inglês

as palavras de textos escritos para português - mostra que o texto foi traduzido automaticamente e, apesar disso, houve um bom desempenho diante dos classificadores. A pesquisa também descreve comparações com outras ferramentas como SentiLex e destaca inúmeras vantagens no *SentiWordNet* em consultas que envolvem grande número de palavras. Isso demonstra a qualidade dos algoritmos e mostra a efetividade da transcrição de texto. Foi concluído que ainda se tem erros de tradução na ferramenta *Sentwordnet*, pois a tradução é automática, no entanto a ferramenta teve um bom desempenho comparado ao *SentiLex*.

2.6 Mineração de textos

A pesquisa de AMO, S. (2003) frisa a diferença entre o significado de tarefa e técnica de mineração e explica que, a tarefa se trata de especificações que se busca nos dados, de regularidade e categoria de padrões. Já a técnica, diz respeito às especificações de cada método. A pesquisa assessora em como avaliar padrões interessantes, seja a partir de uma estrutura ou por meio de estatísticas. E explica a mineração com restrições nos padrões de maneira bem técnica porém tangível.

3. Metodologia

Para o desenvolvimento inicial da aplicação foi necessário percorrer etapas predecessoras, tal como a abstração dos dados que são utilizados na análise de sentimentos. A aplicação trabalha com dados reais, tanto de opiniões de usuários de redes sociais quanto de notícias sobre o investimento escolhido.

A API do Twitter pode ser usada para recuperar e analisar dados, bem como interagir com a conversa no Twitter.

Para a análise de sentimento ser realizada, primeiro, os dados retornados pela API do Twitter ou *Web Crawler* devem passar pela fase de pré-processamento na qual são realizados alguns métodos na base, responsáveis por retirar as palavras e símbolos que não contém significado absoluto para o algoritmo, tal como as *stopword*.

Após retirar os caracteres especiais que não influenciam diretamente na abstração de sentimento, mediante as frases analisadas, é então utilizado o algoritmo de taxa de polaridade, fornecido pela biblioteca *TextBlob*.

Após a geração das taxas de polaridade retornadas pela API do *TextBlob*, as mesmas são representadas de forma gráfica em um dashboard, em diferentes contextos e visões, que tem como objetivo prover uma apresentação detalhada do cenário do investimento escolhido com relação a opiniões expostas na Internet e na mídia.

3.1 Twitter API

A API Twitter pode ser usada para recuperar e analisar dados programaticamente, bem como interagir com a conversa no Twitter. Esta API fornece acesso a uma variedade de recursos diferentes, tal como *Tweets*, *Users*, *Direct Messages*, *Lists*, *Trends*, *Média* e *Places*.

Para apreender os dados das opiniões do twitter, o recurso utilizado foi referente a parte de *tweets*, com o sub-recurso “*api.search*”, que tem como funcionalidade a construção de uma consulta específica por tema da publicação. Já no respectivo lado da

aplicação foi necessário a utilização da biblioteca *Tweepy*, que realiza uma interface facilitadora com a API Twitter. O Twitter gera, por meio de uma subscrição em seu portal de desenvolvedores, as variáveis de autenticação para utilização de sua API, tais como o *access token* e o *Auth Handler*. Após o acesso a essas duas variáveis autenticadoras, basta a criação da requisição e construção da consulta especificando alguns atributos, tal como o recurso da API utilizado, linguagem requerida dos tweets, quantidade de publicação requeridas e tema.

3.2 Web Crawler

Além dos dados do Twitter, a aplicação realiza a análise de sentimentos também em notícias, estas adquiridas por meio de um método de *web crawler*, utilizando a biblioteca *Beautiful Soap* (biblioteca que realiza uma cópia de uma página *.html* específica, passada na requisição, providenciando seus conteúdos no retorno do método).

3.3 Pré-Processamento

Para a análise de sentimentos ser realizada utilizando os dados retornados pela API Twitter, e também os dados retornados pelo *Web Crawler*, primeiro, esses dados devem passar pela fase de pré processamento, são realizados alguns métodos na base, responsáveis por retirar as palavras e símbolos que não contém significado absoluto para o algoritmo, tal como as *stopwords* (“palavra parada” em livre tradução, consideradas irrelevantes, geralmente compostas por conectivos), a partir da retirada de termos irrelevantes, são realizados os outros procedimentos comuns referentes ao contexto de Processamento de Linguagem Natural, tal como a Normalização (Processo no qual estão abarcados ajustes como a tokenização, que abstrai as palavras como token no texto, e demarca as sentenças), para tal é utilizada a plataforma líder para a construção de programas Python para trabalhar com dados de linguagem humana: NLTK (Kit de ferramentas para processamento de linguagem natural), que por sua vez oferece um conjunto de bibliotecas de processamento de texto para classificação, tokenização, análise e raciocínio semântico.

3.4 Resumo do texto

Este passo de processamento é algo exclusivo da feature relacionada a análise de sentimentos sob notícias de um investimento, na qual as notícias são deslocadas para a aplicação por meio do *web crawler*.

Como o conteúdo das notícias adquiridas, são consideravelmente maiores do que simples frases, tal como são os textos adquiridos pela *Twitter API*, se faz necessário a adoção de um método de sumarização de textos, ou resumo de textos, providenciado como uma possibilidade do Processamento de Linguagem Natural.

Para implementação do algoritmo de resumo, as sentenças devem passar pelo procedimento de limpeza e tokenização, assim como na *feature* de análise dos tweets, porém neste cenário, deve-se capturar as sentenças mais importantes do texto, utilizando de um algoritmo de frequência, de modo a capturar as sentenças mais importantes e agrupá-las, para gerar de fato um resumo contendo as ideias principais do texto.

3.5 Influência da Polaridade

A partir do pré-processamento, as informações processadas são armazenadas e então utilizadas no processo de identificação de sentimentos. Geralmente, as abordagens de modelagem para análise de sentimento dependem de dicionários, como as bibliotecas *Text Blob*, *Spacy*, *NLTK* ou modelos treinados em resultados para um domínio específico.

Para esta análise de sentimentos foi utilizada a biblioteca *TextBlob* responsável por retornar um método de polaridade bastante coerente, na qual refere-se ao positivo e negativo com taxas de texto de entrada de -1 até +1, quando o resultado de polaridade é mais próximo ao -1 e menor que -0.3, significa que a opinião tem algum tom negativo, se o resultado da polaridade se encontra entre ser maior que -0.3 e menor que +0.3 tem-se que é uma opinião de teor neutro, que não expressa diretamente ou explicitamente uma opinião que tende a ser considerada negativa ou positiva, já quando o resultado se dá em ser maior que +0.3 trata-se de uma opinião variavelmente positiva.

A abstração para correlação e métrica das opiniões analisadas foi realizada por meio do algoritmo de classificação *Naive Bayes*, no qual desconsidera a correlação entre as variáveis tratando cada sentença de forma independente, consulta em seu dicionário de dados, o algoritmo verifica se uma sentença está de acordo com o contexto de palavras “positivas” ou “negativas” e a partir disso infere uma pontuação de polaridade para cada palavra, no final classificando a frase e texto como um compilado geral das pontuações decorrentes das sentenças.

3.6 Geração dos gráficos

A partir dos resultados da análise de sentimento, foram gerados gráficos para compor um dashboard, facilitando o entendimento da informação que poderá ser consumida pelos os usuários do sistema. Os gráficos então variam de gráficos de barras contabilizadores de diferentes tipos de opiniões ou frequência de termos ou hashtags no caso do fluxo mediante a análise de sentimentos de opiniões do Twitter, até *word clouds* (gráfico de nuvem de palavra), que visam expressar de forma visual os termos mais utilizados. Todos estes cenários visam facilitar a compreensão do usuário final referente ao investimento analisado, pela mídia ou por outros usuários.

4. Resultados

4.1 Construção do Dashboard

A partir da tradução dos sentimentos, os tweets positivos, negativos e neutros podem ser agrupados e então representações dos dados podem ser geradas, para implementação de um relatório de análise de sentimento referente a um investimento específico escolhido pelo usuário. Esse Dashboard, é criado a partir da geração dos gráficos com base nos tweets já pré-processados e polarizados (aplicado o algoritmo de análise de sentimentos), utilizando bibliotecas de ciências de dados do *Python*, tal como a *matplotlib* (biblioteca que gera representações gráficas por meio de expressões matemáticas), *seaborn* (uma biblioteca para visualização de dados baseada na *matplotlib*) e *wordcloud* (tal como o próprio nome diz, é uma biblioteca que implementa

algoritmos de frequência para exibição de gráficos “nuvens de palavras”, que representam as sentenças mais relevantes dado os inputs de texto, neste caso os tweets e notícias analisadas).

A partir da construção dos gráficos, os mesmos são exportados na criação de um arquivo .PDF, que terá o descritivo das análises realizadas, bem como cada um dos gráficos gerados de acordo com a implementação dos algoritmos de análise de sentimentos e agrupamentos de seus resultados, gerando um relatório completo sobre análise de sentimentos da rede social Twitter, e das notícias referentes ao investimento escolhido pelo usuário.

O relatório gerado, contabiliza as seguintes informações:

- Gráfico de barras indicando a contagem dos tweets que expressam opiniões positivas, negativas e neutras, apresentados na Figura 1.
- *Word Clouds* (nuvens de palavras) indicando os termos mais utilizados, tal como na Figura 2, referentes a:
 - Tweets gerais.
 - Tweets com opiniões consideradas positivas pelo algoritmo.
 - Tweets com opiniões consideradas negativas pelo algoritmo.
- Gráficos de barras contendo as Hashtag`s mais utilizadas nos tweets analisados, apresentado na Figura 3, nos seguintes contextos:
 - Tweets gerais.
 - Tweets com opiniões consideradas positivas pelo algoritmo.
 - Tweets com opiniões consideradas negativas pelo algoritmo.
- Tabelas indicando se os textos analisados, retornados pelo *Web Crawler* contém opiniões positivas, neutras ou negativas mediante o algoritmo.

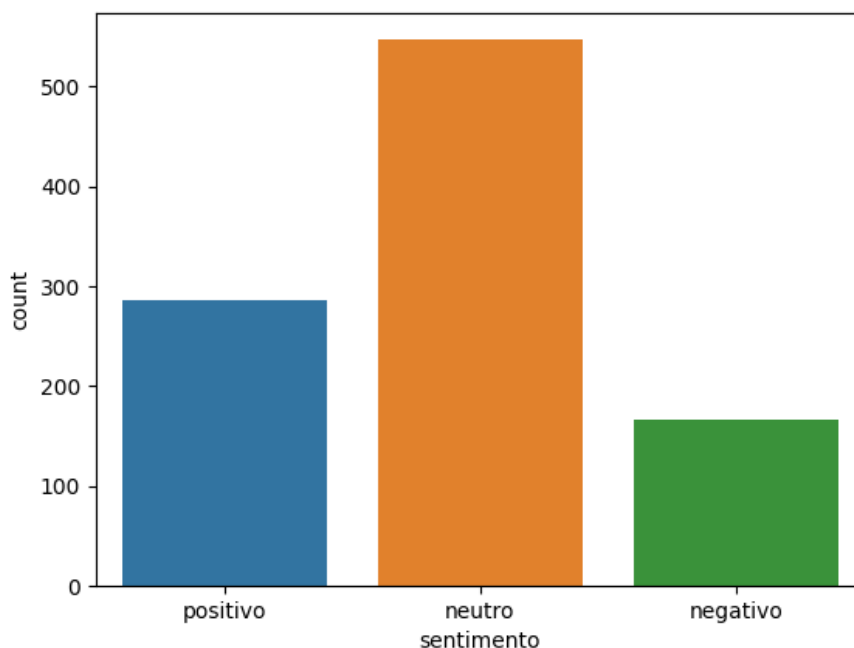


Figura 1. Gráfico contendo o agrupamento de tweets positivos/negativos sobre Bitcoin, ao analisar uma base de 1000 tweets retornados pelo Twitter API.

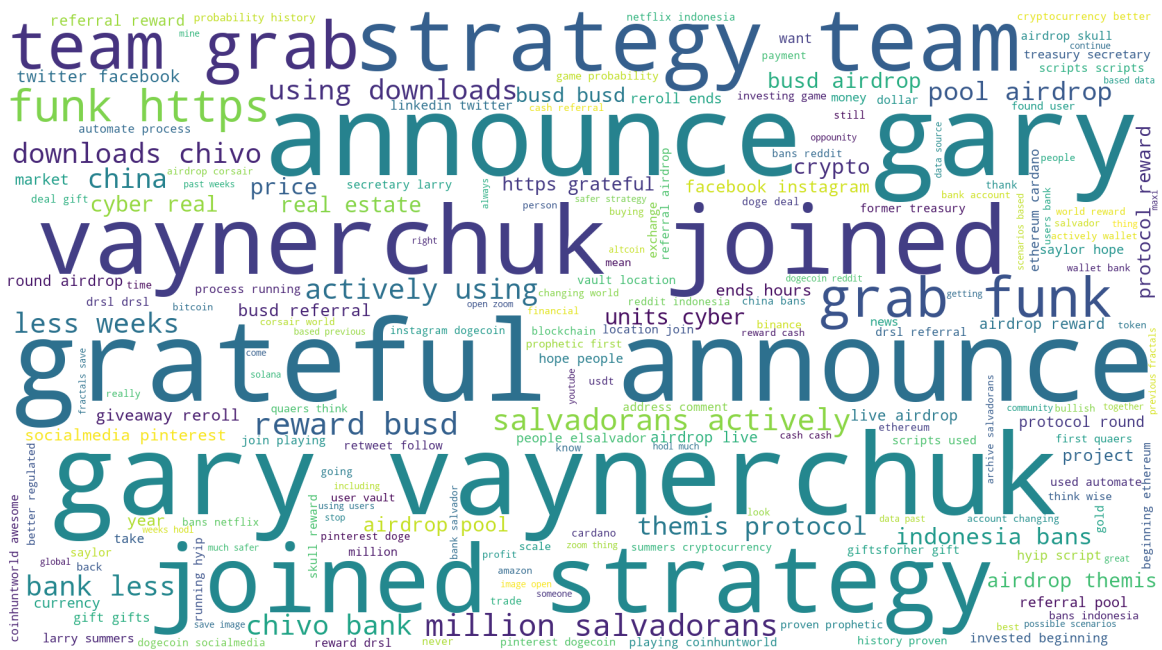


Figura 2. Wordcloud contendo as palavras mais utilizadas nos tweets relacionados à Bitcoin.

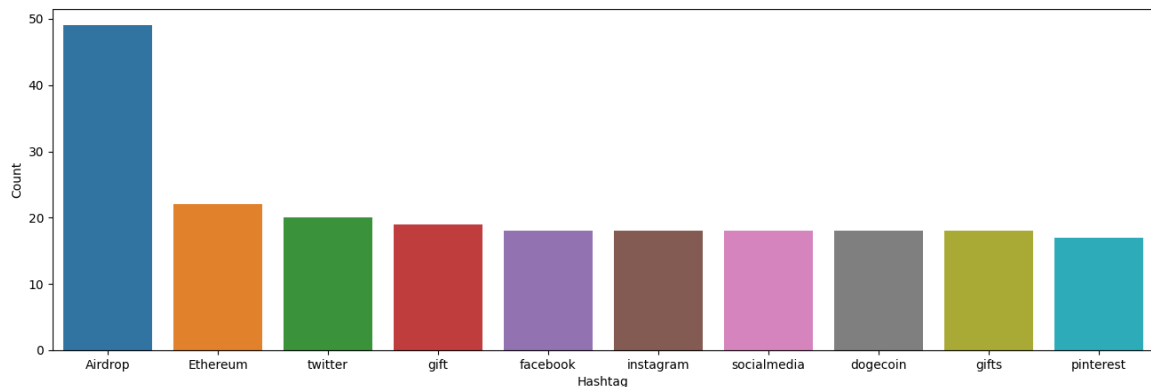


Figura 3. Gráfico de barras contendo as hashtags mais utilizadas nos tweets analisados.

4.2 Avaliação do algoritmo

Para avaliar os resultados do algoritmo, o mesmo foi executado em uma base que descrevia a sumarização e análise de sentimento dos serviços prestados pela *US Airways* (Principal companhia aérea dos Estados Unidos), isto é, tratava-se de um estudo com análise de sentimento já contabilizada, o estudo foi utilizado nesta avaliação para efeitos de comparação da eficácia do algoritmo, no caso, o quanto mais próximo os resultados, maior a eficiência da aplicação em identificar sentimentos, referente ao estudo, o mesmo possui dados compostos de opiniões do Twitter.

Após a execução do algoritmo referente a base que descrevia os sentimentos dos clientes mediante os serviços prestados pela *US Airlines*, pudemos observar uma grande compatibilidade com os dados já incluídos na base referentes aos sentimentos das sentenças, para comparação. Os resultados da análise de sentimento já incluídos na base

apresentavam que 63% de seus tweets eram considerados negativos, enquanto 21% neutros e 16% positivos. Ao executar o algoritmo de taxa polaridade e contabilizar os dados na aplicação, como demonstrado na Figura 4, foi constatado que o algoritmo contabilizou os mesmos tweets como 62.94% como negativos, 20.05 neutros e 17.01% como positivos. Em comparação, gerando uma margem de erro de 1% para cada sentimento, obtendo um resultado consistente e fidedigno.

Apesar da comparação realizada ser referente a um contexto diferente do usual da aplicação, a mesma viabiliza a análise para determinar se o principal recurso está sendo adaptado corretamente, com base nos resultados obtidos, no que diz respeito a compreensão e análise eficiente de sentenças, de modo a abstrair de forma fidedigna o sentimento da mesma, logo podemos constatar mediante o desvio mínimo obtido de 1%, que a representação utilizada corresponde aos resultados reais, evidenciando a eficácia do algoritmo utilizado.

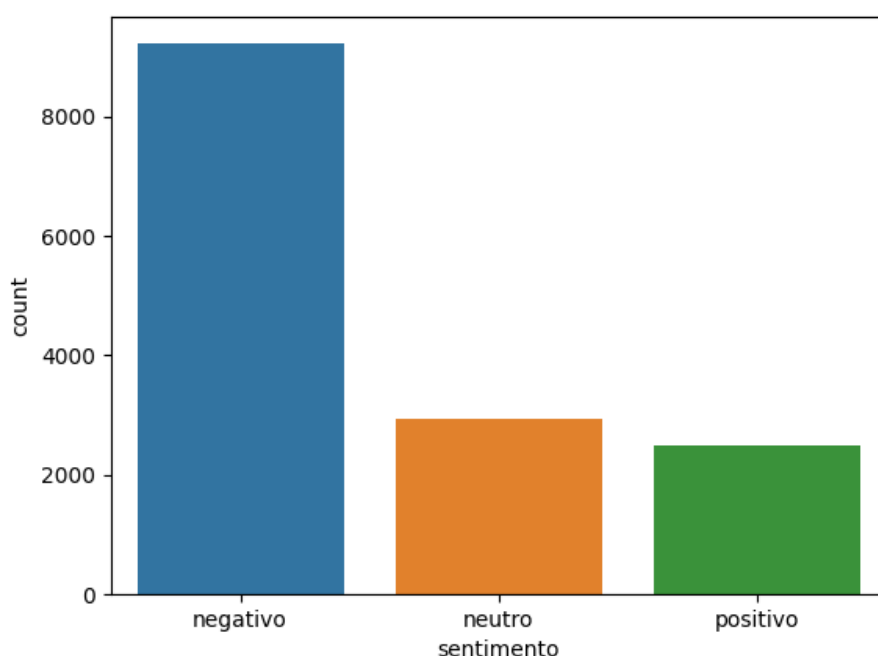


Figura 4. Gráfico de barras de análise de sentimento gerado após a execução do algoritmo de taxa de polaridade sobre a base de sentimento da US Airlines, referente a dados do Twitter.

5. Conclusões e Trabalhos Futuros

A partir da execução do principal algoritmo da aplicação em uma base que detém dados de análise de sentimento já contabilizados, ao se comparar os resultados foi verificada uma diferença mínima de 1% para cada sentimento, o que é um resultado consistente, no qual mostra a eficácia do método utilizado para análise de sentimentos.

Trazendo o método utilizado para o contexto de investimentos, o mesmo se torna um grande aliado que deve indicar com base nos sentimentos dos usuários, qual é de fato o sentimento com relação a determinado investimento, auxiliando na análise do investidor.

Apesar da comparação realizada no tópico "Resultados" ser referente a um contexto diferente do usual da aplicação, foi importante para de fato determinar se o principal recurso está sendo adaptado corretamente, no que diz respeito a compreensão e análise eficiente de sentenças, de modo a abstrair de forma fidedigna o sentimento da mesma, na qual foi constatado que está sendo realizado.

Referente ao contexto de investimentos, geralmente antes de arriscar investir em alguma ação, é realizada uma pesquisa para se saber o momento daquele investimento no mercado adequado, para entender se a compra em questão é viável e pode representar lucro, mesmo que mínimo. Uma ferramenta de análise de investimentos com base em sentimento transcrito por meio das opiniões computadas em diferentes veículos de informações da internet, como redes sociais e site de notícias, acaba informando, de forma confortável e facilitada, ao usuário o contexto e cenário daquele investimento mediante a mídia.

A proposta de solução se dá em automatizar a análise inicial do investidor, buscando informações, e as centralizando em um mesmo lugar, juntamente com a análise de opiniões baseadas em sentimentos, apresentando assim a disposição do usuário opiniões de milhares de pessoas ao redor do mundo, e as contabilizando, para se saber de forma macro, qual o sentimento geral sobre o investimento escolhido, o que geralmente influencia bastante no mercado de ações em geral, fazendo com que esta ferramenta seja de mais valia ao usuário, o auxiliando a investir e obter o melhor retorno possível.

Com relação aos trabalhos futuros, foram discutidos diversos tópicos para melhorias nas funcionalidades existentes, além de funcionalidades novas que poderiam ser apresentadas, tais como:

- Inclusão de novos tipos de investimentos possíveis para escolha do usuário da aplicação, para geração do relatório.
- Inclusão de novos gráficos, apresentando as informações de forma mais diversa, enriquecendo o relatório.
- Implementação de melhorias no *web crawler*, de modo a esta ferramenta consultar em diferentes sites referentes a um mesmo investimento, providenciando conseqüentemente visões alternadas sobre um mesmo cenário, incrementando o conteúdo das informações apresentadas.
- Visões de risco, que demonstram as vantagens e desvantagens analisadas com base nas análises realizadas, apresentando a viabilidade com relação ao investimento escolhido e possibilidade de lucro.
- Inserção de um ranking com base na opinião geral, que demonstra os investimentos que estão “em alta” na bolsa de valores.

6. Referências

AMO, S. “Técnicas de mineração de dados”. Universidade Federal de Uberlândia, Faculdade de Computação (2003). Disponível em: <https://sistemas2012.webnode.com.br/_files/200000095-bf367bfb43/Tecnicas%20de%20Minera%C3%A7%C3%A3o%20de%20Dados.pdf>. Acesso em: 24 de abril de 2021.

- ARANHA, C.N. “Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional”, Tese de Doutorado, Departamento de Engenharia Elétrica, PUC/Rio, 2007.
- ARAÚJO M., GONÇALVES P., BENEVENUTO F. “Métodos para Análise de Sentimentos no Twitter”, UFMG Belo Horizonte, Brasil, 2013. <<https://homepages.dcc.ufmg.br/~fabricio/download/webmedia13.pdf>> Acesso em: 26 de abril de 2021.
- AUGUSTO G., “Análise de Sentimento em Artefatos de Software”, Universidade do Rio Grande do Sul, Instituto de Informática, Porto Alegre, 2018. <<https://www.lume.ufrgs.br/bitstream/handle/10183/185039/001074209.pdf?sequenc e=1>>. Acesso em: 25 de abril de 2021.
- GOMES, H. J. C., “Text Mining: Análise de Sentimentos na classificação de notícias” Universidade Nova de Lisboa, Mestrado de Estatística e Gestão de Informação (2012). Disponível em: <<https://run.unl.pt/bitstream/10362/9182/1/TEGI0325.pdf>>. Acesso em: 26 de abril de 2021.
- LIU, B. HU, M. & CHENG, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. Proceedings of the 14th international conference on World Wide Web. Disponível em: <<https://www.cs.uic.edu/~liub/publications/www05-p536.pdf>>. Acesso em: 26 de abril de 2021.
- RAMOS, P. BENEVENUTO, F. “Uma Análise da Polaridade nas Manchetes de Notícias Brasileiras”, Universidade Federal de Minas Gerais (2016). Disponível em Vista do Uma Análise da Polaridade Expressa nas Manchetes de Notícias Brasileiras <<https://sol.sbc.org.br/index.php/brasnam/article/view/6454>>. Acesso em: 26 de abril de 2021.
- SOUZA, R. “Identificando sentimentos de textos em português com o sentiwordnet traduzido” Universidade Federal do Ceará, Campus Quixadá, 2016. Disponível em: <http://repositorio.ufc.br/bitstream/riufc/24824/1/2016_tcc_rcdesousa.pdf>. Acesso em: 25 de abril de 2021.
- VIEIRA, R. STRUBE, V. “Linguística computacional: princípios e aplicações”, Disponível em: <<https://www.inf.unioeste.br/~jorge/MESTRADOS/LETRAS%20-%20MECANISMOS%20DO%20FUNCIONAMENTO%20DA%20LINGUAGEM%20-%20PROCESSAMENTO%20DA%20LINGUAGEM%20NATURAL/ARTIGOS%20INTERESSANTES/lingu%EDstica%20computacional.pdf>>. Acesso em: 24 de abril de 2021.