

Construção De Uma Aplicação Para Identificar A opinião de Eleitores Brasileiros, Tendo Como Estudo De Caso, O pré-candido Jair Messias Bolsonaro.

André Felipe Gouveia Farias¹, Vinicius de Oliveira Andrade²

Curso de Bacharelado em Sistemas de Informação – Universidade Federal Rural de Pernambuco (UFRPE) – Campus Serra Talhada – PE– Brasil

andrefelipegf@gmail.com, viniciusibsm@gmail.com

Abstract. *Nowadays with a technological globalization, different communications networks are emerging every day, allowing different people from around the world to interact, generating massive amounts of data.*

Resumo. *Na época atual com uma globalização tecnológica, vêm surgindo a cada dia diferentes redes de comunicações, permitindo que diferentes pessoas de vários lugares do mundo interagem, gerando maciças quantidades de Dados.*

1. Introdução

No momento atual um dos temas mais discutidos é sobre política, pois estamos muito próximos de mais uma eleição presidencial, essas discussões estão por toda parte e um lugar onde o tema é bastante argumentado é nas redes sociais como, por exemplo, o Facebook com 117 milhões de usuários brasileiros ativos mensalmente, discutindo os mais diversos assuntos como: política, saúde, segurança e etc.

O presente artigo tem como a missão de minerar opiniões sobre o candidato a presidência do Brasil. Deputado Jair Messias Bolsonaro. a mineração dessas opiniões foram feitas na rede social chamada Facebook.

No momento presente o Facebook é a rede social mais utilizada no mundo gerando dois bilhões de usuários ativos por mês, o Brasil como já citado representa uma parcela significativa desses usuários (117 milhões) e estar em primeiro lugar como o país que mais cresce na rede social citada (Facebook). Levando em consideração o que já foi dito podemos considerar o Facebook como uma ótima rede de dados de informações sendo um ótimo lugar para minerar opiniões.

Mineração de opinião termo que já citado, seria o tratamento de um texto onde o principal objetivo seria extrair a opinião do usuário sobre um

determinado tema, essas opiniões podem ser positivas, negativas, ambas, irrelevante e irônica.

Dando continuidade o que já foi dito, o objetivo do presente trabalho, será a implementação de técnicas de mineração de opinião de usuários do Facebook e consequentemente será feita a avaliação dessas opiniões sobre o Deputado Jair Messias Bolsonaro.

2. Referencial Teórico

“Mineração da Opinião ou Análise de Sentimento (AS) é o estudo computacional de opiniões, sentimentos e emoções expressas acerca de entidades, eventos e seus atributos, que estão em um texto.”(LIU, 2010). Essas opiniões são facilmente encontradas em redes sociais como, por exemplo, Facebook ou Tweeter. Podem-se encontrar opiniões também em sites de notícias ou bloggers.

“Mineração da Opinião é o problema de identificar opiniões expressadas sobre um determinado assunto e avaliar a polaridade dessa opinião.” (TSYTSARAU; PALPANAS, 2012). As opiniões são classificadas em categorias como: negativa, positiva, neutra. Irrelevante, ambas e irônicas.

A geração das bases de dados pode ser através da API (Application Programming Interface) do Facebook e os comentários estão salvos em formatos JSON

3. Método

A Figura 3.1 Mostra as etapas que serão executadas para realizar a mineração de opinião. Cada etapa será descrita nas seções subseqüentes.

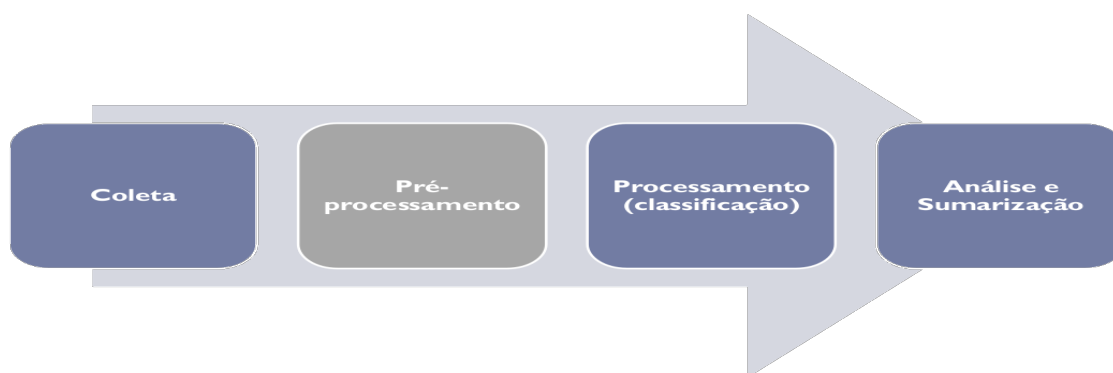


Figura 3.1 - Método proposto
Fonte: Elaborada pelo autor, 2017

3.1 Coleta dos Dados

A criação do Corpus foi realizada em três passos: extração dos comentários do Facebook, filtragem dos mesmo e anotação manual;

Faltando figura

Figura 3.2 – Processo de Coleta dos Dados.
Fonte: Elaborada pelo autor, 2017

A coleta dos comentários deu início com o uso da API do Facebook e da biblioteca Request recebendo respostas em JSON. Inicialmente foram coletados 1500 comentários onde os mesmos foram filtrados, removendo-se 679 comentários do **Facebook que poderiam ser: hastags, emotions e citações**. Os comentários removidos não exerceriam influencia na aprendizagem de máquina. Os 821 comentários do Facebook que restaram foram classificados ~~da seguinte forma~~, em seis classes: positivo que são textos subjetivos contendo apenas palavras positivas sobre o candidato, negativo que são textos subjetivos contendo apenas palavras negativas sobre o candidato, ambas que são textos subjetivos contendo palavras positivas e negativas sobre o candidato, neutra que são textos objetivos sobre o candidato, ou seja, sem qualquer opinião, irrelevante que são textos não relacionados à temática do projeto, entre outros, ironia que são textos com opinião, mas que contém emoticon ou hashtag que indica opinião contrária. com os respectivos

identificadores: 1, 2, 3, 4, 5 e 6. Para a coleção dourada final, foram **pegos** apenas os comentários classificados como positivo e negativo.

A Tabela 3.1 ilustra como ficou dividida a coleção dourada, tanto desbalanceada quanto balanceada final.

Desbalanceada						
	Positivo	Negativo	Ambas	Neutra	Irrelevante	Ironia
Quantidade	414	201	9	122	41	34
Balanceada final						
	Positivo	Negativo	Ambas	Neutra	Irrelevante	Ironia
Quantidade	201	201	0	0	0	0

Tabela 3.1 – Coleção Dourada dividida em 6 classes

3.2 Pré-processamento do Texto

Os textos expostos no **facebook** são em geral curtos e informais escritos em uma linguagem coloquial **e** em uma quantidade pequena de palavras, e de maneira desestruturadas. Os aplicativos de mineração de dados **em sua maioria** assimilam apenas textos estruturados. Para dar início aos métodos de aprendizagem precisam de textos estruturados.

Nessa etapa do processo, as técnicas para o tratamento dos comentários, foram tokenizing , remoção de caracteres especiais, remoção de stopwords, remoção de hashtags, remoção de URLs, remoção de nomes dos políticos e stemming, em seguida junto da sua polaridade, serão convertidos para que se torne possível aplicar o aprendizado de máquina.

A aplicação foi desenvolvida em Python3, para a etapa de pré-processamento foram utilizadas as bibliotecas Unicode Data, Regular Expression e NLTK as quais disponibilizam vários recursos para esse processo.

3.2.1 Tokenização

É o processo de quebra de texto em palavras ou sentenças, para isto foi utilizado o modulo da biblioteca NLTK, TweetTokenizer, por se encaixar melhor com o estilo de tokenização que era necessário para a pesquisa.

3.2.2 Remoção de Caracteres Especiais, Emotions, URL, Bolsonaro e Hashtags

Para esta tarefa foram utilizadas expressões regulares disponíveis no Python3 para a remoção dos artigos desnecessários, assim provendo de um corpus limpo e livres de dados desnecessários.

3.2.3 Stemming

Para esta tarefa foi aplicada a função de stem da biblioteca NLTK para transformação das palavras do corpus em seus radicais.

3.3 Processamento

Nessa etapa será definida a polaridade do texto analisado, utilizando algoritmo de aprendizagem de máquina.

4. Resultado

O resultado será obtido através da análise dos dois tipos de pré-processamento, por Frequência e TF-IDF

6. Referências Bibliográficas

LIU, B. Sentiment Analysis and Subjectivity. **Handbook of Natural Language Processing**, n. 1, p. 1–38, 2010.

TSYTSAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, v. 24, n. 3, p. 478–514, 2012.