



**UNIVERSIDADE FEDERAL DO TOCANTINS  
CÂMPUS UNIVERSITÁRIO DE PALMAS  
CURSO DE CIÊNCIA DA COMPUTAÇÃO  
RELATÓRIO DE APRENDIZADO DE MÁQUINA**

**ANÁLISE DE SENTIMENTOS EM DADOS DO TWITTER**

**DANIEL J. GUIMARÃES**

**PALMAS (TO)**

**2016**

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>3</b>
<b>2</b>	<b>METODOLOGIA . . . . .</b>	<b>4</b>
<b>2.1</b>	<b>Dados . . . . .</b>	<b>4</b>
<b>2.2</b>	<b>Bibliotecas . . . . .</b>	<b>4</b>
<b>2.3</b>	<b>Auxiliares . . . . .</b>	<b>4</b>
<b>2.4</b>	<b>Análise de sentimentos . . . . .</b>	<b>5</b>
<b>2.5</b>	<b>Análise Temporal . . . . .</b>	<b>5</b>
2.5.1	Análise por termos . . . . .	5
<b>3</b>	<b>RESULTADOS . . . . .</b>	<b>6</b>
<b>3.1</b>	<b>Distribuição Temporal de Sentimentos . . . . .</b>	<b>7</b>

## 1 INTRODUÇÃO

Desde a ascensão das redes sociais, políticos têm aproveitado a plataforma para compartilhar suas opiniões, promover suas agendas e se conectar diretamente com o público. Um dos políticos mais proeminentes a utilizar essa forma de comunicação é o ex-presidente dos Estados Unidos, Donald J. Trump. Seus *tweets* foram notórios por sua natureza direta, polarizadora e muitas vezes controversa.

Este artigo se propõe a explorar os *tweets* de Donald Trump por meio de técnicas de análise de sentimentos e remoção de stopwords utilizando a biblioteca NLTK (Natural Language Toolkit). A remoção de stopwords, palavras comuns que geralmente não contribuem significativamente para o conteúdo, permite uma análise mais focada nas palavras-chave e na essência do texto.

Ao remover as stopwords dos *tweets* de Donald Trump, podemos investigar os sentimentos subjacentes em suas mensagens, identificar tópicos recorrentes e entender as tendências de sua comunicação com o público. A análise de sentimentos nos ajudará a discernir se os *tweets* foram positivos, negativos ou neutros, permitindo uma visão mais aprofundada de suas expressões e perspectivas.

Com base na remoção de *stopwords*, poderemos obter uma compreensão mais clara e concisa das mensagens transmitidas pelo ex-presidente, além de explorar a linguagem e o estilo que caracterizaram sua presença nas redes sociais. Essa análise nos permitirá uma visão mais abrangente do impacto político e social que a comunicação via Twitter teve durante seu mandato.

## 2 METODOLOGIA

A metodologia proposta consiste em uma análise dos *tweets* do ex-presidente dos Estados Unidos, Donald Trump, utilizando técnicas de processamento de linguagem natural e análise de sentimentos. O objetivo é explorar os sentimentos expressos nos *tweets*, identificar padrões temporais e investigar o impacto de termos específicos nos sentimentos associados.

### 2.1 Dados

Para a aquisição dos dados para análise dos *tweets* do ex-presidente Donald Trump, foi usado o dataset Trump Tweets, 5/4/09 - 12/5/16, que apesar do nome foi atualizado para conter dados de 2009 até 27/01/2017. Os dados são agrupados pelas seguintes colunas:

- Date - Informa a data de publicação no formato 'MM/DD/AAAA'
- Time - Informa a hora de publicação em formato 12h
- Tweet - Conteúdo do *tweet*
- Client - Dispositivo de envio do twitter
- Client Simplified - Dispositivo de envio do twitter

A planilha usada foi a *2017\_01\_28 - Trump Tweets.xlsx* que gera um dataframe com 30384 registros.

### 2.2 Bibliotecas

A primeira etapa da metodologia envolve a importação das bibliotecas necessárias, como pandas, numpy, matplotlib.pyplot e nltk. Essas bibliotecas fornecem as ferramentas essenciais para manipulação e visualização dos dados, bem como para a análise de sentimentos usando o módulo SentimentIntensityAnalyzer da biblioteca NLTK.

### 2.3 Auxiliares

Em seguida, são definidas funções auxiliares que serão utilizadas ao longo do código. A função *classify()* atribui rótulos aos sentimentos com base na polaridade do sentimento, identificando se é positivo, negativo ou neutro. Outras funções são implementadas para a extração e pré-processamento dos tweets, incluindo a remoção de URLs, caracteres especiais e stopwords, utilizando técnicas como tokenização e limpeza de texto.

## 2.4 Análise de sentimentos

A etapa seguinte consiste na análise de sentimentos propriamente dita. Os *tweets* são submetidos ao *SentimentIntensityAnalyzer*, que retorna uma pontuação composta de sentimento para cada tweet. Com base nessa pontuação, é criada uma coluna de polaridade no DataFrame, atribuindo valores numéricos para representar os sentimentos expressos em cada tweet.

## 2.5 Análise Temporal

A análise temporal dos sentimentos é realizada agrupando os *tweets* por ano. São calculadas as médias das polaridades para cada sentimento (positivo, neutro e negativo) em cada ano. Essas médias são utilizadas para criar séries temporais, que são plotadas em um gráfico utilizando a biblioteca `matplotlib.pyplot`. Esse gráfico permite visualizar as variações dos sentimentos ao longo do tempo, identificando possíveis tendências ou eventos significativos.

### 2.5.1 Análise por termos

Além disso, a metodologia inclui uma análise específica de termos de interesse nos *tweets*. A função `search_in_tweets()` é utilizada para contar a quantidade de vezes que determinados termos aparecem nos *tweets* e identificar o tipo de sentimento associado a esses termos. Uma análise adicional é realizada separando os *tweets* por ano, utilizando a função `search_in_tweets_by_year()`, o que permite investigar possíveis variações dos sentimentos em relação a termos específicos ao longo dos anos.

Por fim, o código exibe informações adicionais, como a porcentagem de *tweets* positivos, neutros e negativos em relação ao total de *tweets*. Também são realizadas pesquisas adicionais nos *tweets* relacionados a palavras-chave específicas, como "russia" e "putin", buscando compreender os sentimentos associados a esses termos.

### 3 RESULTADOS

A análise dos tweets de Donald Trump por meio de técnicas de processamento de linguagem natural e análise de sentimentos revelou insights valiosos sobre os sentimentos expressos pelo ex-presidente ao longo do tempo, bem como o impacto de termos específicos em suas mensagens. As saídas obtidas forneceram uma visão abrangente da polaridade dos tweets, a frequência de ocorrência de determinados termos e os sentimentos associados a eles.

Ao analisar a porcentagem de tweets negativos, positivos e neutros, foi observado que a maioria dos tweets de Donald Trump possui uma conotação positiva, representando aproximadamente 56.76% do total de tweets analisados. Por outro lado, os tweets com polaridade negativa corresponderam a cerca de 19.06% do total, enquanto os tweets neutros representaram aproximadamente 24.18% do conjunto de dados. Esses resultados indicam uma tendência geral de expressão de sentimentos positivos em seus tweets.

Além disso, foram realizadas análises específicas para termos como "Hillary", "México", "Rússia" e "Putin". Ao examinar a ocorrência da palavra "Hillary" nos tweets de 2016, constatou-se que ela apareceu 555 vezes nesse período. Os sentimentos associados aos tweets mencionando "Hillary" foram variados, com 243 tweets expressando polaridade negativa, 169 tweets com polaridade positiva e 165 tweets classificados como neutros.

No caso do termo "México" nos tweets de 2017, verificou-se que ele apareceu 9 vezes nesse período. Os sentimentos associados aos tweets mencionando "México" foram diversificados, com 4 tweets classificados como neutros, 3 tweets com polaridade negativa e 2 tweets com polaridade positiva.

A análise dos termos "Rússia" e "Putin" revelou um maior número de ocorrências nos tweets. A palavra "Rússia" apareceu 120 vezes, enquanto "Putin" foi mencionado 66 vezes. Ao examinar os sentimentos associados a esses termos, observou-se uma distribuição mais equilibrada. No caso de "Rússia", foram encontrados 50 tweets com polaridade positiva, 49 tweets com polaridade negativa e 26 tweets classificados como neutros. Já para "Putin", os sentimentos associados foram mais variados, com 30 tweets positivos, 17 tweets neutros e 17 tweets negativos.

Esses resultados evidenciam a importância de considerar a contextualização dos termos para compreender adequadamente os sentimentos expressos nos tweets de Donald Trump. As análises realizadas demonstram a capacidade das técnicas de processamento de linguagem natural em extrair informações valiosas a partir de grandes volumes de dados não estruturados, permitindo insights sobre as opiniões e atitudes expressas pelo ex-presidente por meio de sua atividade no Twitter.

É fundamental ressaltar que essa análise foi baseada em dados específicos e limitados, representados pelos tweets de Donald Trump e pelos termos selecionados. Portanto,

os resultados obtidos devem ser interpretados dentro desse contexto e não podem ser generalizados para além desse escopo.

### 3.1 Distribuição Temporal de Sentimentos

