Analyzing sentiment of tweets

By Divya Uppal

Aug 2022

Abstract

The goal of this project is to analyze tweets by top 5 leaders and prepare a dashboard analyzing sentiment associated with these tweets. For scope of the project, we will be focussing on tweets of Barack Obama, Joe Biden, Michelle Obama, Narendra Modi and Tim Cook. The project can be easily extended to include additional twitter handles and gather information from various other sources like: Instagram, Facebook, News Article.

Design

This project pulls 123,000 tweets from twitter using Tweepy API. The data is cleaned on Python code running on Amazon EC2 instance. All emoji, https,@twitterhandle name, special characters, “RT” for retweets are removed from twitter text. Sentiment of these twitter text is calculated using Rule based sentiment analyzer named Vader.

Cleaned data frame along with sentiment is downloaded as csv file. Streamlit processes this csv file and develops a web app which displays most positive sentiment tweet, most negative sentiment tweet, most retweeted tweet, favorite tweet, average sentiment associated with tweets from each leader. A word cloud is also generated displaying frequently used words in the tweets.

Data

123K Tweets were collected using Tweepy API and the data was stored in MongoDb Atlas running on cloud. Every data point has a variety of information like text, coordinates, favorite count, retweet count, etc. Data was cleaned before performing sentiment analysis using a Rule Based Sentiment Analyzer named Vader.

Algorithms

Data Collection

1. Pull data from the Tweepy API.
2. Export data to json.
3. Connect to MongoDB.
4. Export data to MongoDB.

Sentiment Analysis

1. Connect to MongoDB and pull raw data
2. Perform text cleaning by removing all emoji, https,@twitterhandle name, special characters, “RT” from text
3. Perform Sentiment Analysis using Vader
4. Store sentiment analysis score into dataframe
5. Download post processed dataframe as csv file

Create streamlit application files

1. Pull post-processed data from csv file
2. Create functions that correspond to application features
3. Update stopwords to generate word cloud for corresponding leaders

Tools

- Requests, JSON, Tweepy for data Acquisition
- Pandas for data manipulation
- Vader for sentiment analysis
- Pymongo to connect to Mongo Atlas
- Streamlit for app deployment
- SSH to connect with python code at EC2 instance