



Visual Analytics Project: Realtime Sentiment Analysis

Shohel Rana
M.Sc Student at Arcada University



Table of Contents

- ❖ Introduction
- ❖ Objective
- ❖ Dataset Reasoning
- ❖ Label Choice
- ❖ Preprocessing Steps
- ❖ ML Algorithm
- ❖ PowerBI Dashboard
- ❖ Challenges



Introduction

- ❖ Significance of Twitter in shaping global opinions.
- ❖ Main Aim: Analyze real-time Twitter data for global insights.





Objective

- ❖ Analyze thought patterns, tweet frequencies, and user engagement on Twitter.
- ❖ Launch a Flask app, work in distributed environment & build a Sentiment Analysis ML model
- ❖ Build an interactive visual dashboard using PowerBI to convey insights



Data



❖ Collected 500000 by Tweepy library

	tweet_id	user_name
1	1594755513945755648	J.E. Dyer 🍀
2	1594755515485143052	Javier Perdomo
3	1594755517574164481	Casey Reilley
4	1594755519868043264	Val Ornelas
5	1594755519981436961	Name Can't be Blank
6	1594755520526778368	The Original Johnboy usuLtRA M

user_location
None
None
Hawaii, USA
Merica
North Coast, Ohio USA
None

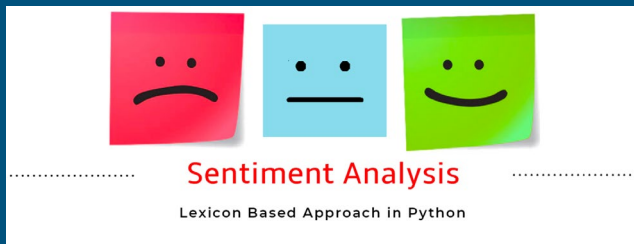
tweet_created_at
Mon Nov 21 18:11:58 +0000 2022
Mon Nov 21 18:11:58 +0000 2022
Mon Nov 21 18:11:59 +0000 2022
Mon Nov 21 18:11:59 +0000 2022
Mon Nov 21 18:11:59 +0000 2022
Mon Nov 21 18:11:59 +0000 2022
Mon Nov 21 18:11:59 +0000 2022

user_screen_name	tweet_text	tweet_retweet
OptimisticCon	Keep in mind, "Trump banned on Twitter!" is an integral element of the "J6" narrative that Trump fomented insurrect... https://t.co/ov7iRF3YSn	9016
Javierperdomo	RT @MattGertz: Elon Musk interacting with sycophantic right-wing influencers this weekend, a thread.	469
caseyreilley	RT @MattGertz: Elon Musk interacting with sycophantic right-wing influencers this weekend, a thread.	180
_surfcowgirl	RT @elizableu: I'd like to make something else clear, I don't work for Twitter, Elon Musk, any government, political party, group etc. I ru...	195
adrenaline1073	RT @disclosetv: JUST IN - Elon Musk has reinstated Rep. Marjorie Taylor Greene's (R-GA) personal Twitter account.	915
johnboy02131989	RT @BehizyTweets: BREAKING: Elon Musk just reinstated Marjorie Taylor Greene's account https://t.co/4hwc46k6Yx	203



Data Labeling

- ❖ A lexicon-based method (NLTK Vader) was used to label tweets as having a positive, neutral or negative sentiment



	tweet_text	tweet_sentiment_label
1	Keep in mind, "Trump banned on Twitter!" is an integral element of the "J6" narrative that Trump fomented insurrect... https://t.co/ov7iRF3YSn	neutral
2	RT @MattGertz: Elon Musk interacting with sycophantic right-wing influencers this weekend, a thread.	neutral
3	RT @MattGertz: Elon Musk interacting with sycophantic right-wing influencers this weekend, a thread.	neutral
4	RT @elizableu: I'd like to make something else clear, I don't work for Twitter, Elon Musk, any government, political party, group etc. I ru...	positive
5	RT @disclosetv: JUST IN - Elon Musk has reinstated Rep. Marjorie Taylor Greene's (R-GA) personal Twitter account.	neutral
6	RT @BehizyTweets: BREAKING: Elon Musk just reinstated Marjorie Taylor Greene's account https://t.co/4hwc46k6Yx	neutral



Data Preprocessing

- ❖ Remove duplicates/ missing values
- ❖ Remove URLs
- ❖ Remove special characters
- ❖ Substituting multiple spaces with single space
- ❖ Lowercase all text
- ❖ Trim the leading/trailing whitespaces



- ❖ Tokenization
- ❖ Stop word removal
- ❖ Countvectorizer & TF-IDF Vectorizer

- ❖ Jenna went back to University



Lowercase & remove non-alphanumeric

- ❖ jenna went back to university



Tokenise

- ❖ 'jenna','went','back','to','university'



Remove Stop Words

- ❖ 'jenna','went','university'



Stem/Lemmatise

- ❖ 'jenna', 'go', 'univers'

QuickSight Dashboard



- ❖ Interactive dashboard by predicted sentiment as well as by right/wrong predictions



TWITTER ANALYTICS DASHBOARD



504K

Total Tweets



7.98K

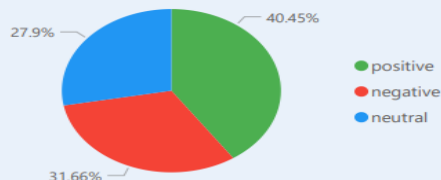
Avg Tweet Per Hour



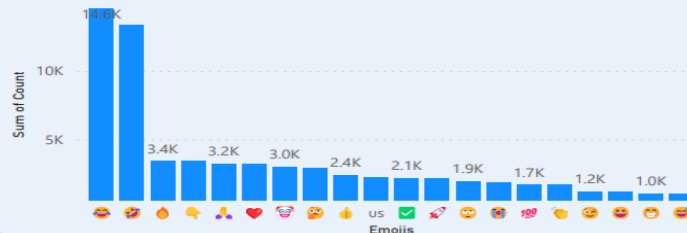
35.32

Average Engagement...

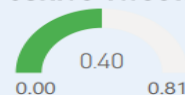
Sentimental Analysis



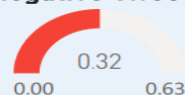
Sum of Tweet by Emojis



Positive Tweets

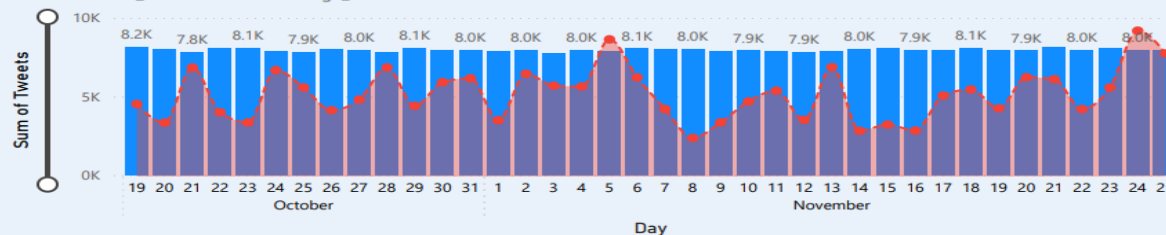


Negative Tweets



Tweets by Month and Day

● Sum of tweet_count ● Sum of average_sentiment



87.90

Average Impressions

87.90

Average Impressions

Challenges & Next Steps



- ❖ Since only recent tweets were scrapped, reply_count, retweet_count & favorite_count were not useable.
- ❖ Some of the attributes collected were not useable. Ex. 'hashtags'.
- ❖ NLTK Vader lexicon based approach may not be the best way to assign initial sentiments to tweets as it negates the emotion/ logic behind the text & focuses only on count of positive or negative words.

RT @lastbubblerrises: Elon Musk isn't even from America. He came from South Africa, a rich white Apartheid embracing family. He never invent...	positive
---	----------

- ❖ Other data labeling methods can be investigated to label the tweets more accurately.

Conclusion



- ❖ **Enhanced Understanding of Twitter Trends:** Showcased Twitter data's capacity to unveil global communication patterns and public sentiment.
- ❖ **Python & Flask in Action:** Demonstrated improvements in processing and visualizing complex datasets effectively.
- ❖ **Springboard for Future Exploration:** Laid the groundwork for further innovative studies in digital analytics and real-time data interpretation.