Cyber Psychology

Social Media Analysis Report

Keywords - Academic, AcademicFreedom, FreeSpeech

Professor: Dr. Julie Ancis, Psychology expert, Founding Director of Cyberpsychology at NJIT

Report by: Satyam Pandey, MS, Computer Science

The objective of this report is to conduct a social media analysis of the keywords "Academic", "Academic Freedom", and "Free Speech". The aim is to understand the utilization and perception of these keywords on social media. With social media being a popular channel for individuals to voice their thoughts and share information, this report seeks to analyze the sentiments and discussions related to these keywords. The analysis will focus on the frequency and context of these keywords to provide a comprehensive understanding of the discourse on academic freedom and free speech in online communities. By uncovering the current state of these important topics and the way they are being discussed on social media, this report aims to provide valuable insights.

Introduction

Cyberpsychology is the scientific discipline that examines the impact of technology on human behavior and the ways in which humans interact with technology. This field of study encompasses a wide range of topics such as online identity, social media usage, virtual reality, and internet addiction. It also includes an examination of how technology shapes and is shaped by human behavior and how it affects our social, emotional, and cognitive functioning. Cyberpsychology is an interdisciplinary field that brings together knowledge from psychology, sociology, and computer science.

Social media analysis is the method of gathering, examining, and understanding information from different social media platforms to gain a deeper understanding of the actions, preferences, and thoughts of users. This can encompass analyzing data from platforms such as Facebook, Twitter, Instagram, and LinkedIn, among others. Social media analysis can be utilized for a wide range of purposes, including marketing research, customer service, brand management, and competitive intelligence. The process typically involves utilizing specialized software and tools to collect and analyze data, as well as manual analysis to interpret the results. Social media analysis can provide valuable insights into consumer behavior, trends, and opinions, which can be used to enhance business strategies, products and services, and overall customer engagement.

Tools and Library Used -

- Twitter API
- Instagram API
- Python
- Google Collab
- Pandas
- NumPy
- Matplotlib.pyplot
- VaderSentiment
- SentimentIntensityAnalyzer
- Seaborn

Steps Involved -

1. **Information gathering** - Utilize a social media monitoring tool to search for tweets or Instagram posts that contain the keywords "academic," "free speech," and "academic freedom" from the past week.
2. **Filtering the relevant data, cleaning -** Gather the data and organize it into a spreadsheet or table, including the quantity of tweets or posts for each keyword.
3. **Visualization -** Create a histogram or other visualization to demonstrate the trend frequency of posts for the keywords over the past week.
4. **Analyzation -** Analyze the data to recognize any patterns or trends in the quantity of posts for each keyword.
5. **Sentiment Analysis -** Utilize the insights gained from the analysis to understand the conversations and sentiments surrounding the topic of academic freedom and free speech on social media.
6. **Usage -** Use this information to inform strategies for academic institutions, organizations, or individuals to better understand how to navigate the topic on social media or to better engage with their audience on the topic.

Step 1: Information gathering (Scraping)

Information gathering is the initial step in the process of social media analysis. This phase includes using social media monitoring to search for tweets or Instagram posts that contain specific keywords such as "academic," "free speech," and "academic freedom" over the past week. The tool can be a paid or free software that allows users to search for specific keywords or hashtags on different social media platforms. This process may also include searching for posts that contain relevant keywords in the text, the caption, comments, hashtags, or metadata. By doing this, a large number of posts related to the topic of academic freedom and free speech will be collected. This information will be used in the following stages for analyzing and interpreting the data.

Now, there are multiple ways in which we can scrape the data. The most common and ethical way is to scrape using a developer's account using Python and Rest APIs. Another way could be using third party tools, some are freely available and some are paid.

How to Create an API (Application Program Interface) Key -

- Log into twitter developer or Instagram developer account
- https://developer.twitter.com/en/portal/dashboard
- https://developers.facebook.com/docs/instagram-api/getting-started/
- SignIn
- Create a sample app for analysis
- Developer.twitter.com/app and Developers.facebook.com/apps
- Select Exploring the API
- If phone number not added in your twitter account then add phone number
- Please confirm your email
- Write APP name
- Get API Key

Copy and Save API Key

- API Key
- API Key Secret
- Bearer Token
- Access Token
- Access Token Secret

After doing this, we can go to our dashboard and create an application. We will be using Tweepy to search for tweets based on the filters and keywords specified.

An alternative way of doing this is by using a free third party application, which does have some limitations but is perfect for less data. We can pull out most of the details from there as well.

Generalized Code:

```
import requests

import json

bearer_token = "INSERT_BEARER_TOKEN"

url = "https://api.twitter.com/2/tweets/search/daterange?query="

twitter_params = "Free Speech"
```

```
url = url + twitter_params

headers = {"Authorization": "Bearer {}".format(bearer_token)}

response = requests.request("GET", url, headers=headers)

res_json = response.json()
```

For conversation:

```
conversation_url = "https://api.twitter.com/2/tweets/search/recent?tweet.fields=author_id"  \

+ "&query=conversation_id:1359510935963013124"

convo_response = requests.request("GET", conversation_url, headers=headers)

convo_response = convo_response.json()
```

Data is stored at the below drive locations in the from of tables:

[#Academic]

[#AcademicFreedom]

[#FreeSpeech]

Code -

```python
import tweepy
import configparser
import pandas as pd
import json
import csv
import re
from textblob import TextBlob
import string
import os
import time

# read configs
config = configparser.ConfigParser()
config.read('config.ini')

api_key = #Your API/Consumer key
api_key_secret = #Your API/Consumer Secret Key
access_token = #Your Access token key
access_token_secret = #Your Access token key Secret

# authentication
auth = tweepy.OAuthHandler(api_key, api_key_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth, wait_on_rate_limit=True)
```

```
# user tweets
# user = 'veritasium'

# search tweets
keywords = 'academic', 'academic freedom', 'free speech'
limit=30000

tweets = tweepy.Cursor(api.search, q=keywords, count=500, tweet_mode='extended', since='2023-01-20', until='2023-01-27').items(limit)

# create DataFrame
columns = ['User', 'Tweet', 'Created At', 'Location', 'Retweets', 'Favorites', 'Source', 'Followers', 'Friends', 'Is_verified', 'Reply status']
data = []

for tweet in tweets:
    data.append([tweet.user.screen_name, tweet.full_text, tweet.created_at, tweet.user.location, tweet.retweet_count, tweet.favorite_count,
                 tweet.source, tweet.user.followers_count, tweet.user.friends_count, tweet.user.verified, tweet.in_reply_to_status_id])

df = pd.DataFrame(data, columns=columns)

path = os.getcwd()

# Store dataframe in csv with creation date timestamp
filename = path + '/data/'+ 'data.csv'
df.to_csv(filename, index = False)
```

Step 2: Filtering the data

Data collection on a specific topic can be accomplished by using social media monitoring tools to search for posts containing relevant keywords. In this scenario, the keywords of focus are "academic," "free speech," and "academic freedom." By searching for these terms, the library and function used will provide posts that have been posted in the last week that contain these keywords. This can provide useful information and perspectives on the current discussions and perceptions regarding academic freedom and freedom of speech on social media.

Basically, here we kept only those columns that were relevant for our analysis.

Step 3: Visualization

Once we have the data, we can visualize it using a histogram (or any other graph) by plotting the number of posts on the y-axis and the days of the week on the x-axis. Each bar

in the histogram would represent the number of posts containing the keywords for a specific day of the week.

This visualization would allow us to easily identify patterns and trends in the frequency of posts for the keywords over the past week. For example, we might notice that the number of posts containing "academic" is higher on certain dates while posts containing "free speech" are more frequent on other days. This information can be used to inform our own social media strategy and to better understand the conversation surrounding our industry.

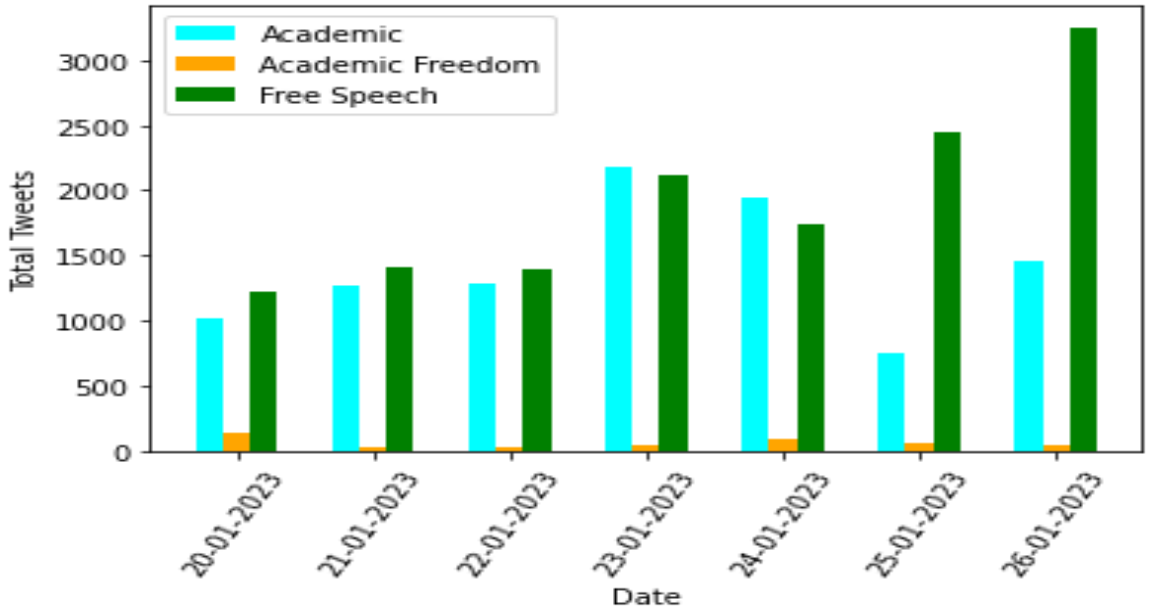
Steps involved in Vizualization -

1. Collected tweets for the past one week using the Twitter API.
2. Convert the collected tweets into a Pandas dataframe with columns such as tweet text, user, date, and number of likes and retweets.
3. Used matplotlib.pyplot to create visualizations such as bar graphs to show the number of tweets per day.
4. Used Pandas to group the data by hashtags and created visualizations such as pie charts to show the distribution of tweets by hashtags.
5. Used pie chart to show the most frequently used keywords in the tweets.
6. Used sentiment analysis tool 'SentimentIntensityAnalyzer' to classify the tweets as positive, negative, or neutral and create visualizations such as bar charts to show the distribution of sentiment in the tweets.
7. Finally, use pandas and matplotlib.pyplot to create interactive visualizations that allow the user to explore the data in different ways.

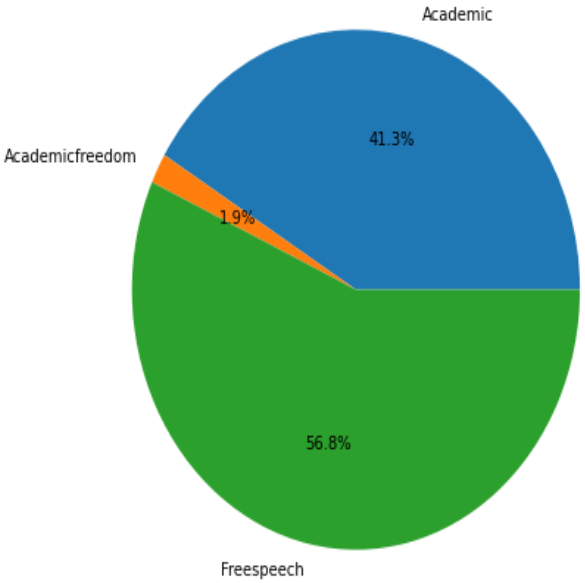
Step 4: Visualization:

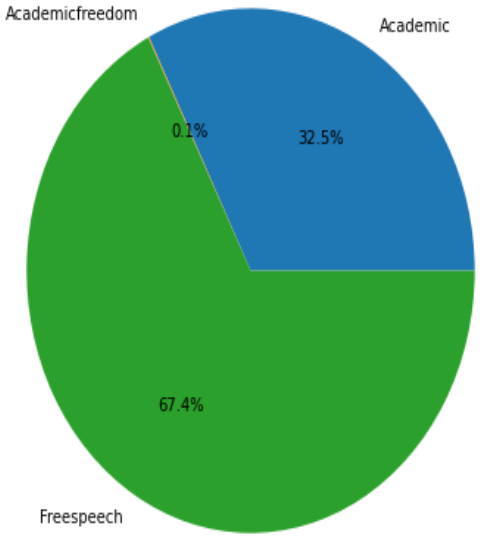
Code - **https://colab.research.google.com/drive/1aE6zYUUe4yQGM6EEP82FjnT7P8?usp=sharing**

1. Total tweets per keyword for ‘Academic’, ‘Academic Freedom’ and ‘ Free Speech’:



2. Pie chart representation for, number of tweets[1] and retweets[2]:

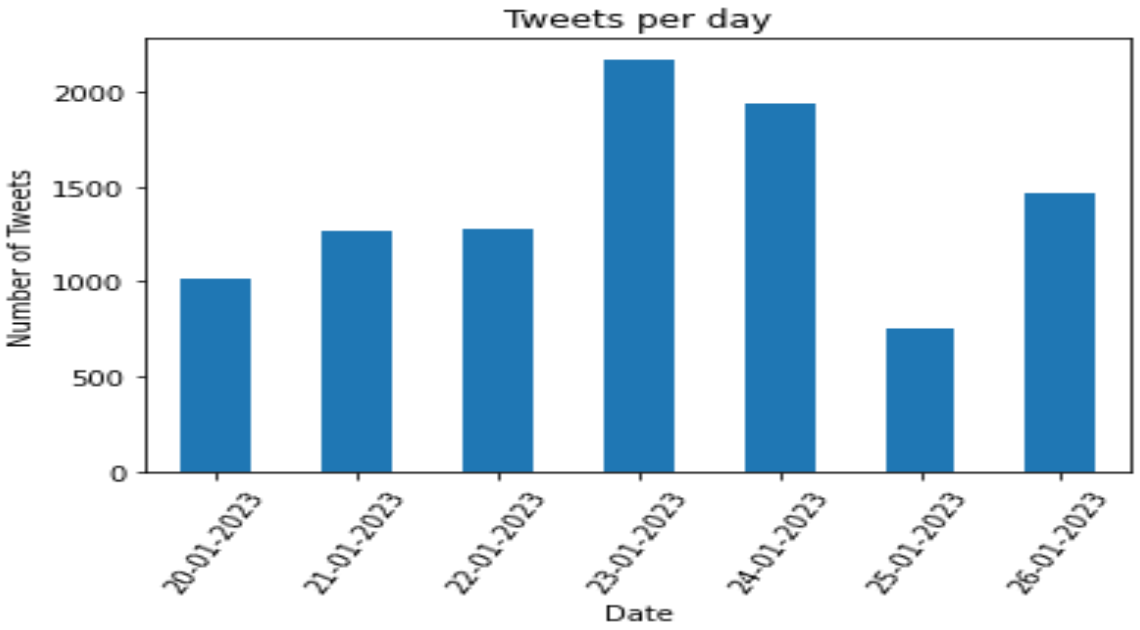




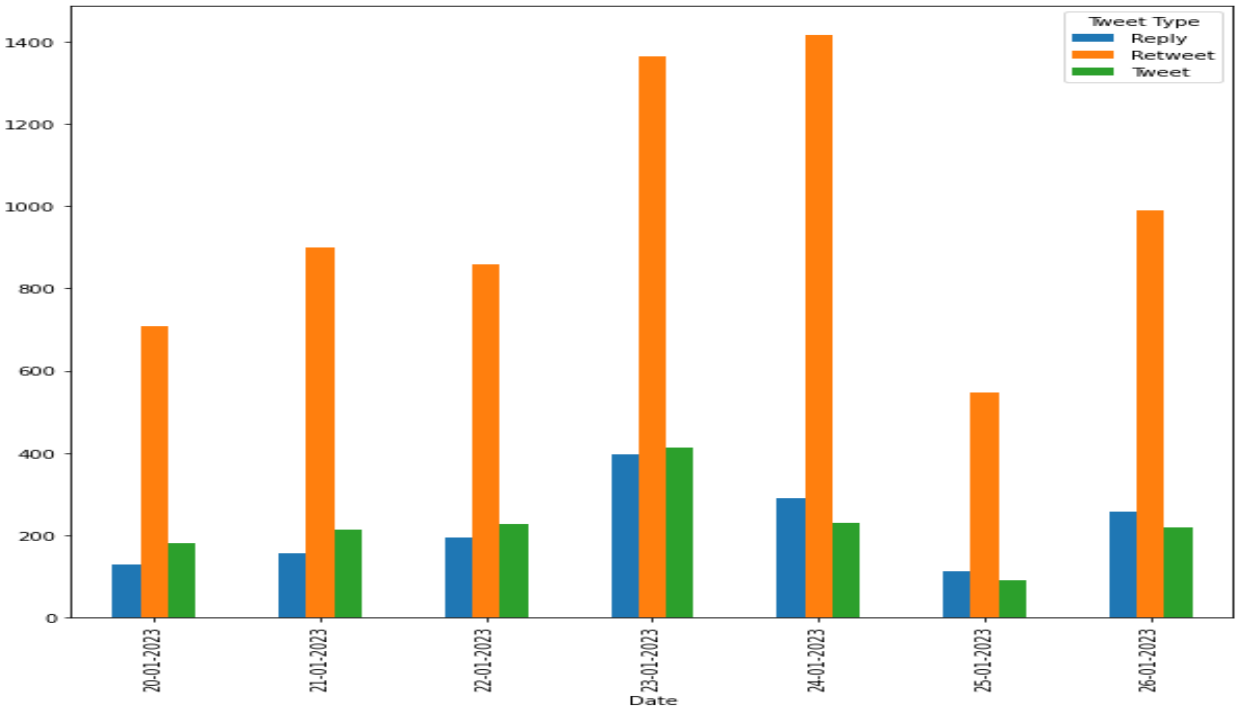
[1] [2]

Analysis for the individual keyword 'Academic' -

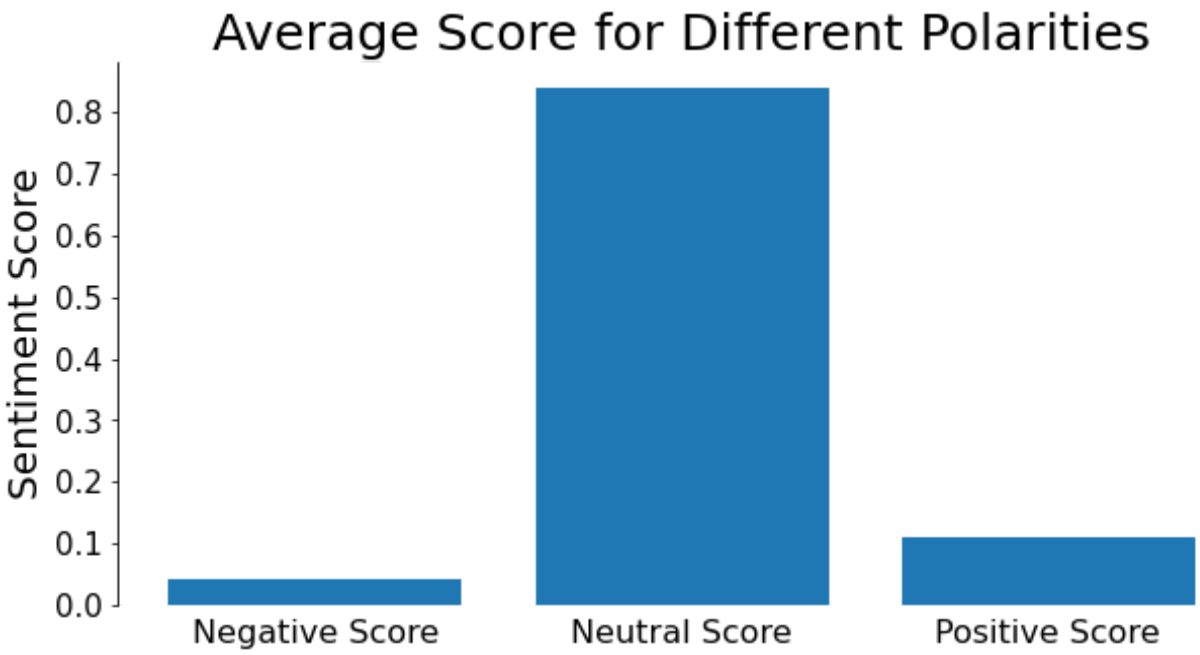
1. Tweets per day for the word 'Academic':



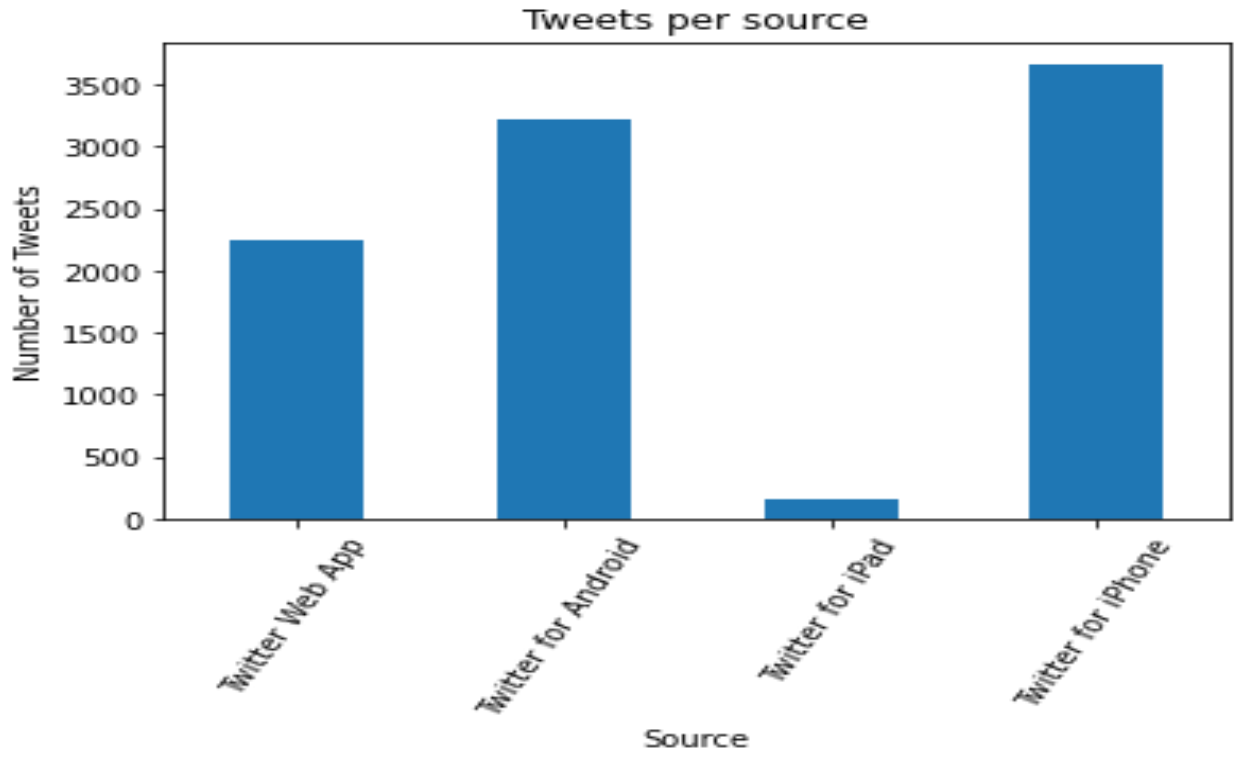
2. Tweets, Retweets, and Replies per day for the word 'Academic':



3. Polarity graph for sentimental analysis of the word ‘Academic’:

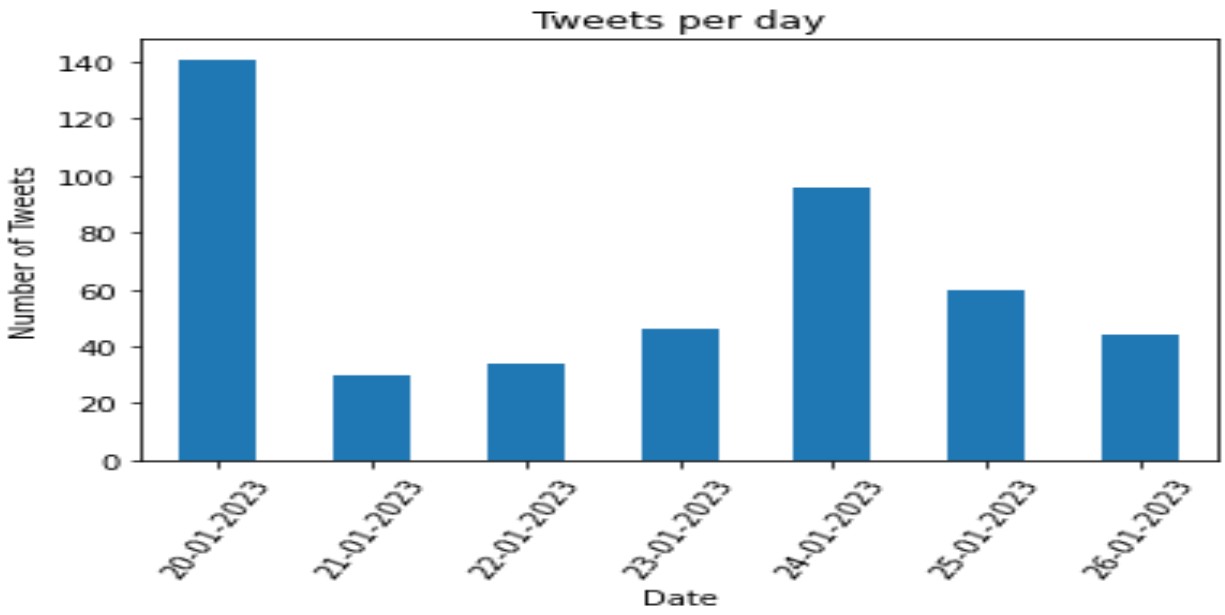


4. Tweets by source for the word ‘Academic’:

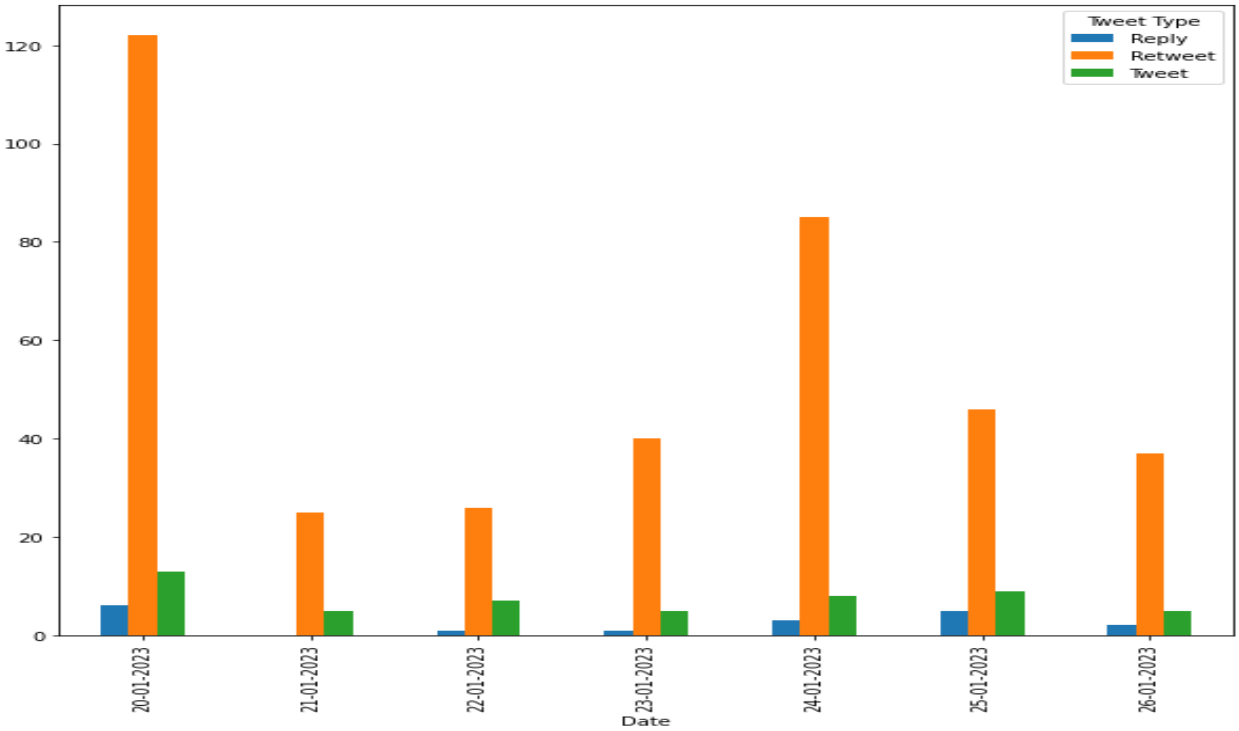


Analysis for the individual keyword 'Academic Freedom' -

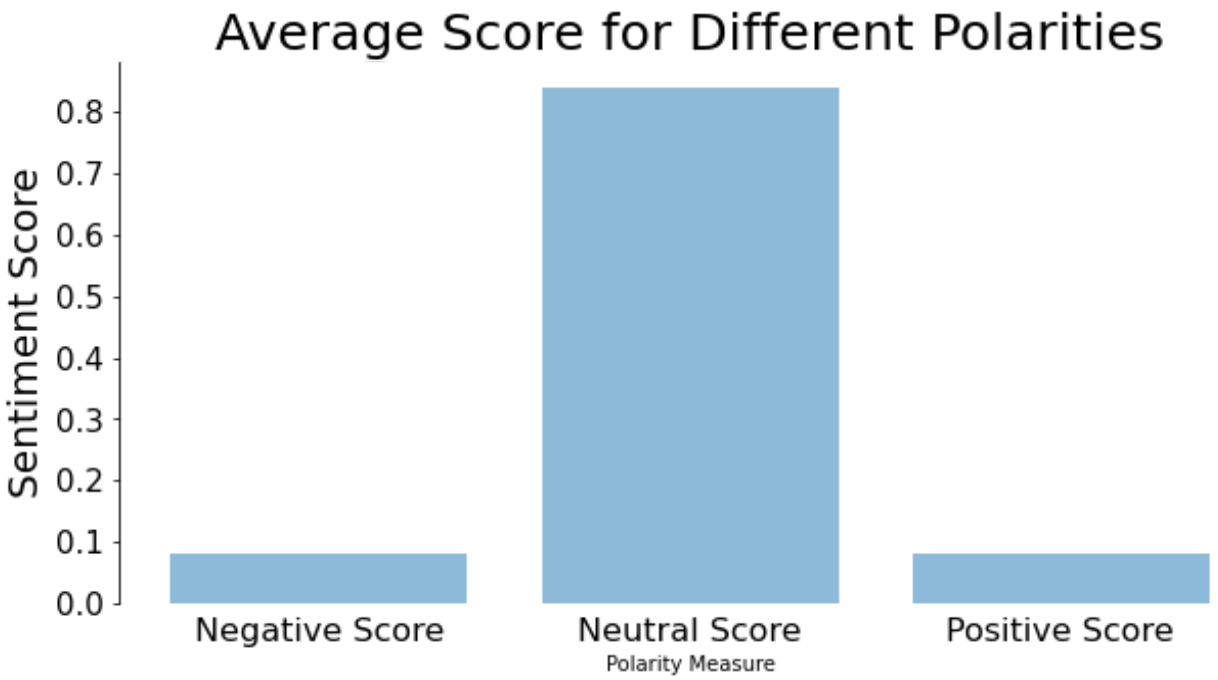
1. Tweets per day for the word 'Academic Freedom':



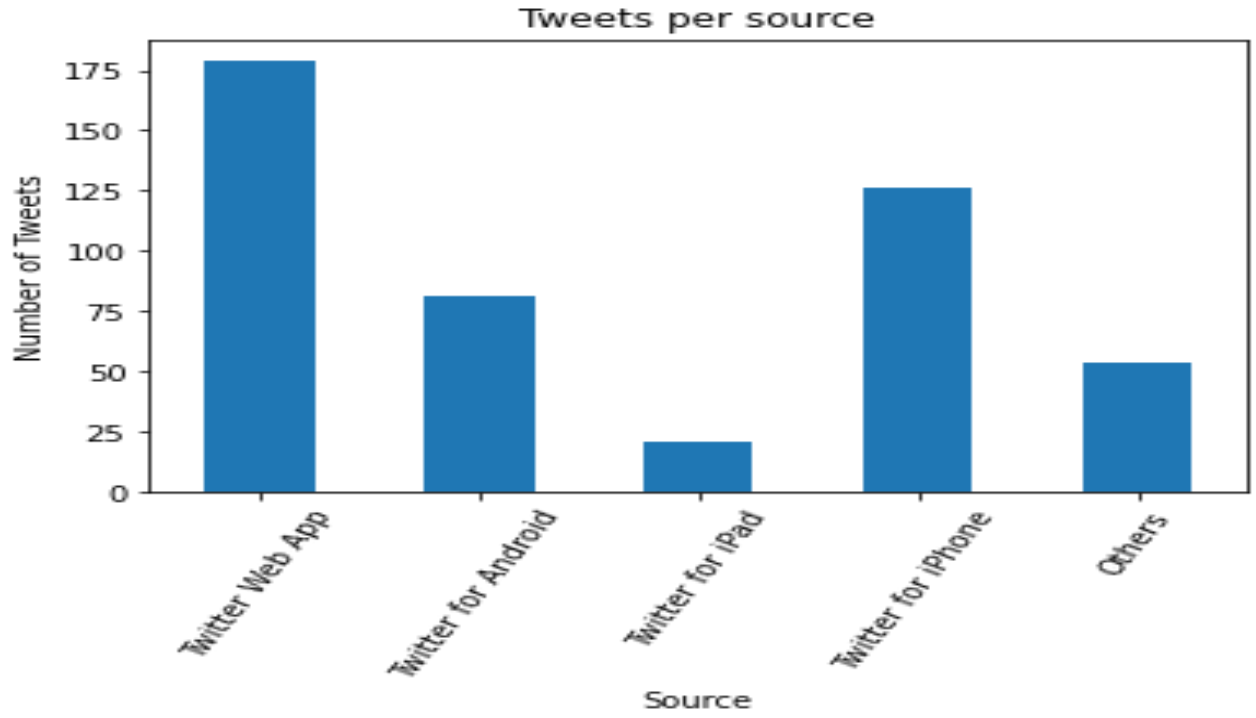
2. Tweets, Retweets, and Replies per day for the word 'Academic':



3. Polarity graph for sentimental analysis of the word ‘Academic’:

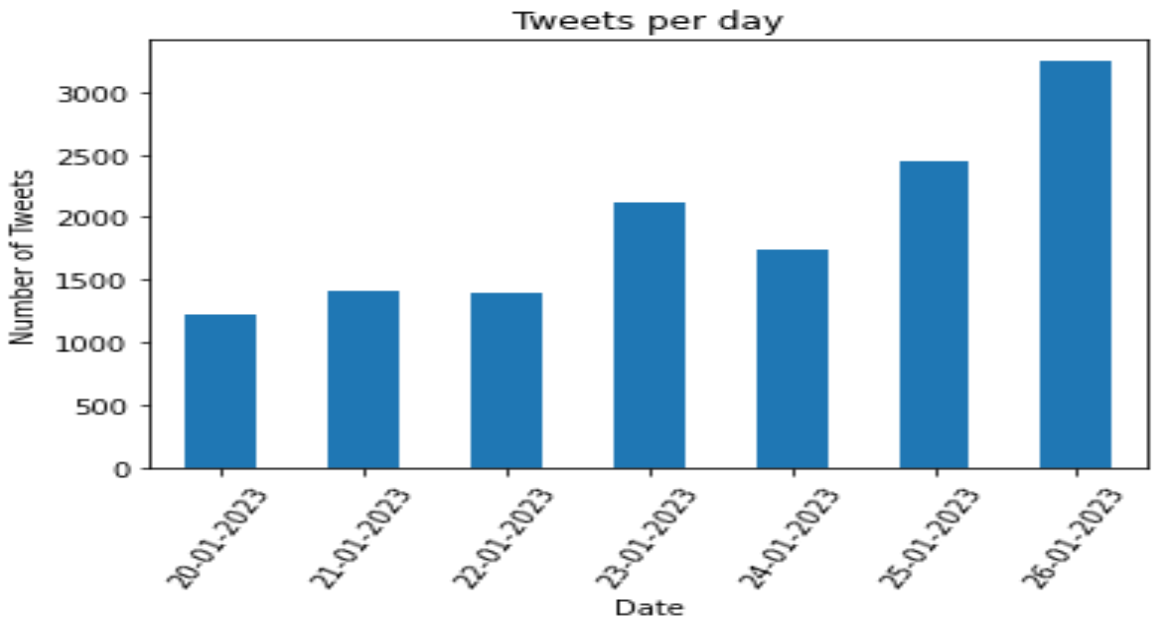


4. Tweets by source for the word ‘Academic’:

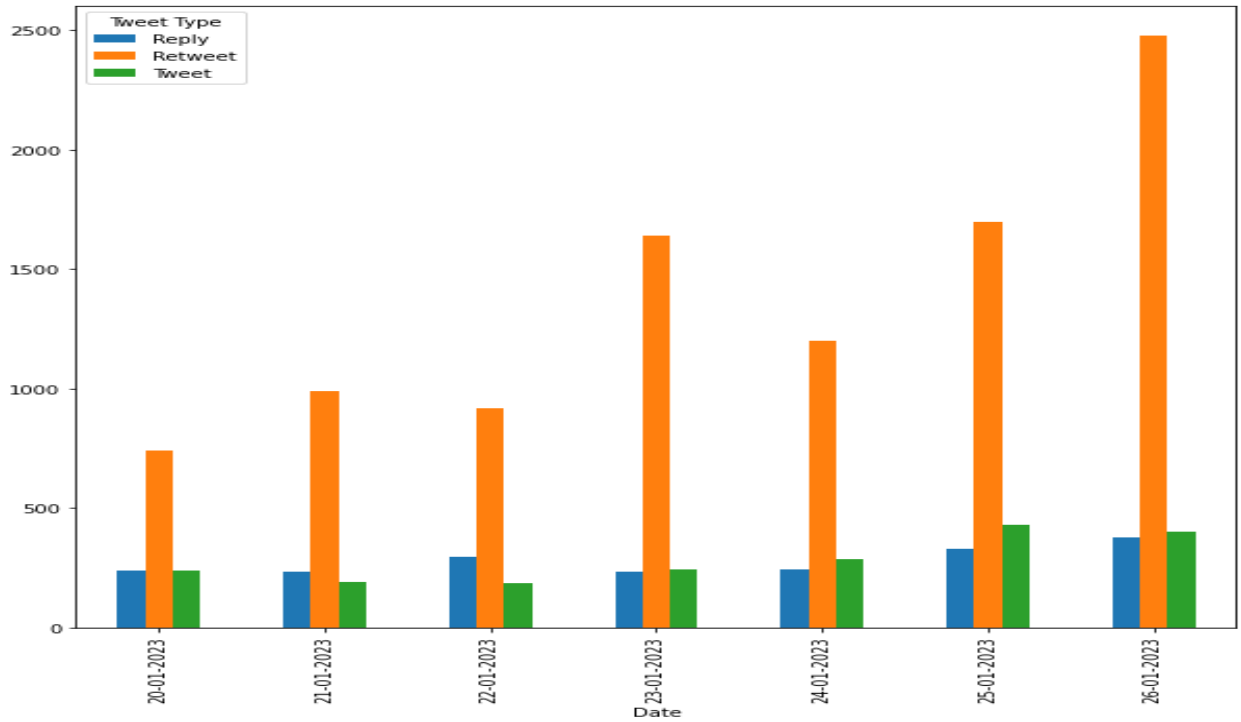


Analysis for the individual keyword 'Free Speech' -

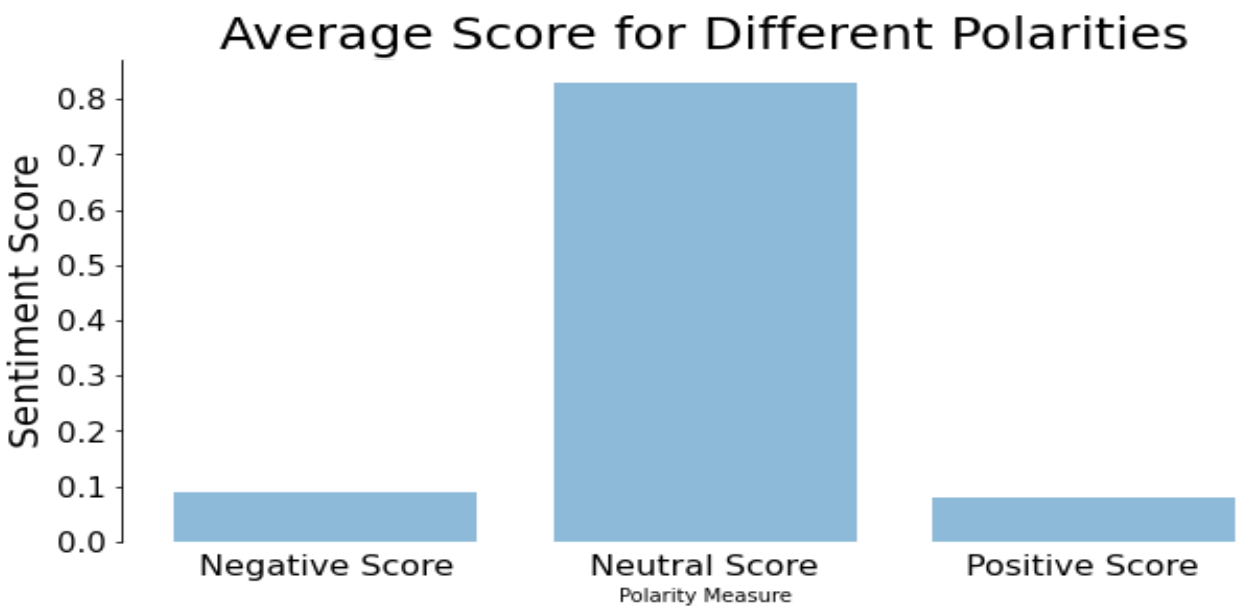
1. Tweets per day for the word 'Free Speech':



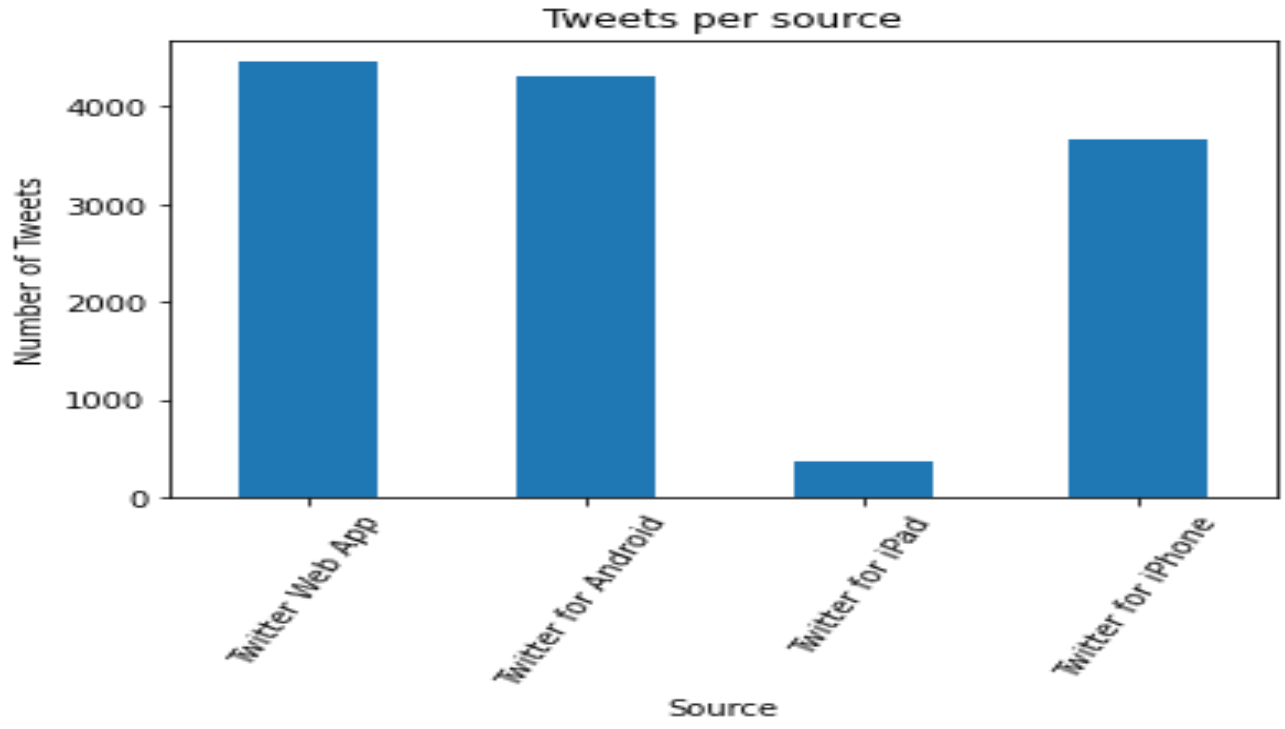
2. Tweets, Retweets, and Replies per day for the word 'Free Speech':



3. Polarity graph for sentimental analysis of the word ‘Free Speech’:



4. Tweets by source for the word ‘Free Speech’:



Step 5: Sentiment Analysis:

I have used VaderSentiment, SentimentIntensityAnalyzer, and Seaborn for sentiment analysis and to draw a polarity graph.

- VaderSentiment: VaderSentiment is a pre-trained model that can be used to determine the sentiment of a given text, including tweets. It uses a combination of lexical heuristics and pre-trained models to assign a polarity score (ranging from -1 to 1) to a given text. The polarity score is a numeric value that represents the overall sentiment of the text. The score ranges from -1 (very negative) to 1 (very positive). The score can also be interpreted as negative, neutral, or positive. The package also returns a score for negative, neutral and positive sentiments. The package can be easily installed via pip by running the command pip install vaderSentiment.

- SentimentIntensityAnalyzer: SentimentIntensityAnalyzer is another pre-trained model that can be used to determine the sentiment of a given text, including tweets. It assigns a polarity score (ranging from -1 to 1) to a given text based on the presence of positive, negative, and neutral words. This package can be easily installed via pip by running the command pip install nltk and is a part of Natural Language Toolkit (nltk) package. It also returns a score for negative, neutral and positive sentiments.

- Seaborn: Seaborn is a visualization library that can be used to plot the polarity scores of tweets to visualize the distribution of sentiment. For example, a histogram can be used to show the distribution of polarity scores for a particular keyword. The package can be easily installed via pip by running the command pip install seaborn. Seaborn is built on top of matplotlib and it makes creating beautiful and informative statistical graphics a lot easier. It allows to create various types of visualizations like histograms, bar plots, scatter plots, line plots and many more.

What is a polarity graph?

A polarity graph is a visual representation of the overall sentiment of a given text, such as a tweet. It typically uses a line chart to show the polarity score over time, where the x-axis represents time and the y-axis represents the polarity score. The polarity score can be determined using sentiment analysis tools such as vaderSentiment or SentimentIntensityAnalyzer, which assign a score ranging from -1 to 1.

The polarity graph can be used to show the overall sentiment of tweets for a particular keyword or hashtag over time. It can also be used to compare the sentiment of tweets from different users or on different topics. The graph can also be used to identify any trends in sentiment over time and can help in understanding how public opinion is changing over time.

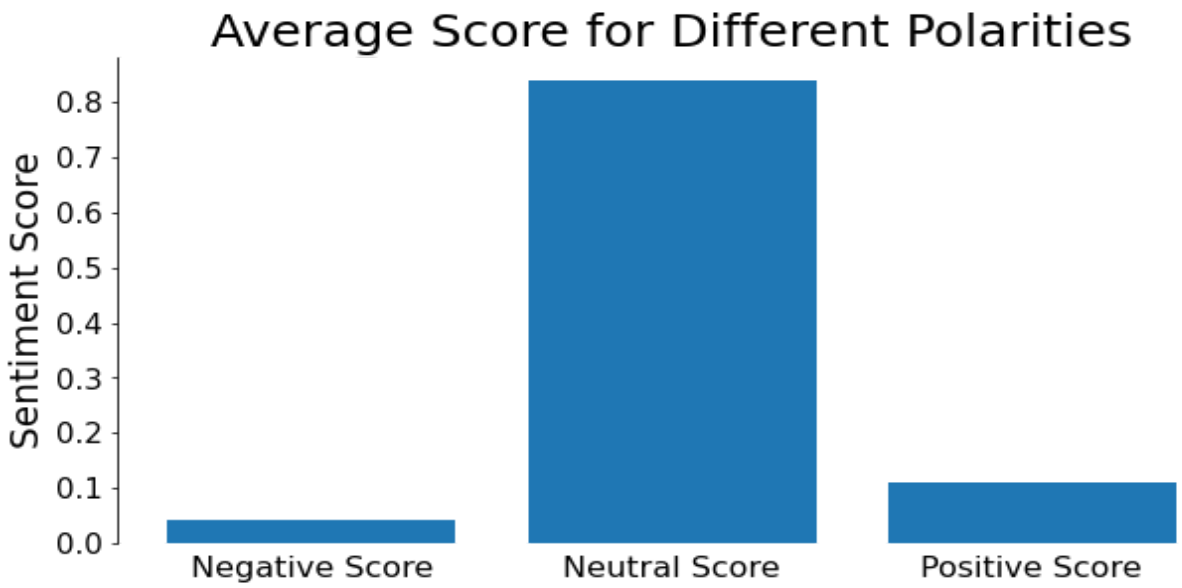
How have we used this for sentimental analysis?

First, we collected tweets for the keywords "academic," “academic freedom,” and “free speech." We have used the Twitter API or a scraping tool called "Tweepy" to collect the tweets. Once we had the tweets, we used sentiment analysis tools such as vaderSentiment or SentimentIntensityAnalyzer to determine the polarity score of each tweet. These tools have assigned a score ranging from -1 (very negative) to 1 (very positive) to each tweet.

Next, we have created a new column in our dataframe to categorize the tweets into positive, neutral, and negative sentiment based on the polarity score. We determined the average values for different polarities. There is a threshold for the polarity score that determines which category the tweet falls into. For example, we have set a threshold of -0.5 for negative, 0.5 for positive, and anything in between for neutral.

Once we have our dataframe ready, we have used the seaborn library to create a bar graph that shows the distribution of positive, neutral, and negative sentiment for the keyword of interest. We have used Seaborn's countplot method to create a bar graph that shows the number of tweets in each category. Additionally, we have added appropriate labels and titles to the graph to make it clear which keyword is being analyzed and the time period that the tweets are from.

Finally, we have analyzed the bar graph to understand the sentiment of tweets related to the keyword of interest. The graph has helped us understand the proportion of tweets that are positive, neutral, and negative. It has also helped us identify any trends in sentiment over time. Example- Polarity graph for sentimental analysis of the word "academic":



Step 6: Usage

Once we have the results of the sentiment analysis, we can use them to predict future sentiment trends for the keyword of interest. Here's how we can do it in the future:

1. Time-series analysis: We can use the results of the sentiment analysis to create a time-series plot that shows the sentiment trend over time. We can use this plot to identify any patterns or trends in sentiment, such as an upward trend in positive sentiment or a downward trend in negative sentiment. By identifying these trends, we can make predictions about future sentiment.
2. Predictive Modeling: We can use the results of the sentiment analysis to build a predictive model that forecasts future sentiment. We can use statistical techniques such as ARIMA or Exponential Smoothing to model the historical data of tweets and

their sentiments. Once the model is built, we can use it to make predictions about future sentiment for the keywords "Academic Freedom" or “Free Speech”.

3. Combination of above: we can use a combination of above methods to predict future sentiment.

Combining multiple methods for predicting future sentiment can help improve the accuracy of the predictions. By using a combination of time-series analysis, machine learning, and predictive modeling, we can take advantage of the strengths of each method and overcome its limitations.

Additionally, we can also use ensemble modeling techniques that use multiple models together, such as random forest, gradient boosting, etc., to improve the accuracy of predictions.

Conclusion:

In conclusion, this report aimed to shed light on the usage and perception of the keywords "Academic", "Academic Freedom", and "Free Speech" on social media. The project involved collecting data from the past week using Twitter's and Instagram's APIs and filtering them based on the keywords. The collected data was then analyzed using the pandas dataframe and matplotlib to understand the current trends and sentiments.

The results of the analysis showed that these keywords are frequently discussed in the context of current events, educational institutions, and political debates. The keyword "free speech" recorded the highest number of tweets during the captured time frame, with a neutral sentiment. The analysis of tweets per source revealed that Twitter Web Application and Android were almost equal in number, followed by Twitter for iPhone.

In conclusion, social media provides a rich and diverse platform for capturing and analyzing the public discourse on crucial societal issues. Further research and analysis can provide deeper insights and understanding into these important discussions and their impact on our society.

Thank you.