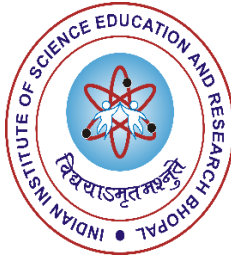


Data Science in Practice – DSE315



Final Project Report

By: Baddepudi E V N M SaiAsrith, 20074

Title: Twitter Scrapping and Twitter Sentiment Analysis

Topic: Twitter Sentiment Analysis on Ukraine-Russia war

Overview: In this report we will discuss what is sentiment analysis, it's applications in the real world and we will also see how people are reacting on Ukraine Russian war through Sentiment Analysis.

Introduction: An overview of Sentiment Analysis and its applications

- Sentiment analysis (or opinion mining) is a Natural Language Processing technique used to determine whether data is positive, negative, or neutral.
- Sentiment analysis is contextual mining of words which indicates the social sentiment of a brand and helps the business to determine whether the product which they are manufacturing is going to make a demand in the market or not.

What is Sentiment Analysis?

- Sentiment analysis is the process of detecting positive or negative sentiment in text.
- It is often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

Types of Sentiment Analysis:

- Sentiment Analysis focuses on the polarity of a text that is (positive, negative, neutral).
- Depending on how you want to interpret customer feedback and queries, you can define and tailor your categories to meet your sentiment analysis needs.
- Popular types of sentiment analysis are
 - (i) Graded Sentiment Analysis
 - (ii) Emotion detection
 - (iii) Aspect-based Sentiment Analysis
 - (iv) Multilingual Sentiment Analysis

The overall benefits of sentiment analysis:

- Sorting data at scale

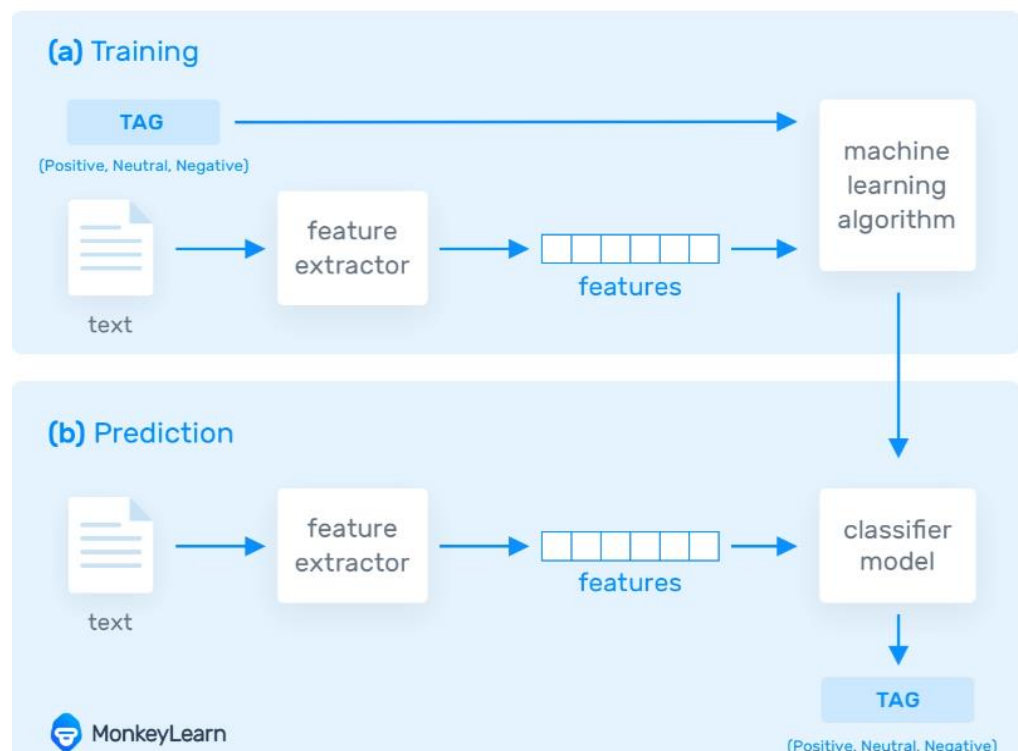
- Real time analysis
- Consistent criteria

Sentiment Analysis examples:

- (i) Netflix has the best collection of movies.
- (ii) I hate waiting for the next series to come out.
- (iii) The final episode was surprising with a terrible twist at the end (negative term used in a positive way).

How sentiment analysis work?

- Sentiment analysis work on Natural language processing (NLP) and Machine learning algorithms to determine the tone of users behind the online conversations.
- There are three types of algorithms for sentiment analysis
 - (i) Rule based
 - (ii) Automatic
 - (iii) Hybrid



- In the training process model learns to associate a particular input to the corresponding output based on the test samples

used for training where the feature vectors are fed into the machine learning algorithm.

- The feature extractor is used to transform unseen text inputs into feature vectors.
- The feature vectors are then fed into the model which generates the tags that is positive, negative, and neutral tags.
- **Classification algorithms include:**
 - (i) Naïve Bayes
 - (ii) Linear Regression
 - (iii) Support Vector Machines
 - (iv) Deep Learning

What are positive, negative, neutral in tags?

- (i) Positive Tags:
 - It depends up on the situation the tag is used.
Examples:
 - (a) Good opinion on product's come under positive tag
 - (b) If there are bad news spreading out and people are tweeting not to do such kind of works then it comes under positive tag
- (ii) Negative tags:
 - It depends up on the situation the tag is used.
Examples:
 - (a) Bad opinion on one of the company products.
- (iii) Neutral Tags:
 - Here are some ideas to help you identify and define *neutral* texts:
 - (i) **Objective texts.** So called *objective* texts do not contain explicit sentiments, so you should include those texts into the neutral category.
 - (ii) **Irrelevant information.** If you haven't pre-processed your data to filter out irrelevant information, you can tag it neutral. However, be

careful! Only do this if you know how this could affect overall performance. Sometimes, you will be adding noise to your classifier and performance could get worse.

(iii) **Texts containing wishes.** Some wishes like, *I wish the product had more integrations* are generally neutral. However, those including comparisons like, *I wish the product were better* are difficult to categorize.

Human Annotator Accuracy

- Sentiment analysis is a tremendously difficult task even for humans. On average, inter annotator agreement (a measure of how well two (or more) human labellers can make the same annotation decision) is low when it comes to sentiment analysis. And since machines learn from **labelled data**, sentiment analysis classifiers might not be as precise as other types of classifiers.
- If you are new to sentiment analysis, then you will quickly notice improvements. For typical use cases, such as ticket routing, brand monitoring, and VOC analysis, you will save a lot of time and money on tedious manual tasks.

Sentiment Analysis applications:

The Sentiment analysis applications are endless and this can be applied to any industry. Some of the popular ways that sentiment analysis is being used in business are:

- Social Media Monitoring
- Brand Monitoring
- Voice of customer
- Customer Service
- Market Research

Sentiment Analysis study from the tweets during Russian-Ukraine war conflict

In this sentiment analysis we got the scrapped data from the Kaggle so in this project we have not done twitter scrapping as data of the tweets is already available.

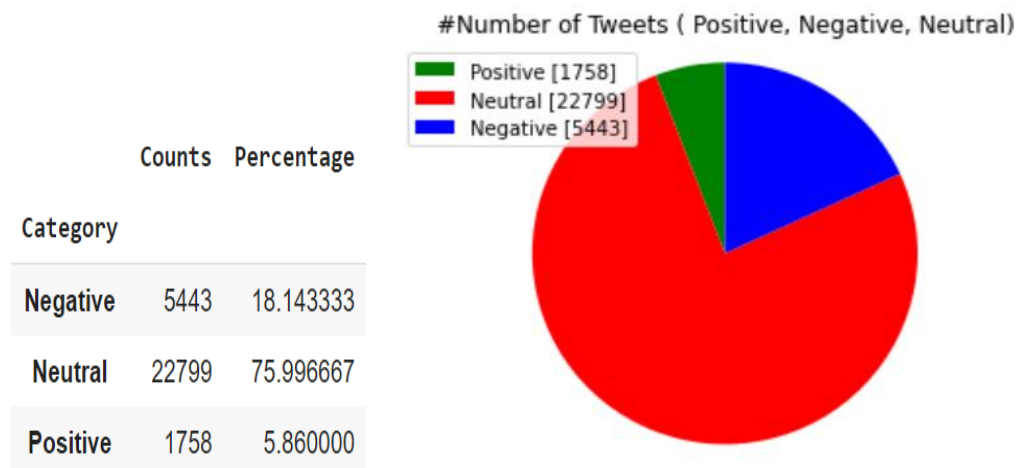
- **Aim:** To analyse how people are reacting to the Ukraine-Russian war conflict by seeing the sentiment scores from the tweets.
- **Libraries used:** NumPy, Pandas, Seaborn, Matplotlib, Nltk, Vader sentiment, sentiment intensity analyzer. (VADER stands for Valence Aware Dictionary and sentiment Reasoner. Vader performs well for the analysis of sentiments expressed in social media. These sentiments must be present in the form of comments, tweets, retweets, or post descriptions, and it works well on texts from other domains also. VADER is a lexicon and rule-based analysis tool)
- Before going to the analysis area we have to clean our data by removing symbols, spaces etc this is known as Data cleaning.

Analysis:

- In this analysis we have extracted all the languages which are used by the people while tweeting
- We have observed that there are tweets in 54 different languages.
- We have also extracted the location of the people from where they had tweeted, it has been observed that people almost across the world are tweeting about their opinion of Ukraine Russia war.
- We have also calculated the sentiment score based on the tweets (sentiment score implies the number of positive, negative, neutral tweets subject to conflict).

Observation and Visualization:

- **Visualization of Number of tweets:**



The table above indicates the positive, negative, and neutral tweets from the tweets data from the data set and the plot to the right is the pie representation of the above table. We can observe that neutral tweets are higher when compared to positive and negative tweets.

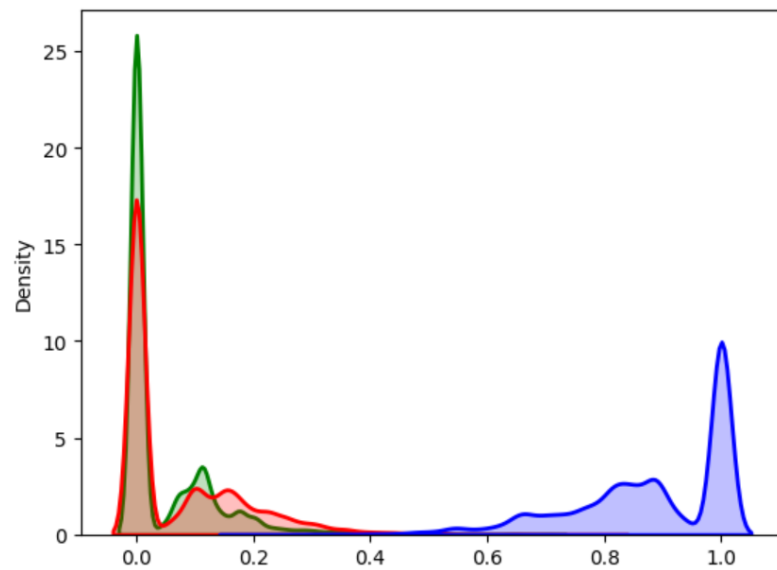
- **Visualization of Sentiment Scores:**

Sentiment scores are a metric for measuring customer sentiment. Scores can range from 0-100, where 100 is the most positive possible outcome and 0 is the least. Positive words are assigned a +1 scoring, while negative words are assigned a -1 scoring in speech analysis software. But here we are scaling from -0.5 to 0.5 were

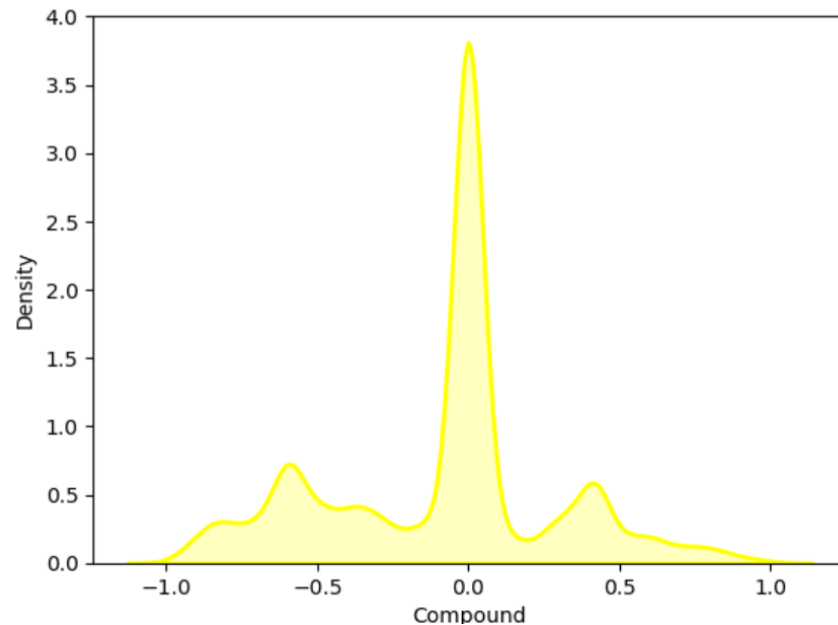
- (a) Positive score if compound value is greater than 0.5
- (b) Negative score if compound value is less than -0.5

(c) Neutral score if compound value is in between(-0.5,0.5)

<matplotlib.axes._subplots.AxesSubplot at 0x7f86003c2a50>



<matplotlib.axes._subplots.AxesSubplot at 0x7f85fdb523d0>



So, from the above plot between compound vs density, we can observe that there is a high peak between -0.5 to 0.5 which implies that neutral score is more than positive and negative score which implies that people are reacting in more neutral way to the war.

- **Top positive, negative, and neutral hashtags:**

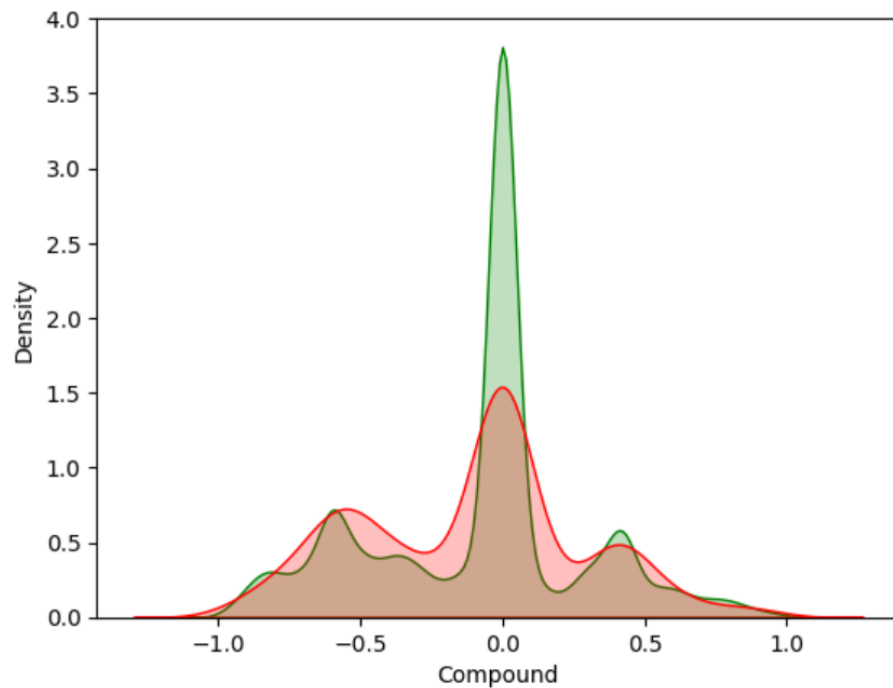

```

[
  'RussiaUkraineWar',
  'Russia',
  'Ukraine',
  'ZOG',
  'Biden',
  'Putin',
  'Russia',
  'America',
  'Ukraine',
  'RussiaUkraineWar',
  'UkraineUnderAttack',
  'SKUY',
  'SEKUYA',
  'Crypto',
  'Binance',
  'SKUYARMY',
  'Putin',
  'Russia',
  'BTC',
  'Bitcoin',
  'Metaverse',
  'RussiaUkraineWar',
  'cryptocurrency',
  'NFT',
  'Marinetti'
]
[
  'Bucha',
  'UkraIndiadiadiaeRussiaWar',
  'UkraIndiadiadiaeWar',
  'ukraIndiadiadiae',
  'RussiaUkraIndiadiadiaeWar',
  'Hostomel',
  'Russian',
  'Izyum',
  'Kharkiv',
  'UkraIndiadiadiae',
  'UkraIndiadiadiaeUnderAttack',
  'RussiaUkraIndiadiadiaeWar',
  'Bucha',
  'Mariupol',
  'Πολεμος_στην_Ουκρανια',
  'WarCrime',
  'UkraIndiadiadiae',
  'RussiaUkra',
  'RussiaUkraIndiadiadiaeWar',
  'EU',
  'Europe',
  'UkraIndiadiadiae',
  'Donbass',
  'NeoNazi',
  'nytimes',
  'wsj',
  'Europe',
  'UkraIndiadiadiae',
  'Zelensky',
  'West',
  'US',
  'EU',
  'Nato',
  'UkraIndiadiadiae',
  'Russia',
  'UkraIndiadiadiaians',
  'KremlIndiadiadia',
  'RussiaUkraIndiadiadiae',
  'RussiaUkraIndiadiadiaeWar',
  'RussiaWarCrimes',
  'Bucha',
  'RussiaUkraIndiadiadiaeWar',
  'UkraIndiadiadiaeUnderAttack',
  'GenocideOfUkraIndiadiadiaians',
  'RussiaWarCrimes',
  'Bucha',
  'RussiaUkraineWar',
  'ukraineUnderAttack'
]

```

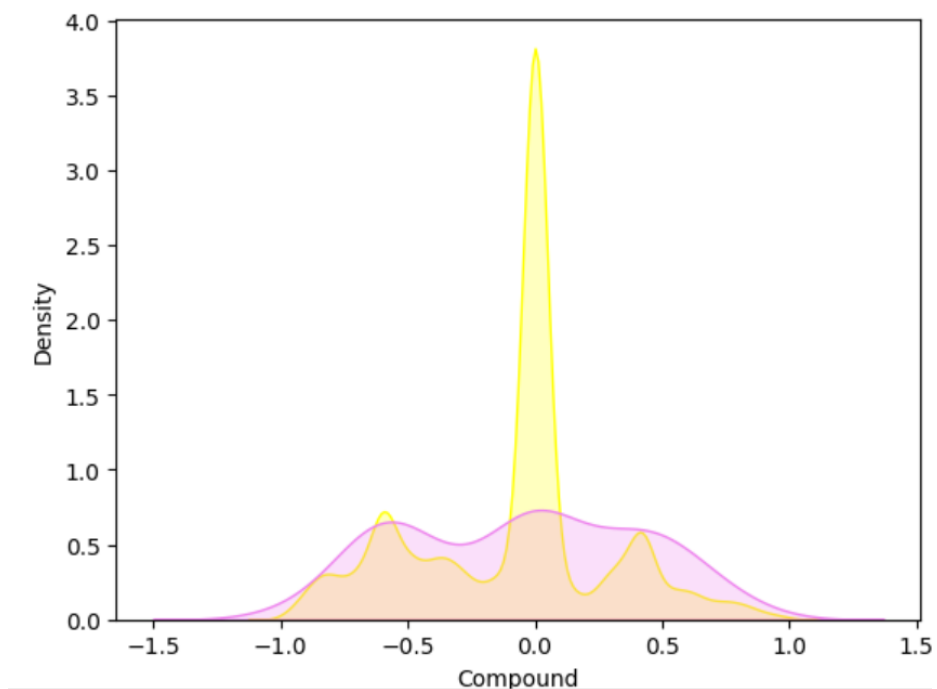
- **Comparison of tweets between Ukrainian and rest of the world**

<matplotlib.axes._subplots.AxesSubplot at 0x7f85fd991690>



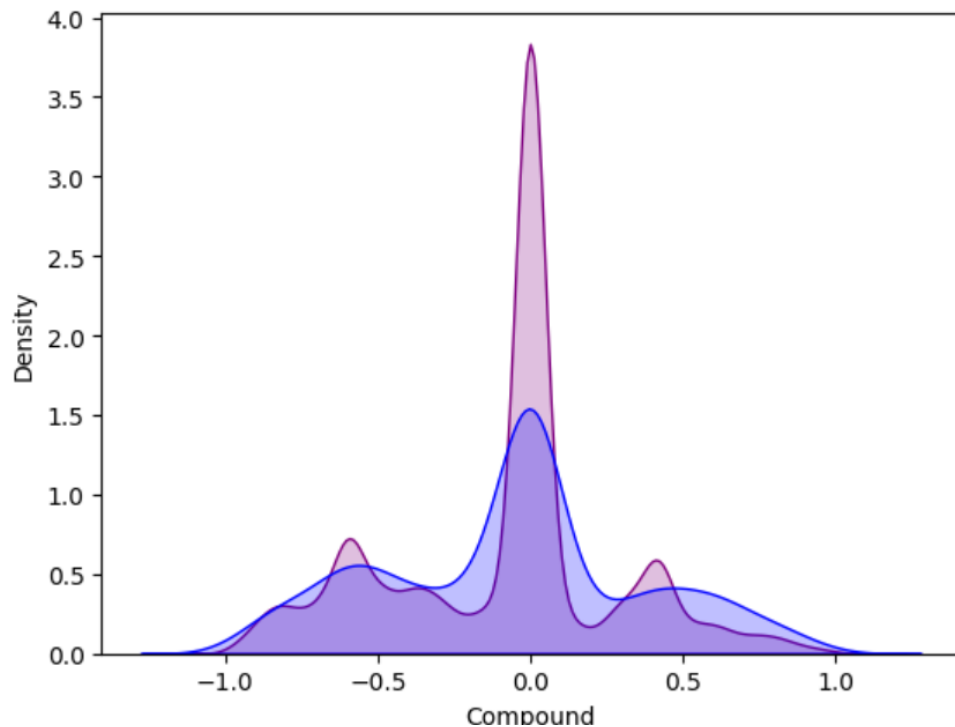
- **Comparison of tweets between the people who belong to Europe vs rest of the world**

<matplotlib.axes._subplots.AxesSubplot at 0x7f85fd7b1f90>



- **Comparison of sentiment score with us citizens vs rest of the world**

matplotlib.axes._subplots.AxesSubplot at 0x7f85fd791710>



- **Word cloud for all sentiments :**



So from the above plots we can observe that the people from all over the world reacting in an neutral way as observed by the high peak value between -0.5 to 0.5 values of compound has high density reacting to that.

Conclusion:

- So from the above analysis we came to conclusion that
 - (i) 75% of the tweets are showing neutral sentiment score implying people that were tweeting had neither a favourable or negative feeling about the war.
 - (ii) 18% of the tweets are showing negative sentiment score.
 - (iii) 5% of the tweets are showing positive sentiment score.

- So, majority of the people are neutral towards war implying that neither a favourable or negative feeling but the people who tweeted from the Ukraine we can observe that they are opposing the war as observed from red coloured area from the plot of sentiment score between the people who tweeted from Ukraine vs the rest of the world.
- So, we can conclude that majority of the people are neutral towards the war.

Resources:

<https://www.kaggle.com/code/avinandandutta/twitter-sentiment-analysis-russia-ukraine-conflict>

[https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20\(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs.](https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs.)