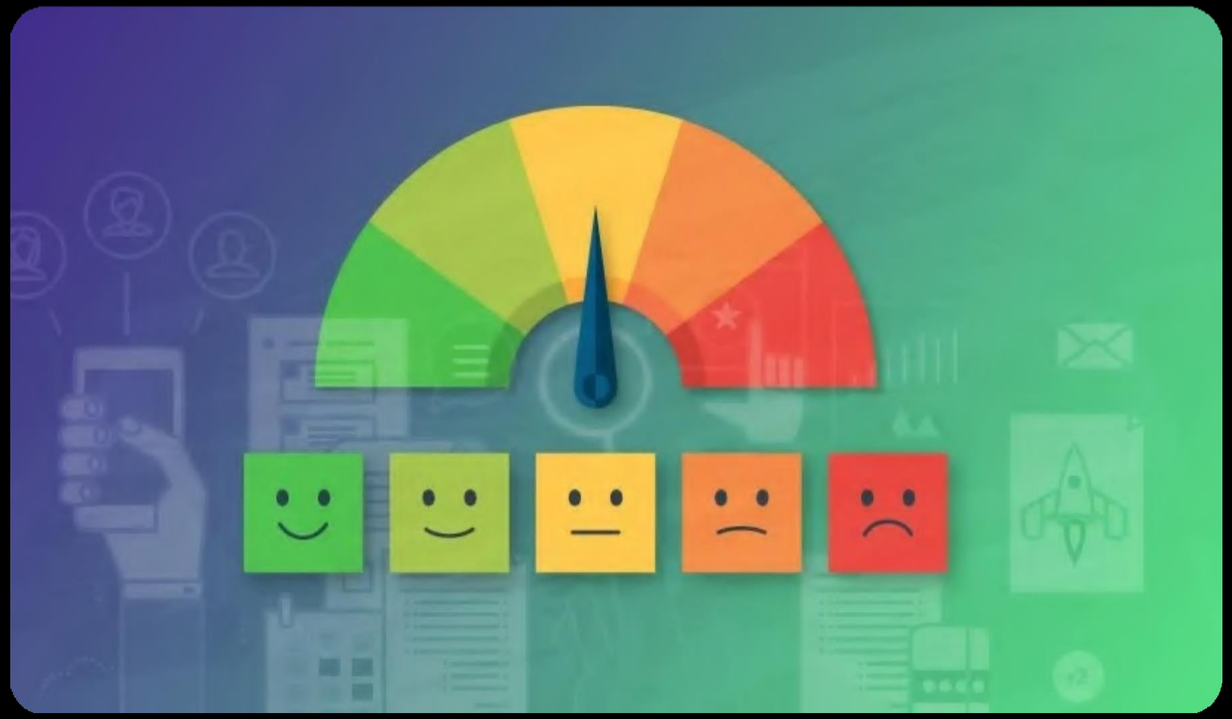
VISHNU
UNIVERSAL LEARNING

Department of
Electronics and Communication Engineering
Vishnu Institute of Technology(A),
Bhimavaram.

Emotion Analysis of Tweet



Under the Guidance of
Mr.D.Durga Prasad
Assistant Professor
Dept. of ECE

Team Members:-

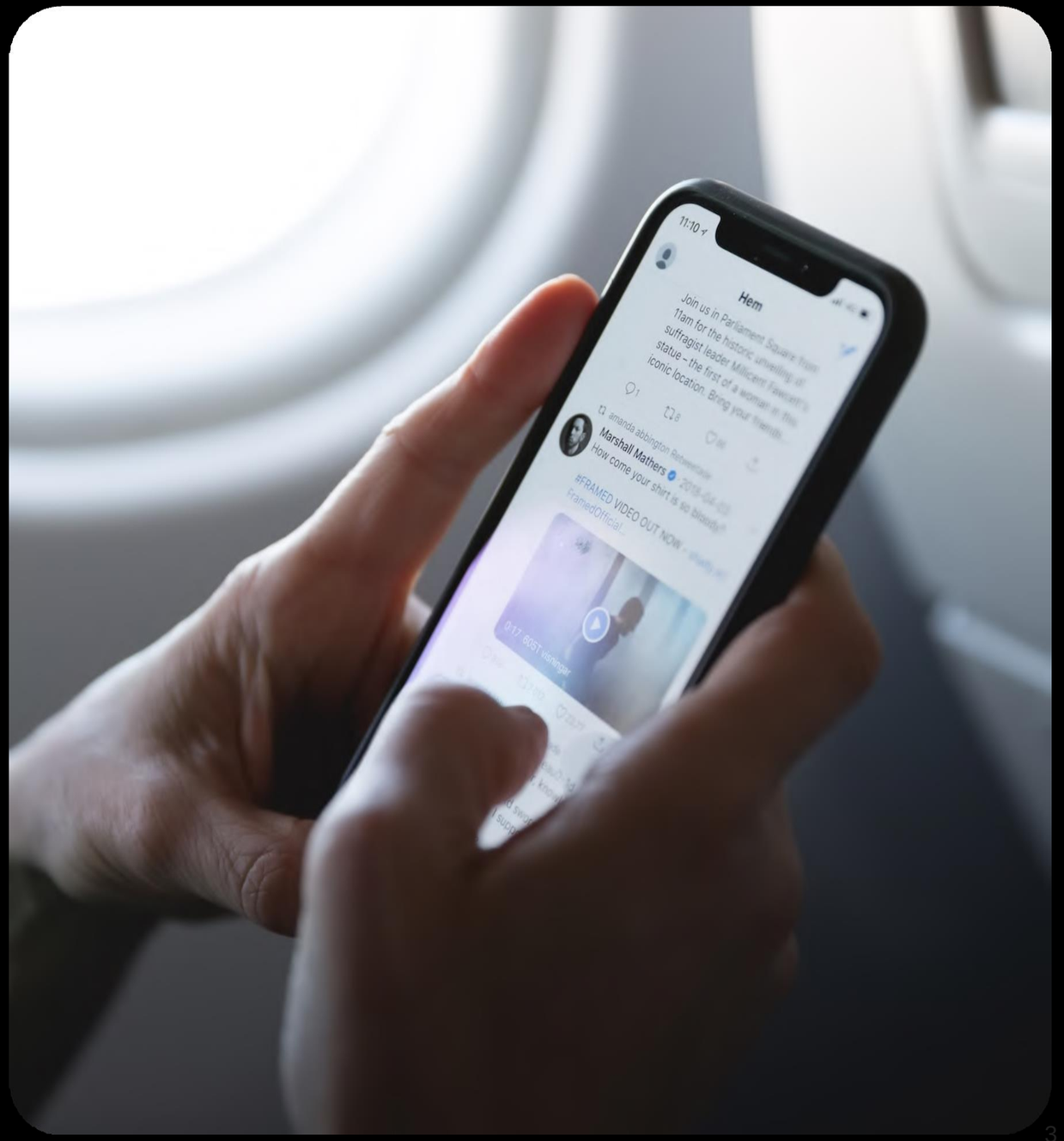
1. G.Pardhu	19pa1a04c7
2.T.Ramya Manasa	19pa1a04f5
3.V.Bavya Sri	19pa1a04h3
4.T.Prashanth	19pa1a04g1
5.P.Niranjan Sai	19pa1a04d2

Table of Contents

Abstract

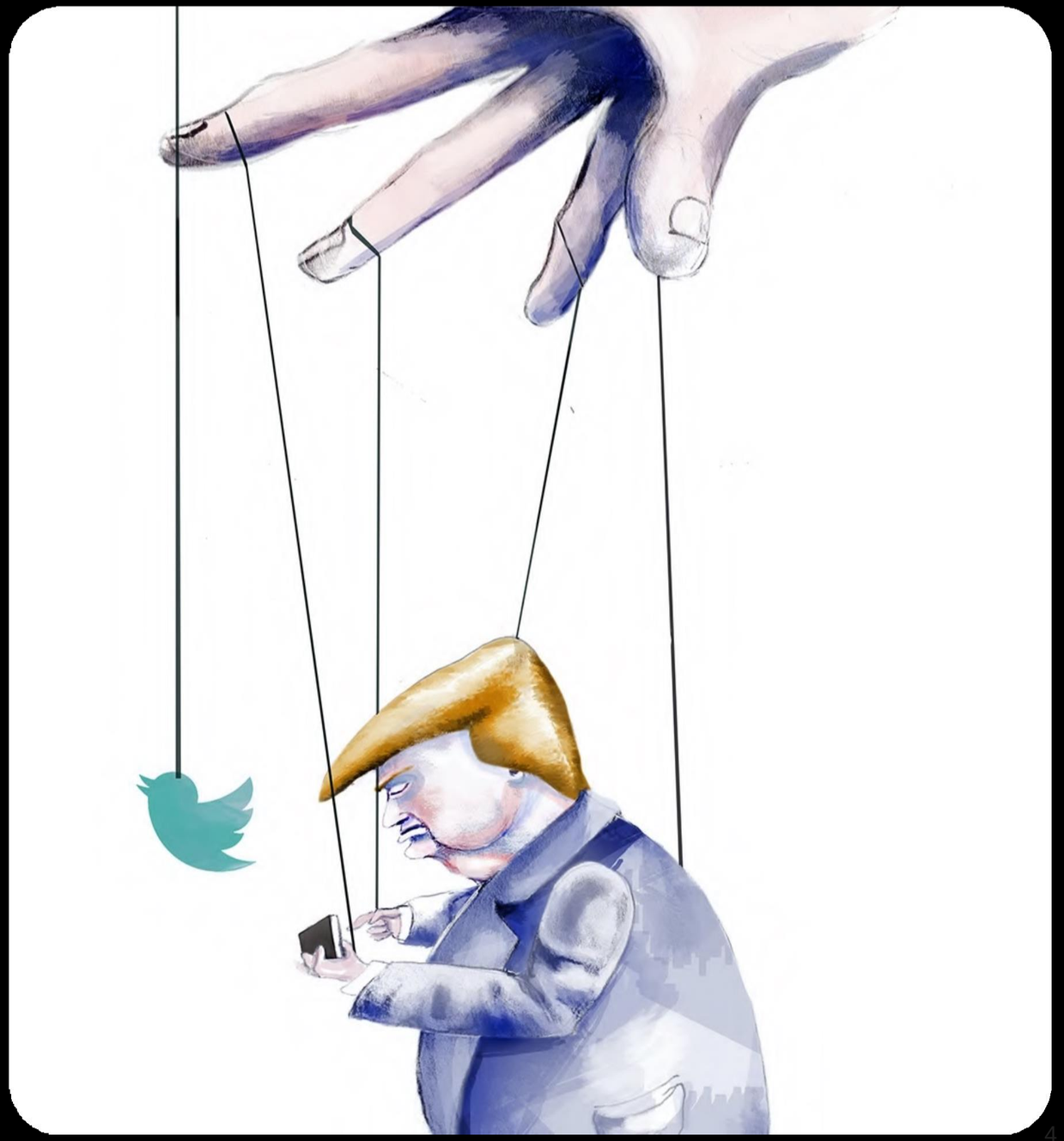
- In recent years, messages and social media has ended up being a very close representation of a person’s life and his mental state.
- People are willing to share their thoughts, stories and their personal feelings, mental states, desires on social network sites, blogging platforms etc.
- This is a huge stockpile of data about a person’s behavior and can be used for detection of various emotion states.

- Emotion Analysis, as the name suggests, it means to identify the view or emotion behind a situation.
- That is done by analyzing expressed opinions, images, sentiments, and other activities.

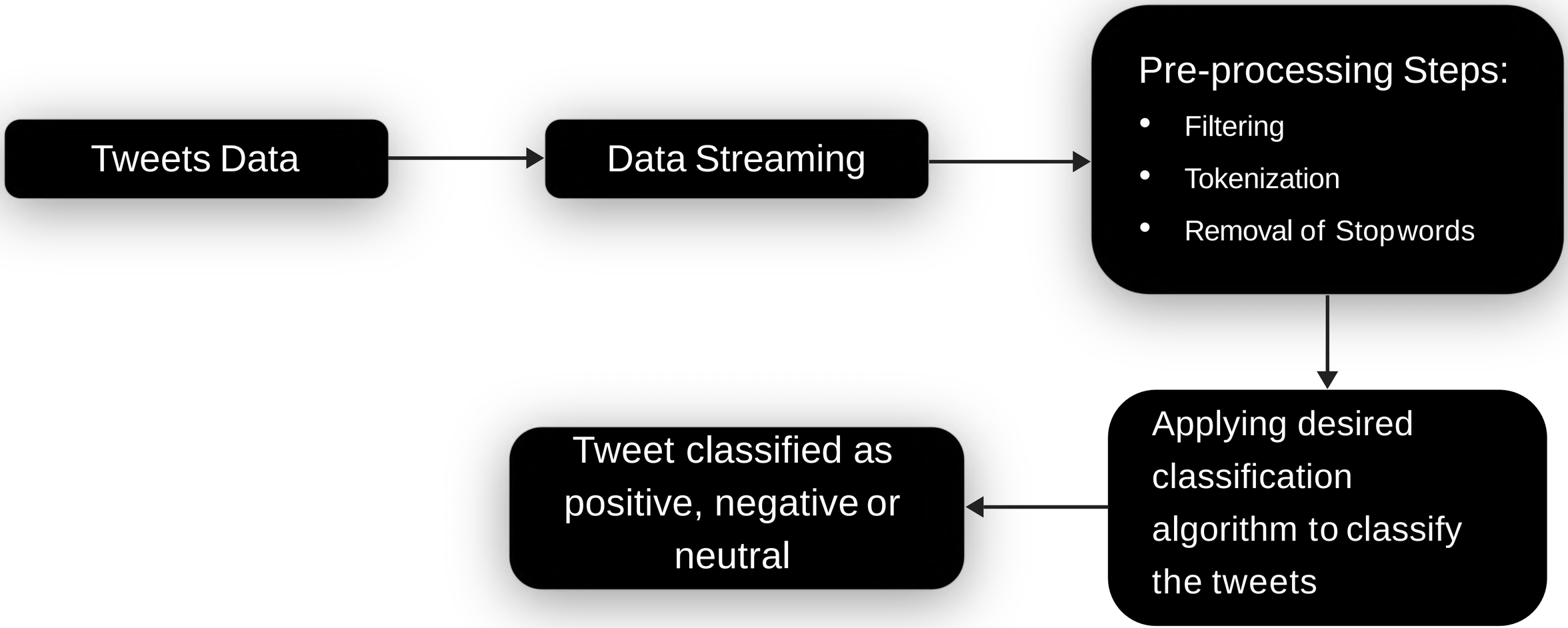


Introduction

- Emotion analysis of tweets used to identify information which helps in businesses to understand their customers' emotion by monitoring online conversations.
- As customers express their reviews and thoughts about the brand more openly than ever before, this emotion analysis has become a powerful tool to monitor and understand online conversations. Analyzing customer feedback and reviews automatically through survey responses or social media discussions allows you to learn what makes your customer happy or disappointed.
- Recent advancements in machine learning and deep learning have increased the efficiency of emotion analysis of tweets using machine learning algorithms.



Flowchart


Tweets Data
Data Streaming
Pre-processing Steps:
Filtering
Tokenization
Removal of Stopwords
Applying desired classification algorithm to classify the tweets
Tweet classified as positive, negative or neutral

Progress

```
# Loading necessary libraries

import numpy as np
import pandas as pd

pd.set_option('display.max_colwidth', 200)

import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline

import re
import string

import warnings

warnings.filterwarnings("ignore", category=DeprecationWarning)
warnings.filterwarnings("ignore", category=FutureWarning)
warnings.filterwarnings("ignore", category=UserWarning)
```

```
[6] train = pd.read_csv('/content/Twitter Sentiment Train.csv')
    test = pd.read_csv('/content/Twitter Sentiment Test.csv')
```

```
[7] print('Training dataset has {} rows and {} columns'.format(train.shape[0], train.shape[1]))
    print('Testing dataset has {} rows and {} columns'.format(test.shape[0], test.shape[1]))
```

```
Training dataset has 31962 rows and 3 columns
Testing dataset has 17197 rows and 2 columns
```

Analyze the datasets by importing necessary libraries.

```
train.head(10)
```

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
1	2	0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in urð□□±!!! ð□□□ð□□□ð□□□ð□□□ð□□¦ð□□¦ð□□¦
4	5	0	factsguide: society now #motivation
5	6	0	[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo
6	7	0	@user camping tomorrow @user @user @user @user @user @user @user dannyâ□¦
7	8	0	the next school year is the year for exams.ð□□¯ can't think about that ð□□ #school #exams #hate #imagine #actorslife #revolutionschool #girl
8	9	0	we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers â□¦
9	10	0	@user @user welcome here ! i'm it's so #gr8 !

```
[9]  test.head(10)
```

	id	tweet
0	31963	#studiolife #aislife #requires #passion #dedication #willpower to find #newmaterialsâ□¦
1	31964	@user #white #supremacists want everyone to see the new â□□ #birdsâ□□ #movie â□□ and hereâ□□s why
2	31965	safe ways to heal your #acne!! #altwaystoheal #healthy #healing!!
3	31966	is the hp and the cursed child book up for reservations already? if yes, where? if no, when? ð□□□ð□□□ð□□□ #harrypotter #pottermore #favorite
4	31967	3rd #bihday to my amazing, hilarious #nephew eli ahmir! uncle dave loves you and missesâ□¦
5	31968	choose to be :) #momtips
6	31969	something inside me dies ð□□¦ð□□¿â□¯ eyes ness #smokeyeyes #tired #lonely #sof #grungeâ□¦
7	31970	#finished#tattoo#inked#ink#loveitâ□¤ï¸□ #â□¤ï¸□â□¤ï¸□â□¤ï¸□â□¤ï¸□ #thanks#aleeee !!!
8	31971	@user @user @user i will never understand why my dad left me when i was so young.... :/ #deep #inthefeels
9	31972	#delicious #food #lovelife #capetown mannaepicure #resturantâ□¦

First ten rows from taken datasets

```
[12] combined = train.append(test, ignore_index=True)
     print('Combined data has {} rows and {} columns'.format(combined.shape[0], combined.shape[1]))

     Combined data has 49159 rows and 3 columns
```

```
[13] combined.head(5)
```

	id	label	tweet
0	1	0.0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
1	2	0.0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked
2	3	0.0	bihday your majesty
3	4	0.0	#model i love u take with u all the time in urð□□±!!! ð□□□ð□□□ð□□□ð□□□ð□□¦ð□□¦ð□□¦
4	5	0.0	factsguide: society now #motivation

```
combined.tail(5)
```

	id	label	tweet
49154	49155	NaN	thought factory: left-right polarisation! #trump #uselections2016 #leadership #politics #brexit #blm >3
49155	49156	NaN	feeling like a mermaid ð□□□ #hairflip #neverready #formal #wedding #gown #dresses #mermaid â□¦
49156	49157	NaN	#hillary #campaigned today in #ohio((omg)) & used words like "assets&liability" never once did #clinton say thee(word) #radicalization
49157	49158	NaN	happy, at work conference: right mindset leads to culture-of-development organizations #work #mindset
49158	49159	NaN	my song "so glad" free download! #shoegaze #newmusic #newsong

```
[25] from nltk.stem.porter import PorterStemmer
     stemmer = PorterStemmer()

     def clean_tweet(tweet):
       clean_handle = re.sub(r'@[\w]*', '', tweet)                                  # Removes twitter handles from tweets
       clean_punc = re.sub(r'[^a-zA-Z#]', ' ', clean_handle)                        # Removes punctuation, special characters(except #tags)
       clean_short_tokenized = [word for word in clean_punc.split() if len(word) > 3] # Remove short words and tokenize
       clean_normalize = [stemmer.stem(word) for word in clean_short_tokenized]     # Stem tokenized words
       return ' '.join(clean_normalize)
```

The implementation of PorterStemmer and creating a function to read and to remove twitter handles, punctuation, short words and stopwords.

```
clean_tweet(combined.tweet.iloc[2])
combined.tweet
df=combined.tweet.to_frame()
df
```

	tweet
0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
1	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked
2	bihday your majesty
3	#model i love u take with u all the time in urð□□±!!! ð□□□ð□□□ð□□□ð□□□ð□□□¦ð□□¦ð□□¦
4	factsguide: society now #motivation
...	...
49154	thought factory: left-right polarisation! #trump #uselections2016 #leadership #politics #brexit #blm >3
49155	feeling like a mermaid ð□□□ #hairflip #neverready #formal #wedding #gown #dresses #mermaid â□¦
49156	#hillary #campaigned today in #ohio((omg)) & used words like "assets&liability" never once did #clinton say thee(word) #radicalization
49157	happy, at work conference: right mindset leads to culture-of-development organizations #work #mindset
49158	my song "so glad" free download! #shoegaze #newmusic #newsong

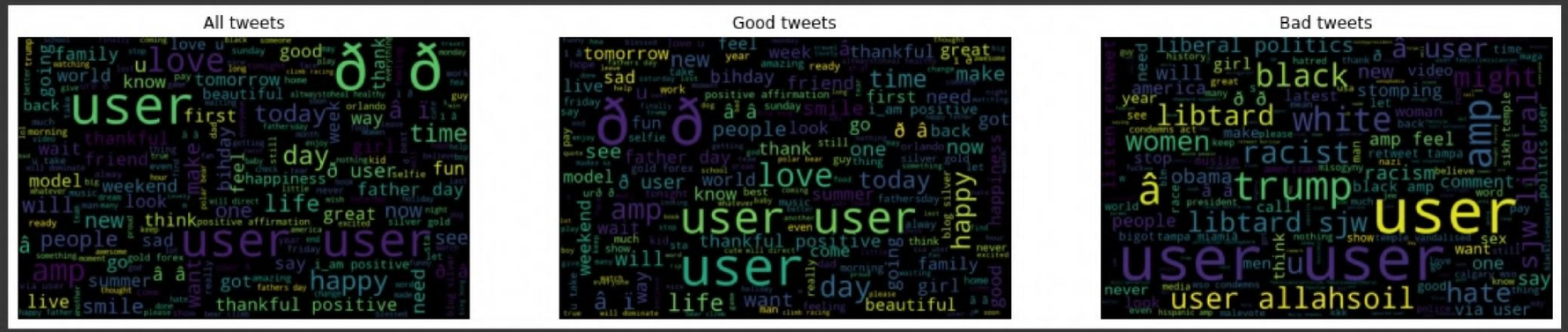
49159 rows x 1 columns

Conversion of combined tweets into dataframe

```
[ ] from wordcloud import WordCloud, STOPWORDS
    wordcloud = WordCloud(width=800, height=500, random_state=42, max_font_size=110, stopwords=STOPWORDS)

[ ] all_tweets = ' '.join(combined.tweet)
    good_tweets = ' '.join(combined[combined.label == 0].tweet)
    bad_tweets = ' '.join(combined[combined.label == 1].tweet)

[ ] bad_tweets

'@user #cnn calls #michigan middle school \'build the wall\' chant \'\' #tcot  no comment!  in #australia  #opkillingbay #seashepherd #helpcovedolphins #thecove  #helpcovedolphins retweet if you agree!  @user @user lumpy says i am
a . prove it lumpy. it\'s unbelievable that in the 21st century we\'d need something like this. again. #neverump  #xenophobia  @user lets fight against  #love #peace  ð\x9f\x98@the white establishment can\'t have blk folx running aro
und loving themselves and promoting our greatness  @user hey, white people: you can call people \'white\' by @user  #race  #identity #medâ\x80¦ how the #altright uses  &amp; insecurity to lure men into #whitesupremacy  @user i\'m
not interested in a #linguistics that doesn\'t address #race &amp; . racism is about #power. #raciolinguistics bringsâ\x80¦ @user why not @user mocked obama for being black.  @user @user @user @user #brexit #people aren\'t protesting
#trump because a #republican won-they do so because trump has fuher…'
```

Creation of word cloud to generate an image of most representative words

```
plt.figure(figsize=(20,12))
plt.subplot(1,3,1)
plt.imshow(wordcloud.generate(all_tweets), interpolation='bilinear')
plt.axis('off')
plt.title('All tweets')
plt.subplot(1,3,2)
plt.imshow(wordcloud.generate(good_tweets), interpolation='bilinear')
plt.axis('off')
plt.title('Good tweets')
plt.subplot(1,3,3)
plt.imshow(wordcloud.generate(bad_tweets), interpolation='bilinear')
plt.axis('off')
plt.title('Bad tweets')
plt.show()
```



Clear visualization of all tweets, good tweets, bad tweets from the dataset

```
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
nltk.download("vader_lexicon")
sentiments = SentimentIntensityAnalyzer()
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
```

```
combined.tweet = combined.tweet.apply(lambda x : clean_tweet(x))
df["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in df['tweet']]
df["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in df["tweet"]]
df["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in df["tweet"]]
df['Compound'] = [sentiments.polarity_scores(i)["compound"] for i in df["tweet"]]
df.head()
```

	tweet	Positive	Negative	Neutral	Compound
0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	0.000	0.385	0.615	-0.8296
1	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked	0.256	0.000	0.744	0.6705
2	bihday your majesty	0.000	0.000	1.000	0.0000
3	#model i love u take with u all the time in urð□□±!!! ð□□□ð□□□ð□□□ð□□□ð□□¦ð□□¦ð□□¦	0.337	0.000	0.663	0.7249

The code for adding 4 columns of positive, negative, neutral and compound to the dataset.

```
score = df["Compound"].values
sentiment = []
for i in score:
    if i >= 0.05 :
        sentiment.append('Positive')
    elif i <= -0.05 :
        sentiment.append('Negative')
    else:
        sentiment.append('Neutral')
df["Sentiment"] = sentiment
df.head()
```

	tweet	Positive	Negative	Neutral	Compound	Sentiment
0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	0.000	0.385	0.615	-0.8296	Negative
1	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked	0.256	0.000	0.744	0.6705	Positive
2	bihday your majesty	0.000	0.000	1.000	0.0000	Neutral
3	#model i love u take with u all the time in urð□□±!!! ð□□□ð□□□ð□□□ð□□□ð□□¦ð□□¦ð□□¦	0.337	0.000	0.663	0.7249	Positive
4	factsguide: society now #motivation	0.000	0.000	1.000	0.0000	Neutral

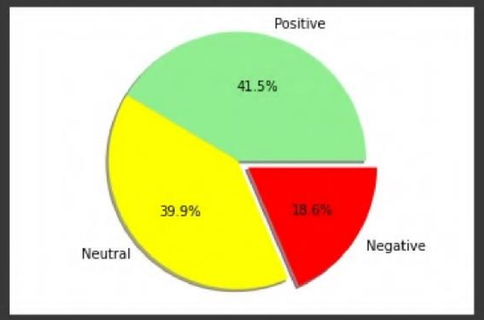
Classification of emotion of tweets by comparing compound score.

```
x=df["Sentiment"].value_counts()
print(x)
```

```
Positive    20402
Neutral     19625
Negative     9132
Name: Sentiment, dtype: int64
```

Frequencies of all labels.

```
import matplotlib.pyplot as plt
%matplotlib inline
labels="Positive", "Neutral","Negative"
sizes=[x]
colors=["lightgreen","yellow","red"]
explode=(0,0,0.1)
plt.pie(sizes,labels=labels,explode=explode,colors=colors,autopct="%1.1f%%",shadow=True)
plt.axis('equal')
plt.show()
```



Number of positive, negative, neutral tweets present in the given dataset.

Output - Models Accuracy

Support vector machine

```
[41] from sklearn.svm import SVC
     svc = SVC(kernel='linear', C=1, probability=True)
```

```
[47] svc.fit(x_bow_train, y_bow_train)
     bow_pred_prob = svc.predict_proba(x_bow_test)
     bow_pred_thresh = bow_pred_prob[:, 1] >= 0.3
     bow_pred = bow_pred_thresh.astype(np.int)
     print('F1 Score : ',f1_score(y_bow_test, bow_pred))

     F1 Score :  0.5078776645041705
```

```
[48] svc.fit(x_tfidf_train, y_tfidf_train)
     tfidf_pred_prob = svc.predict_proba(x_tfidf_test)
     tfidf_pred_thresh = tfidf_pred_prob[:, 1] >= 0.3
     tfidf_pred = tfidf_pred_thresh.astype(np.int)
     print('F1 Score : ',f1_score(y_tfidf_test, tfidf_pred))

     F1 Score :  0.5109489051094891
```

The code for implementing SVM algorithm.

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(n_estimators=400, random_state=11)
```

```
[50] rfc.fit(x_bow_train, y_bow_train)
     bow_pred = rfc.predict(x_bow_test)
     print('F1 Score : ',f1_score(y_bow_test, bow_pred))
```

```
F1 Score :  0.5494853523357086
```

```
[51] rfc.fit(x_tfidf_train, y_tfidf_train)
     tfidf_pred = rfc.predict(x_tfidf_test)
     print('F1 Score : ',f1_score(y_tfidf_test, tfidf_pred))
```

```
F1 Score :  0.5592592592592593
```

The code for Random Forest model and finding the F1 score of this model.

Code link: https://colab.research.google.com/drive/1I3pZV4yoCgQZaNKa9HIsnZ18hAfzAeoF

Conclusion

- We learned how to approach an emotion analysis problem by preprocessing and exploration of data. Then we extracted features from the cleaned text using Bag-of-Words and TF-IDF algorithms.

- The best model fit is calculated by finding F1 score of these two models and based on the F1 scores we took maximum value of F1 score algorithm and we treat that algorithm as our main priority model to predict the emotion of tweet.

Bibliography

- Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, “Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Media,2011 , pp. 30-38
- Pang, B.and Lee, L. “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts”. 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, 271-278.
- Neethu M,S and Rajashree R,” Sentiment Analysis in Twitter using Machine Learning Techniques” 4th ICCCNT 2013,at Tiruchengode, India. IEEE – 31661
- A. Pak,and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” Special Issue of International Journal of Computer Application, France:Universitede Paris-Sud, 2010
- Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacherjee, S. K., & Viswanathan, A. (2015, October). Sentiment analysis of Facebook data using Hadoop based open source technologies. In Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on (pp. 1-3). IEEE.


THANK
YOU