# TWITTER SENTIMENT ANALYSIS.


## TEAM MEMBERS:

- SAI TEJA BALUSU (**TEAM LEAD**)
- RAKESH NATH DHULIPALLA
- JAI SAI MALAKALAPALLI

# GOALS AND OBJECTIVES:

- **MOTIVATION:**
  It keeps on tracking the latest updates on social media which includes markets, products, podcasts, political and celebrity news, etc.… and detects the positives and negatives in the twitter tweets, posts or comments which can prevent from a great surge.

- **SIGNIFICANCE:**
  We picked twitter because for sentiment analysis twitter is can provide a lot of data that includes people expressions like opinions, feelings, monitoring's on different fields and arguments. So, with this data you can find different trends, interests and behaviors of people and helps the sentiment analysis tool to improve customer service.

- **OBJECTIVES:**
  The main objective is to know the people's opinion on different topics which they have targeted. So, we analyse the text from n no of sources to provide sentiment analysis on different topics by using hashtags.

- **FEATURES:**
  It is automating the categories of the sentiments of messages tweeted on twitter by resisting and fixing the num of tasks which includes characters restrictions, unnatural styles of writings, etc...

# INCREMENT 1.

- **Related Work (Background):**
  We use data processing for the text and then we pass the data to the sentiment classifier to extract the sentiment of the data. We might not be able to get the best probability but the relative probability of the entire data itself will be enough to understand the sentiment of the given tweet.

- **Dataset:**
  We perform twitter authentication and use tweepy to request over 2500 tweets for particular hash tag. So, we test and train the data in the dataset to make it easier for system to analyse the data with zero null values.

- **Detail design of Features:**
  Is to analyse the sentiment of the people on given topic to predict the percentage of positives and negatives to provide better customer support before escalating the topic.

- **Analysis:**
  In our analysis not everyone has the same opinion on the given topic so from this we can understand. The general trend of the people's opinion on different if we analyse the topic on regular basis which can give the general trend of the entire population Example: In a cricket tournament people might support one team while the tournament in going on if the team is not performing well so the opinion of the people will change so we can predict the people's opinion only today but not in the future.

- **Implementation:**

  We need to import:

  1. Sys:(helps us accessing the system specific functions and parameters),
  2. Tweepy: (helps us in accessing twitter API),
  3. Matplotlib: (used for plotting graphs),
  4. pandas: helps in bringing data to respective environment.
  5. NumPy: numerical expressions.
  6. Os: to interact with operating systems to operate a folder.
  7. nltk: it is natural language tool kit to program with human language data.
  8. pycountry: it operated the databases.
  9. re: it's a regular expression.
  10. String:(helps in functioning standard python strings.

  **Steps involved:**

  I. Install packages and import libraries.
  II. Authentication of twitter API'S.
  III. Now we need to calculate paraments like compound, polarity , positive , neutral and negative by using text blog and get text you need to use tweepy.
  IV. Getting tweets with hashtags,
  V. Here you need to press a keyword and no of tweets to get the sentiment analysis based on [TOTAL, POSITIVE, NEUTRAL, NEGATIVE] in the form of percentages.
  VI. To represent it we use barographs, pie charts, etc…

- **Preliminary Results:**

### Importing libraries:

```
from textblob import TextBlob
import sys
import tweepy
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import os
import nltk
import pycountry
import re
import string

from wordcloud import WordCloud, STOPWORDS
from PIL import Image
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from langdetect import detect
from nltk.stem import SnowballStemmer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from sklearn.feature_extraction.text import CountVectorizer
```

### Authentication:

```
consumerKey = "Type your key"
consumerSecret = "Type your secret"
accessToken = "Type your  authentication token"
accessTokenSecret = "Type your authentication token secret"

auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokenSecret)
api = tweepy.API(auth)
```

## Sentiment analysis:

```
def percentage(part,whole):
    return 100 * float(part)/float(whole)

keyword = input("enter keyword or hashtag: ")
noOfTweet = int(input ("enter how many tweets you need to analyze:"))


tweets = tweepy.Cursor(api.search, q=keyword).items(noOfTweet)
positive  = 0
negative = 0
neutral = 0
polarity = 0
tweet_list = []
neutral_list = []
negative_list = []
positive_list = []

for tweet in tweets:

    #print(tweet.text)
    tweet_list.append(tweet.text)
    analysis = TextBlob(tweet.text)
    score = SentimentIntensityAnalyzer().polarity_scores(tweet.text)
    neg = score['neg']
    neu = score['neu']
    pos = score['pos']
    comp = score['compound']
    polarity += analysis.sentiment.polarity

    if neg > pos:
        negative_list.append(tweet.text)
        negative += 1

    elif pos > neg:
        positive_list.append(tweet.text)
        positive += 1

    elif pos == neg:
        neutral_list.append(tweet.text)
        neutral += 1

positive = percentage(positive, noOfTweet)
negative = percentage(negative, noOfTweet)
neutral = percentage(neutral, noOfTweet)
polarity = percentage(polarity, noOfTweet)
```

```
neutral = percentage(neutral, noOfTweet)
polarity = percentage(polarity, noOfTweet)
positive = format(positive, '.1f')
negative = format(negative, '.1f')
neutral = format(neutral, '.1f')
```

**No of tweets:**

```
tweet_list = pd.DataFrame(tweet_list)
neutral_list = pd.DataFrame(neutral_list)
negative_list = pd.DataFrame(negative_list)
positive_list = pd.DataFrame(positive_list)
print("total : ",len(tweet_list))
print("positive: ",len(positive_list))
print("negative: ", len(negative_list))
print("neutral: ",len(neutral_list))
```

**Representation:**

```
labels = ['Positive ['+str(positive)+'%]' , 'Neutral ['+str(neutral)+'%]','Negative ['+str(negative)+'%]']
sizes = [positive, neutral, negative]
colors = ['yellowgreen', 'blue','red']
patches, texts = plt.pie(sizes,colors=colors, startangle=90)
plt.style.use('default')
plt.legend(labels)
plt.axis('equal')
plt.show()
```

- **Project Management:**
  Here we are taking raw data from twitter and we are doing simple text processing using text blog and tweepy. We use positive, neutral and negative from the processed text data to get the output i.e., view of people on different topics. We are using twitter as base because of its loyal customer base who tweets every day on current topic which provides a lot of data.

# IMPLEMENTATION STATUS REPORT.

- **WORK COMPLETED:**
  Twitter authentication, Processing of data, estimating the trend

- **DESCRIPTION:**
  We will gather data for twitter using hashtags and test and train the data, analyse the data and finding the present trends of positive, neutral and negative.

- **RESPONSIBILITY:**
  I. Sai Teja Balusu ----------Data Processing
  II. Rakesh Nath Dhulipalla ---- Data Visualization
  III. Jai Sai Malakalapalli ---------Twitter Authentication

- **CONTRIBUTIONS:**
  I. Sai Teja Balusu ----------35/100
  II. Rakesh Nath Dhulipalla ---- 35/100
  III. Jai Sai Malakalapalli ---------30/100

- **WORK TO BE COMPLETED:**
  Increasing the efficiency of the code to attain accuracy.

- **DESCRIPTION:**
  We will do the code to find the most using words

- **RESPONSIBILITY (TASK, PERSON):**
  I. Sai Teja Balusu ---------- increasing the probability
  II. Rakesh Nath Dhulipalla ---- Finding ways to Improve efficiency of the code
  III. Jai Sai Malakalapalli --------- testing different test cases.

- **ISSUES/CONCERNS:**
    I. It's hard-to-get twitter authentication.
    II. The data will change every day which impacts the credibility of the report

- **REFERENCES:**
- https://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d6f650ade58d

- https://developer.twitter.com/en/docs/tutorials/how-to-analyze-the-sentiment-of-your-own-tweets

- https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/