PROJECT

Introduction to Data Science

Sentimental Analysis

Presented by:

**Student Reg#**  **Student Name**

L1f17bscs0181 Moeez Mujahid

L1f17bscs0180 Usama Sohail

L1f17bscs0210 Abdul Moaeed

**Faculty of Information Technology**

**University of Central Punjab**

**Note:**

**The Code was done on Google collab ipnyb files are shared**

**The main dataset folder contain our main task**

**Other folder contain other dataset visualization**

**Note**: The datasets we are interested on are based on sentimental analysis and its influence based on twitter tweets or any other social media.

(Using tweets to predict various sentimental behavior positive, neutral, negative etc.)

**C) Questions to be answered:**

1. List of words can help to identify negative tweets?

Python Dictionary are used to store words as a key alongside their frequency representing the impact, and frequency of that particular word.

Furthermore in CNN classifier, Stanford nlp pertained vectors were used in order to determine the nature of the word.

Some list of words possible in tweets:

Worst, shit, not good, hate, Injurious, hellish, terrible, iniquitous, harmful, rotten, hateful, disgusting, dreadful, unpleasant, cursed, horrible, unpleasant, annoying, worthless.

For depression: Blue, Cheerless, Dejected, Rejected, Depressed, Desperate, Discouraged, Disgusted, Disappointed, Heartbroken, Upset, Miserable, Sorrowful, and Unhappy.

1. How to differentiate between sarcasm and serious tweets?

In micro texts, such as Twitter.com sarcasms, the hatch tag #sarcasm is frequently expressly marked to avoid the sarcastic message being understand in its unintentional literal significance. We exaltedly decided to use only the tweets with the #Airline name symbol thus avoiding as much sarcasm as possible.

1. Find the standard deviation among the users and negative tweets.

(Dataset Bitcoin 17.7 million Tweets and price, Bitcoin Tweets sentiment analysis and the price)

165.6945263939214 // Count Negatives

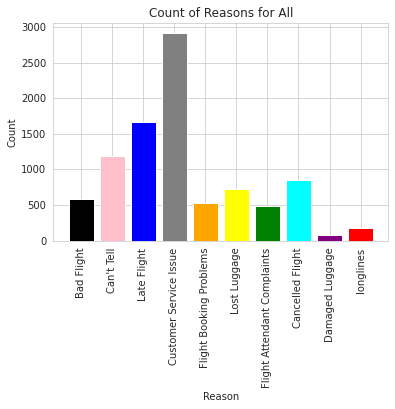
0.03998731766422437 // Sent Negatives

For twitter airline dataset:

Sentiments Standard dev: 0.756084

1. Find the significance of relation among Negative tweets and the reasons for negative tweets

(From dataset airline tweets)



(Fig 1)From twitter airline dataset, count of reasons for negative tweets label

5) Effect of tweets on the business (airline, bitcoin)..?

# 🡪 “Tesla will no longer accept Bitcoin over climate concerns”, says Musk

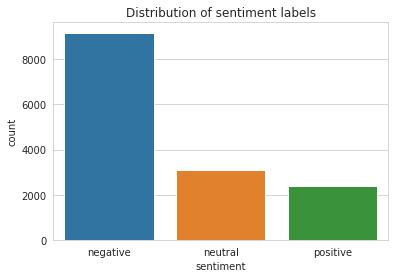
# Bitcoin fell by more than 10% after this tweet. Tweets can have important impact on the ups and downs price of bitcoin and crypto, specially tweets from the big fish in the bitcoin market like Elon Musk.

🡪 For years now, as has been well documented, travelers have been using the social media platform to express their compliments, concerns and downright outrage over everything from delayed flights and lost baggage to far more serious issues, including racially motivated incidents, mistreatment of passengers by airline staff and more.

6) What is the ratio of negative tweets over positive?

From depression dataset: 77.56% negative tweets

22.44% positive tweets



(Fig 2): ratio / count of sentimental labels .

7) What marketing strategies can be imposed using sentimental analysis?

By gauging the public's opinion of an event or product through analysis of data on a scale no human could achieve, it gives your team the ability to figure out what people really think. Using sentiment analysis tools allows you to evaluate the attitudes of your target consumer’s attitudes that can make or break your brand's reputation. Imagine your business just released a product and everyone is talking about it on social media. There are thousands of Instagram posts, Facebook posts, and tweets.

**D) Algorithms to be used**

We will use Naive **Bayes** as it is some of the most used algorithms in text classification and text analysis.

Also we will use CNN classifier alongside pertained vectors of Stanford nlp to test the results

Accuracy for Multinomial Naïve Bayes Accuracy: 0.764344262295082

Accuracy for CNN classifier: Accuracy : 0.6270492

**Dataset 1**: [Twitter US Airline Sentiment | Kaggle](https://www.kaggle.com/crowdflower/twitter-airline-sentiment)

**Title: Twitter US Airline Sentiment**

(Analyze how travelers in Feb 2015 expressed their feeling on twitter)

**Description**:

The dataset is targeted to analyze sentiments about the problem of each major airline.

The tweets are classified as negative, positive and neutral tweets, and further classified

By negative reason e.g. (late flight, bad weather, rude service) etc.

The customers tweets are from various major airlines e.g., Virgin America, US airways etc.

The ratio about airline sentiment confidence and negative reason confidence is also provided.

**Dataset 2**: [Bitcoin 17.7 million Tweets and price Bitcoin Tweets sentiment analysis and the price](https://www.kaggle.com/jaimebadiola/bitcoin-tweets-and-price)

**Title: Bitcoin 17.7 million Tweets and price Bitcoin Tweets sentiment analysis and the price**

Description:

This dataset contains the average sentiment of all tweets about bitcoin from 01/08/2017 until 21/01/2019. It also contains the financial data of bitcoin for that same period.

## How did I gather the tweets?

To collect all tweets, I used this GitHub. Some days have missing data but I think it is minimal. I collected over 17.7 million tweets

## How did I do sentiment analysis?

I used the library Vader Sentiment. I added about 30 expressions and words to the dictionary. To score the expressions I used the same methodology as the authors described in their paper.

## What do I want to accomplish?

I would like to create a predictive model for bitcoin's price using only twitter's data.

**Dataset 3:** [**Sentimental Analysis for Tweets**](https://www.kaggle.com/gargmanas/sentimental-analysis-for-tweets/tasks?taskId=4255)

**Title: Sentimental Analysis for tweets**

**Description:**

The dataset is for finding if a person is depressed from their use of words on social media can definitely help in the cure!and stop desasters like suicide and mental depression.

The dataset simply contains tweets from the twitter users and classified as 1 and 0 indicating weather the tweet is depressing or not.NLP can also be used on the data set to exrtact various features .