**Twitter Sentimental Analysis**

Submitted in partial fulfilment of the requirements of the degree of

Bachelor of Engineering

by

Masood Khan (3117021)

Ashhar Shaikh (3117052)

Faraz Shaikh (3117053)

Needa Shaikh (3117057)

Supervisor (s):

Prof. Waheeda Dhokley

Computer Engineering/M.H Saboo Siddik College of Engineering.

Mumbai University

2019-20

**CERTIFICATE**

This is to certify that the project entitled **“Twitter Sentimental Analysis”** is a bonafide work of **“Masood Khan (3117021)”, ”Ashhar Shaikh(3117052)”, ”Faraz Shaikh(3117053)”, ”Needa Shaikh (3117057)”** submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **“Undergraduate”** in **“Bachelor of Computer Engineering”.**

Prof. Waheeda Dhokley

Supervisor

Dr. Zainab Pirani

Head of Department

# Thesis Approval for Project Report Approval for B. E.

This project report entitled Twitter Sentimental Analysis by **Masood Khan, Ashhar Shaikh, Faraz Shaikh, Needa Shaikh** is approved for the degree of Bachelor in Computer Engineering.

##### Examiners

1. Prof. Waheeda Dhokley

Date: 23 April 2020

Place: Mumbai

# Declaration

##### We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Masood Khan (3117021)

Ashhar Shaikh (3117052)

Faraz Shaikh (3117053)

Needa Shaikh (3117057)

Date: 23 April 2020

Place: Mumbai

**Chapter 1)**

**Introduction**

Social media has profound impact in capturing the potential customers and thus there are a lot of consulting firms that operate in the digital strategy space. Whether it is to design a marketing campaign or look at the effect of marketing campaigns on user engagement or sentiment, it is a very valuable tool.

Natural Language Processing (NLP) is a hotbed of research in data science these days and one of the most common applications of NLP is sentiment analysis. From opinion polls to creating entire marketing strategies, this domain has completely reshaped the way businesses work, which is why this is an area every data scientist must be familiar with.

Thousands of text documents can be processed for sentiment (and other features including named entities, topics, themes, etc.) in seconds, compared to the hours it would take a team of people to manually complete the same task.

We will do so by following a sequence of steps needed to solve a general sentiment analysis problem. We will start with preprocessing and cleaning of the raw text of the tweets. Then we will explore the cleaned text and try to get some intuition about the context of the tweets. After that, we will extract numerical features from the data and finally use these feature sets to train models and identify the sentiments of the tweets.

We have worked upon sentimental analysis of twitter comments and predicting results upon classification as 0,1,2,3 as positive, negative, neutral and can’t say respectively. The Machine Learning Algorithm used is Support Vector Machine (SVM) .We have demonstrated all plots such as Word Cloud, Count plot, Training vs Testing Accuracy plot and metrics count including confusion metrics.

**Chapter 2)**

**Sentimental Analysis**

Sentiment analysis is a text analysis method that detects polarity (e.g. a positive or negative opinion) within text, whether a whole document, paragraph, sentence, or clause.

Understanding people’s emotions is essential for businesses since customers are able to express their thoughts and feelings more openly than ever before. By automatically analyzing customer feedback, from survey responses to social media conversations, brands are able to listen attentively to their customers, and tailor products and services to meet their needs.

For example, using sentiment analysis to automatically analyze 4,000+ reviews about your product could help you discover if customers are happy about your pricing plans and customer service.

Why Perform Sentiment Analysis?

It’s estimated that 80% of the world’s data is unstructured, in other words it’s unorganized. Huge volumes of text data (emails, support tickets, chats, social media conversations, surveys, articles, documents, etc), is created every day but it’s hard to analyze, understand, and sort through, not to mention time-consuming and expensive.

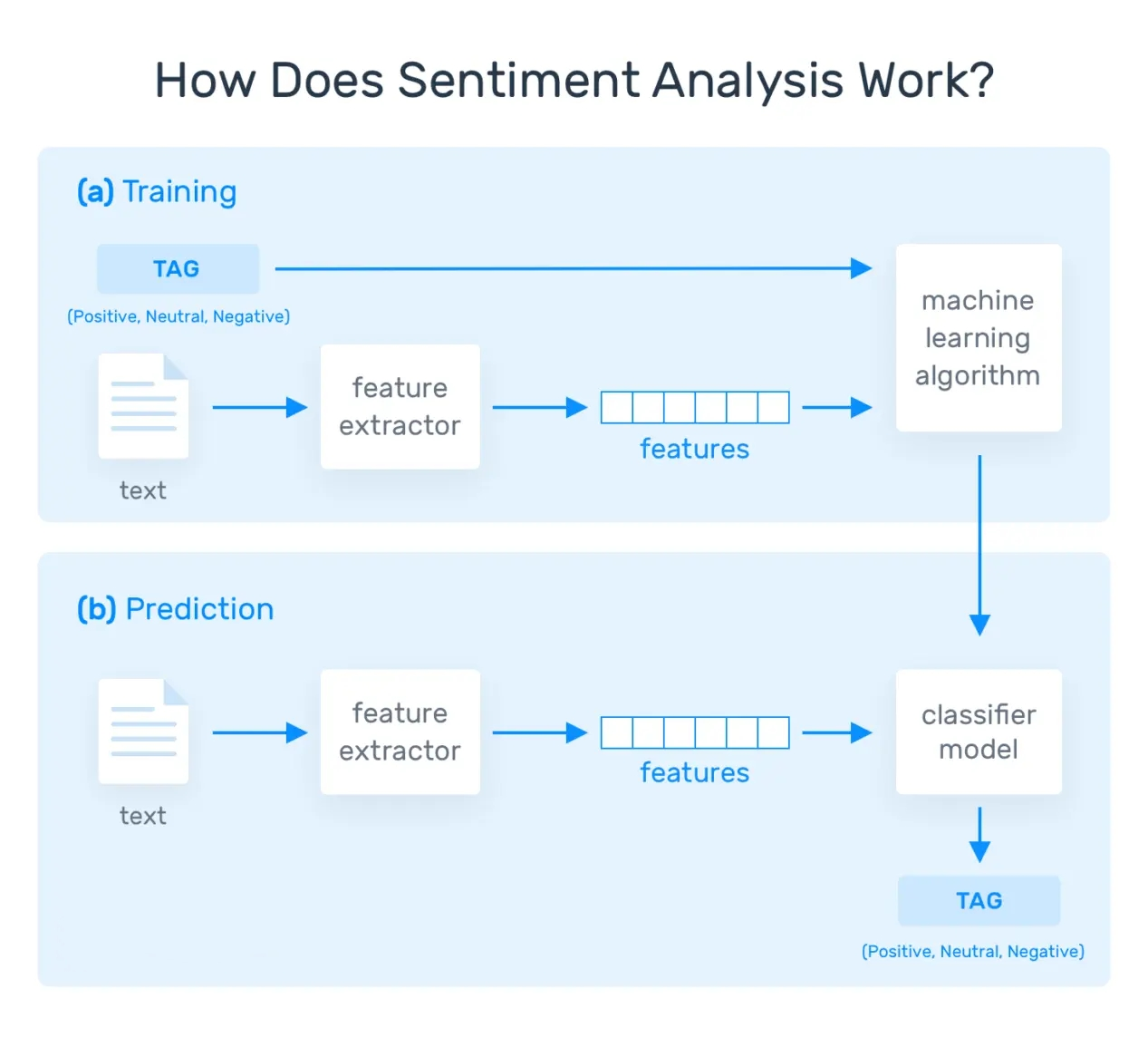
Sentiment analysis, however, helps businesses make sense of all this unstructured text by automatically tagging it.

Benefits of sentiment analysis include:

Sorting Data at Scale Can you imagine manually sorting through thousands of tweets, customer support conversations, or surveys? There’s just too much data to process manually. Sentiment analysis helps businesses process huge amounts of data in an efficient and cost-effective way.

Real-Time Analysis Sentiment analysis can identify critical issues in real-time, for example is a PR crisis on social media escalating? Is an angry customer about to churn? Sentiment analysis models can help you immediately identify these kinds of situations and gauge brand sentiment, so you can take action right away.

Consistent criteria It’s estimated that people only agree around 60-65% of the time when determining the sentiment of a particular text. Tagging text by sentiment is highly subjective, influenced by personal experiences, thoughts, and beliefs. By using a centralized sentiment analysis system, companies can apply the same criteria to all of their data, helping them improve accuracy and gain better insights.



How Accurate Is Sentiment Analysis?

Here’s what sentiment analysis is: it’s a tremendously difficult task even for human beings. That said, sentiment analysis classifiers might not be as precise as other types of classifiers. Remember that inter-annotator agreement is pretty low and that machines learn from the data they are fed with (see above).

That said, you might be saying, is it worth the effort? The answer is simple: it sure is worth it! Chances are that sentiment analysis predictions will be wrong from time to time, but by using sentiment analysis you will get the opportunity to get it right about 70-80% of the times you submit your texts for classification.

If you or your company have not used sentiment analysis before, then you’ll see some improvement really quickly. For typical use cases, such as ticket routing, brand monitoring, and VoC analysis (see below), this means you will save a lot of time and money -which you are likely to be investing in in-house manual work nowadays,- save your teams some frustration, and increase your (or your company’s) productivity.

**Chapter 3)**

**The Training and Prediction Processes.**

In the training process (a), our model learns to associate a particular input (i.e. a text) to the corresponding output (tag) based on the test samples used for training. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g. positive, negative, or neutral) are fed into the machine learning algorithm to generate a model.

In the prediction process (b), the feature extractor is used to transform unseen text inputs into feature vectors. These feature vectors are then fed into the model, which generates predicted tags (again, positive, negative, or neutral).

**Feature Extraction from Text**

The first step in a machine learning text classifier is to transform the text extraction or text vectorization, and the classical approach has been bag-of-words or bag-of-ngrams with their frequency.

More recently, new feature extraction techniques have been applied based on word embeddings (also known as word vectors). This kind of representations makes it possible for words with similar meaning to have a similar representation, which can improve the performance of classifiers.

**Classification Algorithms**

The classification step usually involves a statistical model like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks:

Naïve Bayes: a family of probabilistic algorithms that uses Bayes’s Theorem to predict the category of a text.

Linear Regression: a very well-known algorithm in statistics used to predict some value (Y) given a set of features (X).

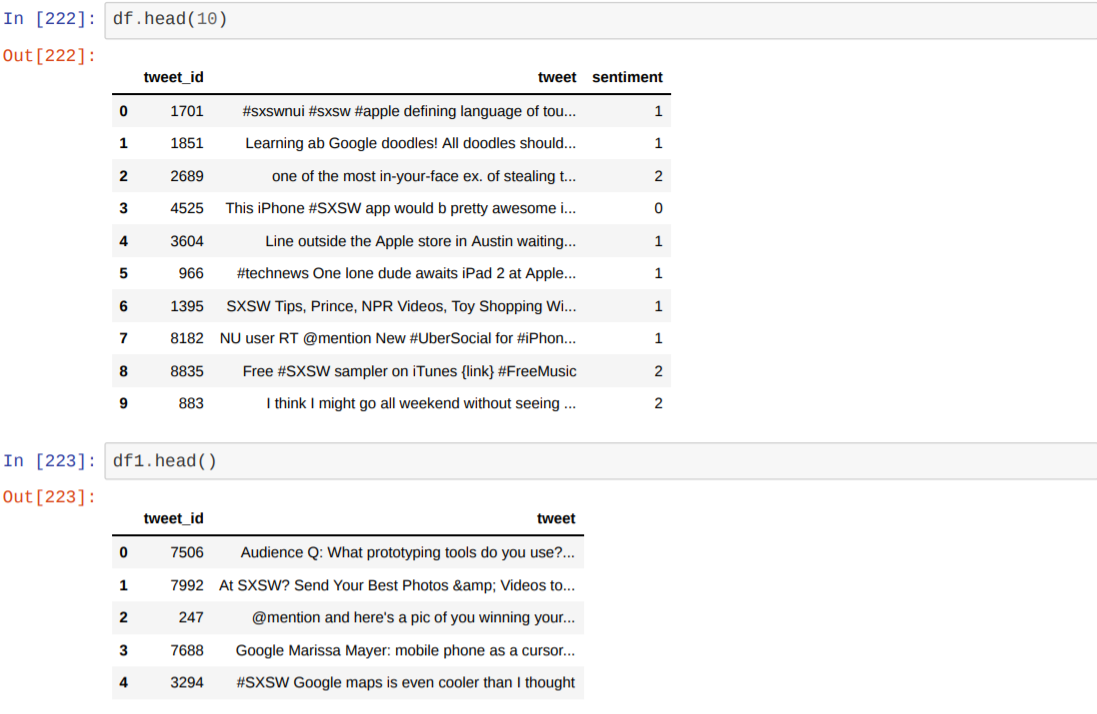
**Support Vector Machines**: a non-probabilistic model which uses a representation of text examples as points in a multidimensional space. Examples of different categories (sentiments) are mapped to distinct regions within that space. Then, new texts are assigned a category based on similarities with existing texts and the regions they’re mapped to.

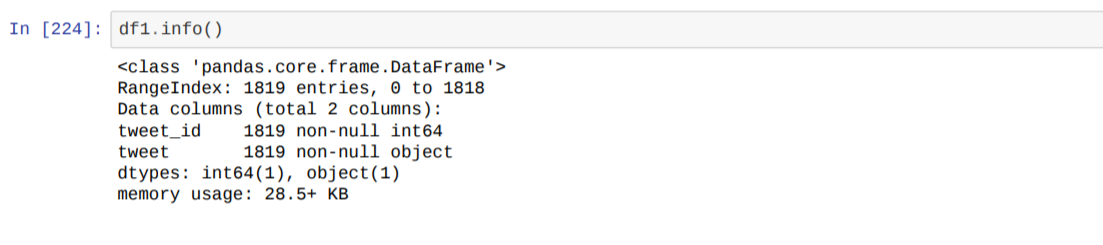
Deep Learning: a diverse set of algorithms that attempt to mimic the human brain, by employing artificial neural networks to process data.

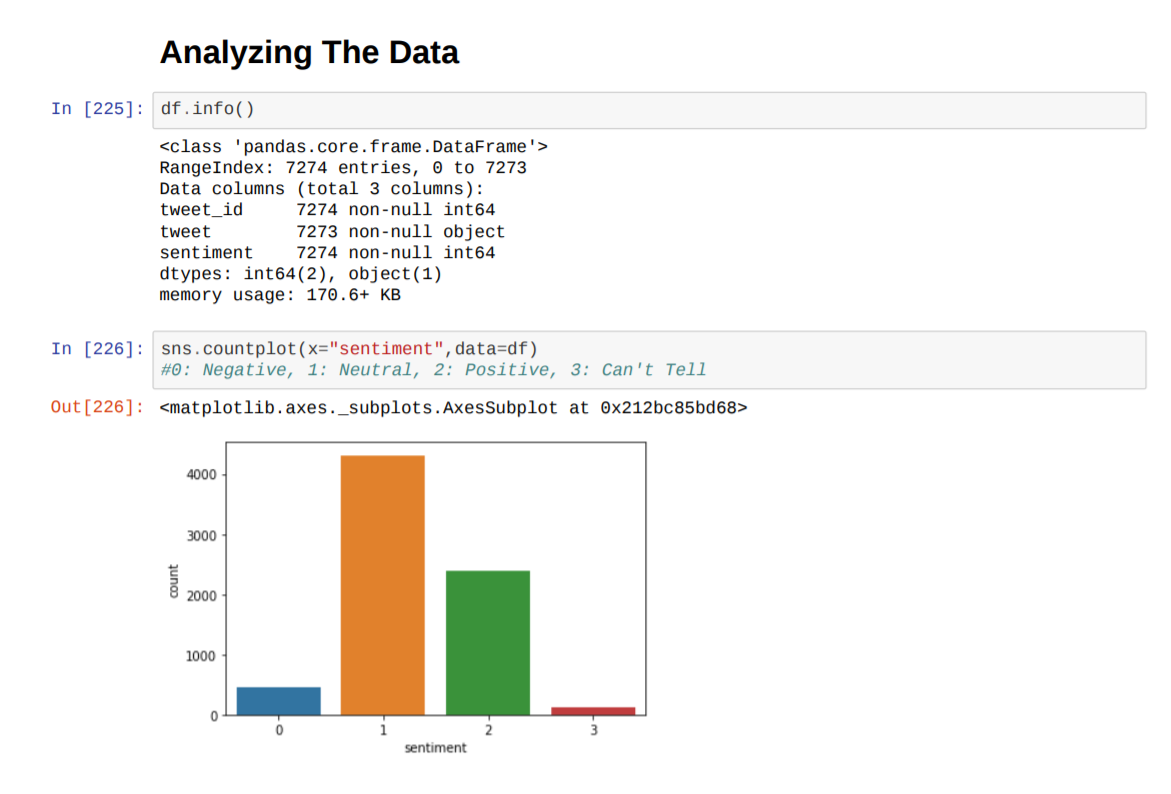
**Chapter 4)**

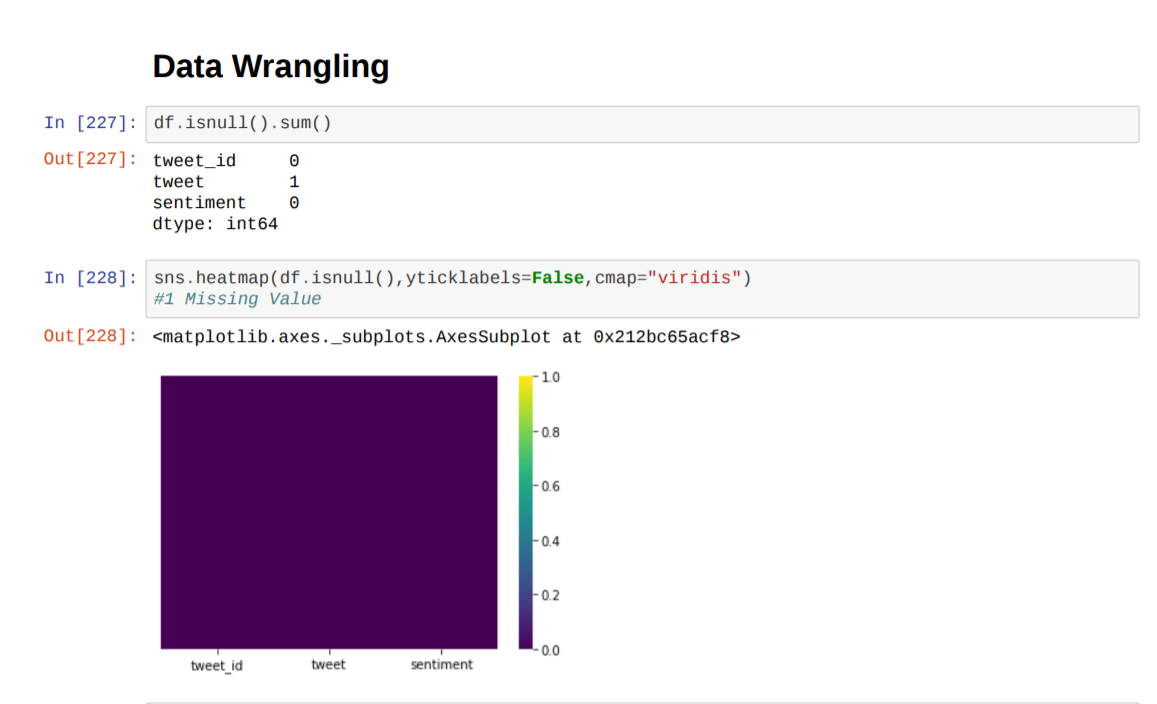
**Data Analysis and Data Wrangling.**



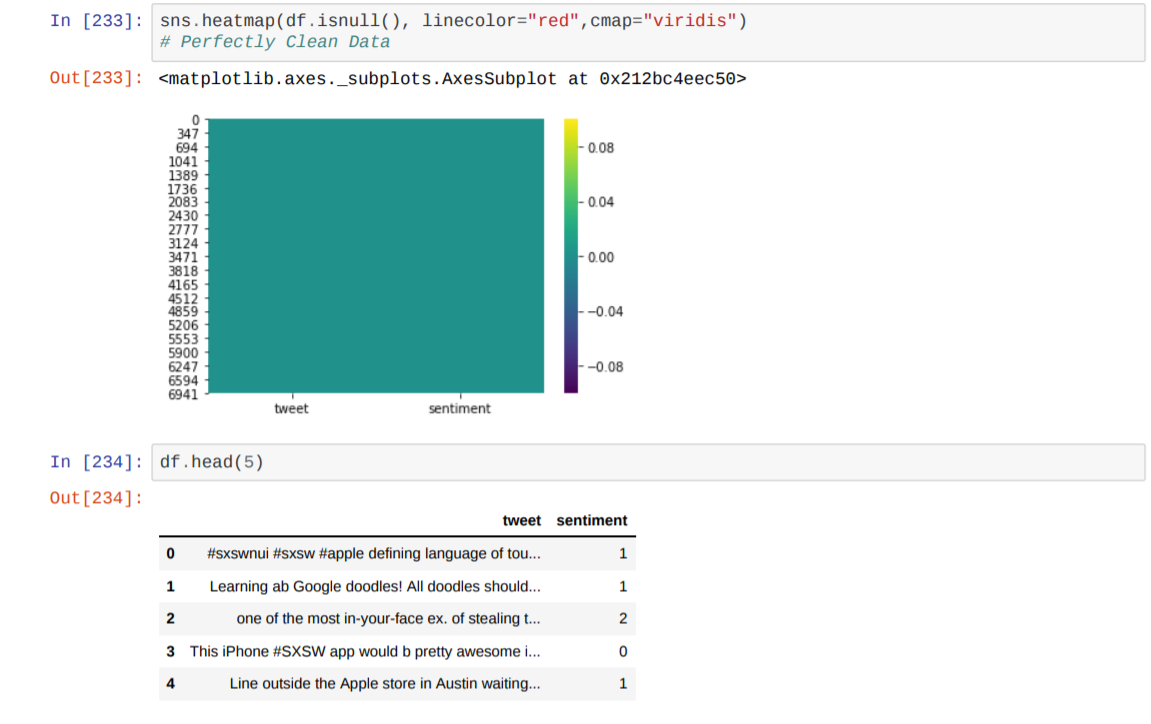












**Chapter 7)**

**Conclusion :**

Thus we have successfully implemented Twitter Sentimental Analysis

Using Support Vector Machine and have achieved an accuracy of 75.78% . Importing Packages, Data Analyzing, Data Wrangling, Exploratory Data Analysis , Data Preprocessing And Cleaning, Advanced Text Processing, Model and Testing are the procedures followed in this project.

**References**