Twitter Sentiment Analysis

Submitted in partial fulfillment of the requirements

for the degree of

Bachelor of Engineering (SEM VIII)

Project Report

by

Prathamesh Chikane

Roll No. 06

Subodh Halpatrao

Roll No. 18

Sutej Kulkarni

Roll No. 35

Under the Supervision of

Prof. J.P.Patil

DEPARTMENT OF INFORMATION TECHNOLOGY
KONKAN GYANPEETH COLLEGE OF ENGINEERING
KARJAT-410201
MAY 2021

Certificate

This is to certify that the project entitled **Twitter Sentiment Analysis** is a bonafide work of Prathamesh Chikane **(Roll No. 06), Subodh Halpatrao (Roll No. 18), Sutej Kulkarni (Roll No. 35)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Undergraduate** in **DEPARTMENT OF INFORMATION TECHNOLOGY.**

Supervisor/Guide

Prof. J.P.Patil

Department of Information Technology

Head of Department

Prof. J.P.Patil

Department of Information Technology

Principal

Dr. M.J. Lengare

Konkan Gyanpeeth College of Engineering

Project Report Approval

This project report entitled **Twitter Sentiment Analysis** by **Prathamesh Chikane (Roll No. 06), Subodh Halpatrao (Roll No. 18), Sutej Kulkarni (Roll No. 35)** is approved for the degree of **DEPARTMENT OF INFORMATION TECHNOLOGY.**

Examiner.

……

……..

DATE.

PLACE.

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I 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. I 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

Prathamesh Chikane (Roll No. 06)

Signature

Subodh Halpatrao (Roll No. 18)

Signature

Sutej Kulkarni (Roll No. 35)

Date.

Abstract

Nowadays, people from all around the world use social media sites to share information. Twitter for example is a platform in which users send, read posts known as `tweets' and interact with different communities. Users share their daily lives, post their opinions on everything such as brands and places. Companies can benefit from this massive platform by collecting data related to opinions on them. The aim of this paper is to present a model that can perform sentiment analysis of real data collected from Twitter. Data in Twitter is highly unstructured which makes it difficult to analyze. However, our proposed model is different from prior work in this field because it combined the use of supervised and unsupervised machine learning algorithms. The process of performing sentiment analysis as follows:
Tweet extracted directly from Twitter API, then cleaning and discovery of data performed. After that the data were fed into several models for the purpose of training. Each tweet extracted classified based on its sentiment whether it is a positive, negative or neutral. We also are performing a real time Emotion analysis of the tweets and text classifying them into Worry, Happy, Love and Hate based tweets or sentence or text.

Acknowledgements

Success is nourished under the combination of perfect guidance, care blessing. Acknowledgement is the best way to convey. We express deep sense of gratitude brightness to the outstanding permutations associated with success. Last few years spend in this estimated institution has molded us into condent and aspiring Engineers. We express our sense of gratitude towards our project guide Prof. J. P. Patil. It is because of his valuable guidance, analytical approach and encouragement that we could learn and work on the project. We will always cherish the great experience to work under their enthusiastic guidance. We are also grateful to our principle Dr. M.J. Lengare who not only supporting us in our project but has also encouraging for every creative activity. We extend our special thanks to all teaching and non-teaching staff, friends and well wishers who directly or indirectly contributing for the success of our maiden mission. Finally, how can we forget our parents whose loving support and faith in us remains our prime source of inspiration. Lastly we would like to thank all those who directly and indirectly helping to complete this project. We would also like to acknowledge with much appreciation the crucial role of the staff of Information Technology Department, who gave the permission to use the all required software/hardware and the necessary material to completing to the project.

Contents

List of Figures

List of Tables

Abbreviations

SVM	**S**upport **V**ector **M**achine
NN	**N**eural **N**etwork
ML	**M**achine **L**earning

Chapter 1

INTRODUCTION

1.1 Introduction

Sentiment Analysis is a technique widely used in text mining. Twitter Sentiment Analysis, therefore means, using advanced text mining techniques to analyze the sentiment of the text in the form of positive, negative and neutral. It is also known as Opinion Mining, is primarily for analyzing conversations, opinions, and sharing of views (all in the form of tweets) for deciding business strategy, political analysis, and also for assessing public actions.

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics and biometrics to systematically identify, extract, quantity and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction or event. The attitude may be judgement or evaluation, affective state (emotional state of the author or speaker), or the intended emotional communication.

1.2 Objectives

In this section we mention objectives of our project:

- To build a system/software using machine learning techniques that analysis the sentiments of users on the twitter platform.

- The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

1.3 Motivation

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. More ever the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment and Emotion analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value.

1.4 Purpose, Scope, and Applicability

1.4.1 Purpose

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral.

Sentiment analysis helps data analysts within large enterprises gauge public opinion, conduct nuanced market research, monitor brand and product reputation, and understand customer experiences whereas the Emotion analysis model gives more specified review and result or a specialised sentiment

1.4.2 Scope

Business:

Companies use Twitter Sentiment Analysis to develop their business strategies, to assess customer's feelings towards products or brand, how people respond to their campaigns or product launches and also why consumers are not buying certain products.

Politics:

In politics Sentiment Analysis Dataset Twitter is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. Sentiment Analysis Dataset Twitter is also used for analyzing election results.

Public Actions:

Twitter Sentiment Analysis also is used for monitoring and analyzing social phenomena, for predicting potentially dangerous situations and determining the general mood of the blogosphere.

1.4.3 Applicability

Sentiment Analysis in business, also known as opinion mining is a process of identifying and cataloging a piece of text according to the tone conveyed by it. This text can be tweets, comments, feedback, and even random rants with positive, negative and neutral sentiments associated with them. Every business needs to implement automated sentiment analysis. If you doubt it, here's a little perspective. The accuracy can never be 100%. And of course, a machine does not understand sarcasm. However, according to a research, people do not agree 80% of the time. It means that even if the machine accuracy does not score a perfect 10, it will still be more accurate than human analysis. Also, when the corpus is huge manually analyzing is not an option. Hence, sentiment analysis in business is more than just a trend.

The applications of sentiment analysis in business cannot be overlooked. Sentiment analysis in business can prove a major breakthrough for the complete brand revitalization. The key to running a successful business with the sentiments data is the ability to exploit the unstructured data for actionable insights. Machine learning models, which largely depend on the manually created features before classification, have served this purpose fine for the past few years.

Business analysts can use this information to track feelings and opinions of people with respect to their products. The problem with most of the Sentiment Analysis that is done today is that the analysis only informs whether the public reaction is positive or negative but fails to describe the exact feelings of the customers and the intensity of their reaction. With our emotional analysis, they can have a more profound analysis of their markets than the naive 2-way or 3-way Sentiment Analysis, which itself has turned their businesses more profitable. Business leaders can analyse the holistic view of people in response to their actions or events and work accordingly. Also, health-analysts can study the mood swings of individuals or masses at different times of the day or in response to certain events. It can also be used to formulate the mental or emotional state of an individual, studying his/her activity over a period of time, and possibly detect depression risks

Chapter 2

LITERATURE SURVEY AND PAPER REVIEW

Literature Survey: Is the process of analyzing, summarizing, organizing, and presenting novel conclusions from the results of technical review of large number of recently published scholarly articles. In this chapter we survey previous research done on automatic image annotation, we have studied about following papers published by some experts

2.1 Literature Survey

1. Twitter Sentimental Analysis

Author: Aliza Sarlan, Chayanit Nadam, Shuib Basri 2016

In this paper, author suggested an method of analyzing tweets using Natural Language Processing (NLP) for classifying tweets as positive, negative or neutral. The central idea of NLP techniques is based on machine learning and especially statistical learning which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules. Sentiment analysis has been handled as a Natural Language Processing denoted NLP, at many levels of granularity. Starting from being a document level classification task , it has been handled at the sentence level and more recently at the phrase level.

2. A Study of Sentiment Analysis of Twitter Data of tweets

Author: Abdullah Alsaeedi, Mohamad Zubair Khan

This paper briefly describes commonly used sentimental analysis methods including machine learning based approach and global and local feature extraction. This paper is organized as follows. Section 2 briefly describes commonly used twitter sentimental analysis methods including global and local feature extraction. Section 3 covers the lowlevel features which are generally extracted from tweets. Section 4 overviews the mostly used supervised learning methods for the classification or analyzing the sentences or tweets

These methods often rely on supervised classification approaches where sentiment detection is framed as a binary which are positive and negative. This approach requires labeled data to train classifiers. This approach, it becomes apparent that aspects of the local context of a word need to be taken into account such as negative (e.g. Not beautiful) and intensification (e.g. Very beautiful)

3.REAL TIME SENTIMENT ANALYSIS OF TWITTER POSTS

Author:Prakruthi V, Sindhu D, Dr S Anupama Kumar 2018

In paper the author discusses about the systems are designed to retrieve information using twitter data and then classify them based on the semantics of knowledge contained. Authors Lokmanyathilak Govindan Sankar Selvan and Teng-Sheng Moh have developed the framework in paper which makes use of real-time Twitter data stream, that are cleaned and analyzed and then fast feedback is acquired through opinion mining. Paper deals with the opinion extracted and collected from the popular social media platform named Twitter. For Comparison of market status of two enterprises they copy the two dictionary files one for positive and other for negative words from the repository in the backend as those files are used for analyzing and scoring terms from tweets. The paper mainly conveys about the sentimental analysis of tweets using R language which is helpful for collecting the sentimental information in the form of either positive score, negative score or someplace in between them. Then they execute the analysis of tweets that are in size of TBs which means big data using R language and Rhadoop Connector. In paper author has considered a way of advancing the present sarcasm detection algorithms by including improved pre-processing and text mining techniques like emoji and slang detection.

To analyze bulky live data streams, three considered key features are density distribution, negativity and influence. To do this in paper authors Ming Hao, Christian Rohrdantz and others have considered Pixel cell-based sentiment calendar where every opinion is represented in form of cell. The cell color is the sentiment value, i.e, green for positive, gray for natural and red for negative. In paper Web-based tool named SWAB (Social Web 30 Analysis Buddy) which is integrating qualitative analysis and large-scale data mining techniques together is proposed here. Prototype of this tool is demonstrated by analyzing matter posted by student on Twitter is demonstrated.

4.REAL TIME TWITTER SENTIMENT ANALYSIS USING 3-WAY CLASSIFIER

Author: Alaa S, Al Shammari 2018

In, authors used twitter to collect a corpus to preform linguistic analysis of the corpus and building a sentiment classifier using nave Bayes to determine the polarity of the documents. The experimental evaluation shows improvement of the proposed techniques than the previous methods. In, authors collect 3 different datasets (HASH, EMOT and iSieve) and do a pre-processing of it. Then, evaluating the training data which found that some features may not be good to be combined with other features. While, proposed a unigram model as baseline which gain 4 % at both binary classification and 3-way classification. Then, they investigated of both tree kernel model and feature based models which exceeded the unigram model. The combination of parts-of-speech and prior polarity features outperforms other features based on features analysis. In, authors focused at political sentiment analysis problem at real time using nave Bayes to classify tweets under 4 categories (positive, negative, neutral and unsure) where the system applicable to analyze tweet at election events. The method that used is generic which can be used in other domains such as movies events. While, studied the sentiment analysis problem for electronic products domain where tweet posts related to that field was analyzed. The machine learning techniques outperforms symbolic techniques based on sentiments identification. The evaluation of enhanced vector feature was tested using several classifiers such as nave Bayes, Maximum Entropy, SVM and Ensemble classifiers where the accuracy of these classifiers was almost similar. The performance of the proposed feature vector improved at the electronic products opinion domain. Another insignificant effort has been directed under the development of sentiment analysis systems to support the research community. Microsoft

azure developed a real-time twitter sentiment analysis demo that helps to represent the polarity of tweets of Bing maps where it is expressed as opinion as positive, negative or neutral

There are several classifiers used for opinion mining and sentiment analysis of microblogging platforms. I used Simple Voter algorithm and Naïve Bayes algorithm which classify tweets as positive, negative, or neutral opinion. The positive and negative dictionaries have been downloaded from the internet with total of 2014 positive words and 4783 negative words. In the proposed system, it works based on sentence level. So, each tweet decomposed into number of separated words. At this level, we score each tweet based on this equation:

Score = Number of positive words - Number of negative words

If Score > 0, then the sentence is positive.

If Score < 0, then the sentence is negative.

If Score = 0, then the sentence is neutral.

As shown above, each tweet will be tokenized into separated words where each word will be compared based on matching of positive dictionary term or even negative dictionary term. Then, the score will be assigned for each tweet based on the probability using Naïve Bayes algorithm or the high majority using Simple Voter algorithm. Therefore, the tweets will be classified into 3 categories (positive, negative, or neutral) based on the score.

5. EMOTION DETECTION AND ANALYSIS ON SOCIAL MEDIA

Author: Bharat Gaind, Varun Syal, Sneha Padgalwar 2019

Emotions are described as intense feelings that are directed at something or someone in response to internal or external events having a particular significance for the individual. And the internet, today, has become a key medium through which people express their emotions, feelings and opinions. Every event, news or activity around the world, is shared, discussed, posted and commented on social media, by millions of people.

In this paper, the authors address the problem of detection, classification and quantification of emotions of text in any form. They have considered English text collected from social media like Twitter, which can provide information having utility in a variety of ways, especially opinion

mining. Social media like Twitter and Facebook is full of emotions, feelings and opinions of people all over the world. However, analyzing and classifying text on the basis of emotions is a big challenge and can be considered as an advanced form of Sentiment Analysis. This paper proposes a method to classify text into six different Emotion-Categories: Happiness, Sadness, Fear, Anger, Surprise and Disgust. In our model, we use two different approaches and combine them to effectively extract these emotions from text. The approach is based on Natural Language Processing, and uses several textual features like emoticons, degree words and negations, Parts Of Speech and other grammatical analysis.

There are plenty of research works that have focussed on Sentiment Analysis and provide a 2-way classification of text. But few have actually focussed on mining emotions from text. However, machine analysis of text to classify and score it on the basis of emotions poses the following challenges:

- Lack of manually annotated data to train classifiers to label data into six categories.
- Unavailability of a comprehensive bag of Emotion-words labeled and scored according to Emotion-Categories (Happiness, Sadness, etc.) and their intensities, that can be used to detect Emotion-words in text

2.2 Paper Comparison

Sr. No.	Paper Title	Author (Pub. Year)	Description
1	Twitter Sentiment Analysis	Aliza Sarlan, Chayanit Nadam, Shuib Basri (2016)	In this paper, author suggested an NLP based technique for classification of the sentences from the dataset also known as tweets used In Twitter Natural Language Processing (NLP) for classifying tweets as positive, negative or neutral.
2	A Study of Sentiment Analysis of Twitter Data of tweets	Abdullah Alsaeedi, Mohamad Zubair Khan	This paper briefly describes commonly used sentimental analysis methods including machine learning based approach and global and local feature extraction. This paper is organized as follows. Section 2 briefly describes commonly used twitter sentimental analysis methods including global and local feature extraction. Section 3 covers the low-level features which are generally extracted from tweets.
3	REAL TIME SENTIMENT ANALYSIS OF TWITTER POSTS	Prakruti Sindhu Dr. Anupama K (2018)	In paper the author discusses about the systems are designed to retrieve information using twitter data and then classify them based on the semantics of knowledge contained. The paper mainly conveys about the sentimental analysis of tweets using R language which is helpful for collecting the sentimental information in the form of either positive, negative score or neutral.
4	Real time Twitter Sentiment Analysis using 3-way classifier/Lexicon based approach	Alaa S, Al Shammari (2018)	In this the author proposed system works based on sentence level. So, each tweet decomposed into number of separated words. At this level, we score each tweet based on this equation: Score = Number of positive words - Number of negative words If Score > 0, then the sentence is positive. If Score < 0, then the sentence is negative. If Score = 0, then the sentence is neutral.
5	Emotion Detection and Analysis on Social Media	Bharat Gaind Varun Syal Snehal Padgalwar (2019)	In this paper, we propose a method to classify and quantify tweets according to six standard emotions suggested..In this paper, we have addressed the problem of classifying text into the six basic Emotion-Categories, rather than just labeling them as positive or negative.

Figure 2.1: Comparison Table

Chapter 3

SURVEY OF METHODOLOGY

3.0.1 Naïve Bayes Classification

A classifier is a machine learning model that is used to discriminate different objects based on certain features. A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem, Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem

Types of Naive Bayes Classifier:

Multinomial Naive Bayes:

This is mostly used for document classification problem, i.e whether a document belongs to the category of sports, politics, technology etc. The features/predictors used by the classifier are the frequency of the words present in the document.

Bernoulli Naive Bayes:

This is similar to the multinomial naive bayes but the predictors are boolean variables. The parameters that we use to predict the class variable take up only values yes or no, for example if a word occurs in the text or not.

Gaussian Naive Bayes:

When the predictors take up a continuous value and are not discrete, we assume that these values are sampled from a gaussian distribution.

3.0.2 Natural Language Processing (NLP)

Natural Language Processing, usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language.The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable.Most NLP techniques rely on machine learning to derive meaning from human languages.

NLP entails applying algorithms to identify and extract the natural language rules such that the unstructured language data is converted into a form that computers can understand. When the text has been provided, the computer will utilize algorithms to extract meaning associated with every sentence and collect the essential data from them. Sometimes, the computer may fail to understand the meaning of a sentence well, leading to obscure results.

3.0.3 Support Vector Machine (SVM)

The objective of the support vector machine algorithm is to find a hyperplane in an Ndimensional space(N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3.

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of

the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

Chapter 4

REQUIREMENTS AND ANALYSIS

4.1 Problem Definition

Twitter Sentiment and Emotion Analysis is the automated process of analyzing text data and sorting it into sentiments positive, negative or neutral and emotions worry, happiness, love, hate. Performing Sentiment Analysis on data from Twitter using machine learning can help companies understand how people are talking about their brand. With more than 321 million active users, sending a daily average of 500 million Tweets, Twitter allows businesses to reach a broad audience and connect with customers without intermediaries. Monitoring Twitter allows companies to understand their audience, keep on top of what's being said about their brand and their competitors, and discover new trends in the industry. Are users talking positively or negatively about a product? Well, that's exactly what sentiment analysis determines but this 3 point analysis in present time is quite outdated so along with the sentiment some more specialised information is required to predict the outcome, Hence emotion analysis along with sentiment gives a perfect overview of the tweet, sentence or the text.

4.2 Requirements Specification

For implementation, in software we will require the following software and hardware specifications:

- **Software Specification**

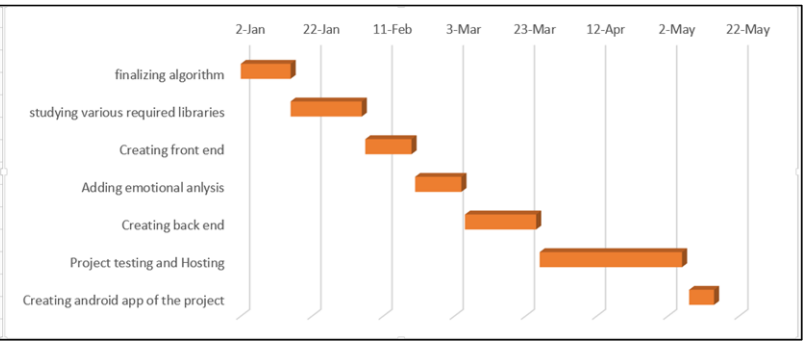
For implementation we will require applications such as, Jupiter Notebook, Python IDE, Django,android studio etc. That can be used to build an environment and to train machine based on it.

- **Hardware Specification**

To implement the project we will require a computer with specification such as multicore CPU, graphics card, hard disk upto 500GB, upto 8GB RAM. Input devices such as keyboard, optical mouse.

4.3 Planning and Scheduling

task	start date	end date	days
finalizing algorithm	2-Jan	15-Jan	14
studying various required libraries	16-Jan	05-Feb	20
Creating front end	6-Feb	19-Feb	13
Adding emotional anlysis	20-Feb	05-Mar	13
Creating back end	6-Mar	26-Mar	20
Project testing and Hosting	27-Mar	07-May	40
Creating android app of the project	8-May	14-May	7



Chapter 5

Implementation

5.1 Coding

```
import re
from textblob import TextBlob
from nltk.stem.wordnet import WordNetLemmatizer
import itertools
import numpy as np
import nltk

class sentiment_analysis_code():
lem = WordNetLemmatizer()
def cleaning(self, text):
 txt = str(text)
 txt = re.sub(r"http\S+", "", txt)
 if len(txt) == 0:
return 'no text'
else:
```

```
txt = txt.split()
index = 0
for j in range(len(txt)):
if txt[j][0] == '@':
index = j
txt = np.delete(txt, index)
if len(txt) == 0:
return 'no text'

else:
words = txt[0]
for k in range(len(txt)-1):
words+= " " + txt[k+1]
txt = words
txt = re.sub(r'[^\w]', ' ', txt)
if len(txt) == 0:
return 'no text'
else:
txt = ''.join(''.join(s)[:2] for _, s in itertools.groupby(txt))
txt = txt.replace("'", "")
txt = nltk.tokenize.word_tokenize(txt)

#data.content[i] = [w for w in data.content[i] if not w in stopset]

for j in range(len(txt)):
txt[j] = self.lem.lemmatize(txt[j], "v")
if len(txt) == 0:
return 'no text'
else:
return txt

def get_tweet_sentiment(self, tweet):

#cleaning of tweet
tweet = ' '.join(self.cleaning(tweet))
analysis = TextBlob(tweet)
if analysis.sentiment.polarity > 0:
return 'Positive'
elif analysis.sentiment.polarity == 0:
return 'Neutral'
```

```
else:
return 'Negative'
```

Emotion Analysis

```
import pandas as pd
import numpy as np
import nltk
import re
import pickle
import itertools
from nltk.stem.wordnet import WordNetLemmatizer
from django.conf import settings
import os

class emotion_analysis_code():
lem = WordNetLemmatizer()
def cleaning(self, text):
txt = str(text)
txt = re.sub(r"http\S+", "", txt)
if len(txt) == 0:
return 'no text'
else:
txt = txt.split()
index = 0
for j in range(len(txt)):
if txt[j][0] == '@':
index = j
```

```
txt = np.delete(txt, index)
if len(txt) == 0:
return 'no text'
else:
words = txt[0]
for k in range(len(txt)-1):
words+= " " + txt[k+1]
txt = words
txt = re.sub(r'[^\w]', ' ', txt)
if len(txt) == 0:
return 'no text'
else:

txt = ''.join(''.join(s)[:2] for _, s in itertools.groupby(txt))
txt = txt.replace("'", "")
txt = nltk.tokenize.word_tokenize(txt)
#data.content[i] = [w for w in data.content[i] if not w in stopset]
for j in range(len(txt)):
txt[j] = self.lem.lemmatize(txt[j], "v")
if len(txt) == 0:
return 'no text'
else:
return txt
def predict_emotion(self, tweet):
tweet_in_pandas = pd.Series(' '.join(self.cleaning(tweet)))
path_vec = os.path.join(settings.MODELS, 'vectorizer.pickle')
path_model = os.path.join(settings.MODELS, 'finalized_model.sav')
# load vectorizer
# vec_file = 'vectorizer.pickle'
vectorizer = pickle.load(open(path_vec, 'rb'))
# load trained model
# filename = 'finalized_model.sav'
model = pickle.load(open(path_model, 'rb'))

test = vectorizer.transform(tweet_in_pandas)
predicted_sentiment = model.predict(test)
final_sentiment = (predicted_sentiment[0])
```

```
if final_sentiment == 'worry':
return 'Worry'
elif final_sentiment == 'sadness':
return 'Sadness'
elif final_sentiment == 'happiness':
return 'Happiness'
elif final_sentiment == 'love':
return 'Love'
elif final_sentiment == 'hate':
return 'Hate'
```

Chapter 6

CONCLUSIONS

6.1 Conclusion

We have successfully developed, a system that can analyse twitter sentiments and emotions. Various experiments were being conducted using different methodologies, the best results are seen in the methods that are based on SVM in sentiment analysis and NLP for emotion analysis. We have studied, tested and implemented both of them. Both methodologies and algorithms gave a significant accuracy for the final model that was deployed using Django framework.

Bibliography

Journals and Conferences:

[1] M. Rambocas , and J. Gama, “Marketing Research :The Role of Sentiment Analysis”. The 5th SNA-KDD Workshop’11. University of Porto, 2013

[2] A. K. Jose, N. Bhatia, and S. Krishna, “Twitter Sentiment Analysis”. National Institute of Technology Calicut,2010.

[3] P. Lai, “Extracting Strong Sentiment Trendfrom Twitter”. Stanford University, 2012.

[4] M. Comesaña, A. P. Soares , M. Perea, A.P. Piñeiro, I. Fraga, and A. Pinheiro, “ Author ’ s personal copy Computers in Human Behavior ERP correlates of masked affective priming with emoticons,” Computers in Human Behavior, 29, 588–595, 2013. [6] A.H.Huang, D.C. Yen, & X. Zhang.

[5] 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.

[6] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R.Passonneau, “Sentiment Analysis of Twitter Data,” Annual International Conferences. New York:Columbia University, 2012.

[7] H. Saif, Y. He and H. Alani, “Alleviating Data Scarcity for Twitter Sentiment Analysis”. Association for Computational Linguistics, 2012.

[8] R. Prabowo, and M. Thelwall, “Sentiment Analysis:A Combined Approach,” International World Wide Web Conference Committee (IW3C2), 2009. United Kingdom: University of Wolverhamption.

[9] H. Saif, Y.He, and H. Alani, “Semantic Sentiment Analysis of Twitter,” Proceeding of the Workshop on Information Extraction and Entity Analytics on Social Media Data. United Kingdom: Knowledge Media Institute, 2011.

[10]J. Spencer and G. Uchyigit, “Sentiment or: Sentiment Analysis of Twitter Data,” Second Joint Conference on Lexicon and Computational Semantics. Brighton:University of Brighton, 2008

Websites:

https://ieeexplore.ieee.org/Xplore/home.jsp

https://en.wikipedia.org/ https://sci-hub.tw/

https://skymind.ai/wiki/neural-network

https://www.kaggle.com/langkilde/linear-svm-classification-of-sentiment-in-tweets

https://www.sciencedirect.com/science/article/pii/S1389041718300482