T.C.
BAHCESEHIR UNIVERSITY

TWITTER SENTIMENT ANALYSIS

Master Thesis

BİRCAN ENGÜLLÜ

ISTANBUL, 2018

T.C.
BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
COMPUTER ENGINEERING

TWITTER SENTIMENT ANALYSIS

Master Thesis

BİRCAN ENGÜLLÜ

Thesis Supervisor: Assist. Prof. TEVFİK AYTEKİN

ISTANBUL, 2018

THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
COMPUTER ENGINEERING

Name of the thesis: Twitter Sentiment Analysis
Name/Last Name of the Student: Bircan Engüllü
Date of the Defense of Thesis: 28.08.2018

The thesis has been approved by the Graduate School of Natural and Applied Sciences.

Signature
Assist. Prof. Yücel Batu SALMAN
Graduate School Director

I certify that this thesis meets all the requirements as a thesis for the degree of Master of Sciences.

Signature
Assist. Prof. Tarkan AYDIN
Program Coordinator

This is to certify that we have read this thesis and we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Sciences.

Examining Comittee Members	Signature
Thesis Supervisor Assist. Prof. Dr. Tevfik AYTEKİN	……..…………………………..
Member Assoc. Prof. Dr. M. Alper TUNGA	……..…………………………..
Member Assoc. Prof. Dr. Burcu TUNGA	……..…………………………..

ACKNOWLEDGEMENT

Firstly, I would like to thank my thesis advisor Assist. Prof. Tevfik Aytekin of the Computer Engineering department at Bahcesehir University for his advice and support throughout my master thesis.

I would also like to thank my family, especially for my parents Mesude Engüllü and Suat Engüllü for giving birth to me and for their endless love, support and encouragement during my life.

Last but not the least, I would also like to thank my wife Ebru Baycan Engüllü for her continuous love, support and patience despite her pregnancy period.

Finally, I would like to dedicate this thesis to my daughter who is going to come into the world probably at the end of the November and all of the children around the world to feel positive every time during their lives.

Istanbul, 2018 Bircan ENGÜLLÜ

ABSTRACT

TWITTER SENTIMENT ANALYSIS

Bircan Engüllü

Computer Engineering

Thesis Supervisor: Assist. Prof. Tevfik AYTEKİN

August 2018, 62 pages

Social media is an internet-based environment that gives people the opportinity of interacting with each other; constructing virtual networks and communities; sharing information, videos, photos, news, ideas etc. Twitter, Facebook, Instagram and Linkedin are some commonly used examples of social media platforms. After the usage of such social media platforms and data generated towards these platforms have made a notable increase in recent years, it has started to attract more and more attention in many areas such as business, politics and scientific researches for obtaining people opinions about some specific topics for future foresights.

This thesis is focused on the topic of Twitter sentiment analysis using supervised machine learning algorithms, such as Multinomial Naive Bayes, Support Vector Machine, Random Forest and Logistic Regression classifiers. Lexicon based techniques, such as Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis tool and my own Domain Based Lexicon approach were also used. These algorithms and techniques were applied on five well-known datasets with different properties, such as “Stanford Twitter Sentiment”, “Sentiment Strength Twitter” and “IMDB Movie Rewiews” datasets etc. to get evaluation results and investigate their performance.

Keywords: Twitter, Sentiment Analysis, Machine Learning, Lexicon Based

ÖZET

TWITTER DUYGU ANALİZİ

Bircan Engüllü

Bilgisayar Mühendisliği

Tez Danışmanı: Dr. Öğr. Üyesi Tevfik AYTEKİN

Ağustos 2018, 62 sayfa

Sosyal medya, insanlara birbirleriyle etkileşim kurma; sanal ağlar ve topluluklar inşa etme; bilgi, video, fotoğraf, haber ve fikirlerini paylaşma gibi olanaklar tanıyan internet tabanlı bir ortamdır. Twitter, Facebook, Instagram ve Linkedin yaygın bir şekilde kullanılan sosyal medya platformlarına örnek olarak gösterilebilir. Bu ve benzeri sosyal medya platformlarının ve bu platformlar aracılığı ile üretilen verilerin son yıllarda hatırı sayılır bir artış göstermesinin ardından, iş, politika ve bilimsel araştırmalar gibi alanlarda, gelecek ile ilgili öngörüler oluşturabilmek adına insanların belli konulardaki fikirlerini gözlemleyebilmek, gün geçtikçe daha fazla ilgi çeken bir olgu hâline gelmiştir.

Bu tez çalışmasında, gözetimli (supervised) öğrenme algoritmalarından Multinomial Naive Bayes, Support Vector Machine, Random Forest, Logistic Regression sınıflandırıcılarının yanı sıra, sözlük tabanlı (lexicon based) teknikler ile alakalı olarak Valence Aware Dictionary and Sentiment Reasoner (VADER) duygu analizi aracı ve kendi oluşturduğum Domain Based Lexicon yaklaşımı kullanılarak, Twitter duygu analizi konusuna odaklanılmıştır. Bahsedilen algoritma ve teknikler, "Stanford Twitter Sentiment", "Sentiment Strength Twitter" ve "IMDB Movie Reviews" gibi farklı özelliklere sahip beş iyi bilinen veri kümesi üzerinde uygulanmış, elde edilen değerlendirme sonuçları, algoritmaların performanslarını gözlemleyebilmek amacıyla kullanılmıştır.

Anahtar Kelimeler: Twitter, Duygu Analizi, Makine Öğrenmesi, Sözlük Tabanlı

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ACC	:	Accuracy
AI	:	Artificial Intelligence
API	:	Application Programming Interface
DBL	:	Domain Based Lexicon
FN	:	False Negative
FP	:	False Positive
F1	:	F-Measure
LR	:	Logistic Regression
MAP	:	Maximum a Posteriori
ML	:	Machine Learning
MT	:	Modified Tweet
MNB	:	Multinomial Naive Bayes
NLP	:	Natural Language Processing
NLTK	:	Natural Language Tool Kit
POS	:	Part-of-Speech
PREC	:	Precision
REC	:	Recall
RF	:	Random Forests
RT	:	Retweet
SA	:	Sentiment Analysis
STS	:	Stanford Twitter Sentiment
SVM	:	Support Vector Machines
TN	:	True Negative
TP	:	True Positive
TSA	:	Twitter Sentiment Analysis
VADER	:	Valence Aware Dictionary and Sentiment Reasoner

1. INTRODUCTION

“What other people think” has always been an important information for most of us in our lives during the decision-making process about some specific topics. Before the invention of World Wide Web, people asked their friends to recommend a restaurant, or to explain for whom they are going to vote in elections, or consulted some surveys to decide what products to buy etc. However, after the rise of web, now it is possible to find out about the opinions and experiences of many people that are both familiar to us and completely stranger to us.

In the early stages of the Web, most of its content was generated by owners of the websites and included traditional and well-defined information. Also, the content usually was about objective information than subjective one. After the rise of Web 2.0 platforms at the begining of 2000s (O’Reilly, 2007), not only the owners of the websites but also their users were started to generate content. Nowadays, most of the content is generated by users and usually is about subjective information than objective one. Especially, after the appearance of social network and microblogging platforms that give people the opportinity of interacting with each other, constructing virtual networks and communities and sharing information, videos, photos, news, ideas etc. subjective information generated by users has made a notable increase and this situation started to attract more attention in many areas such as business, politics and scientific researches for obtaining people opinions about some specific topics for future foresights.

Social media is a part of the web in which billions of people are active and interact with each other by sharing their stories, lifestyles, ideas and some other personal informations. There are many social media platforms such as Twitter, Facebook, Instagram, Google+, Linkedin etc. One of the most popular of them is Twitter, a social microblogging platform that allows users to write textual entries of up to 140 characters, commonly referred to as tweets. Hitz, L. & Blackburn, B., 2017 states that Twitter has over 328 million monthly active users, 100 million daily active users and 500 million tweets sent per day on 2017. As a consequence, data created by Twitter such as personal information of users (location, gender, age etc.) and twitter posts called as tweet that generated by these users when analyzed properly, provide a massive source of useful information. This has gave rise of a new sub-field of sentiment analysis (SA): Twitter sentiment analysis (TSA).

In this thesis, it was focused on the topic of twitter sentiment analysis using both the supervised machine learning algorithms, such as Multinomial Naive Bayes (MNB) classifier, Support Vector Machines (SVM) classifier, Random Forests (RF) classifier, Logistic Regression (LR) classifier, and lexicon based techniques, such as Valence Aware Dictionary and Sentiment Reasoner (VADER) and my own Domain Based Lexicon (DBL). Some well known datasets such as Stanford Twitter Sentiment dataset, Sentiment Strength Twitter dataset, US Airline dataset, Self-Driving Car dataset and IMDB Movie Rewiew dataset were used to investigate the performance of the these algorithms and techniques.

2. LITERATURE REVIEW

Sentiment analysis, or opinion mining, is the process of analayzing text context using Natural Language Processing (NLP) techniques, such as machine learning, lexicon-based and rule based approaches to obtain people opinions. After the usage of social media platforms and data generated towards by this usage has made notable increase in recent years, it is started to attract more attention in many areas such as product sales (Liu, 2006; Forman C. et al., 2008), stock returns (Das and Chen, 2007; Zimbra et al., 2015), political elections (Tumasjan et al., 2010; O'Connor et al., 2010) and scientific researches (Adrover C. et al., 2014; Kim D. & Kim J.W., 2014) in terms of obtaining people opinions about some specific topics for future foresights. Therefore, a sub area is rised, social media sentiment analysis or in our case it is Twitter Sentiment Analysis (TSA).

Sentiment analysis of twitter data when take the structure of tweets into account is a harder process than the sentiment analysis of conventional text such as review documents. Limitation of length in tweets and the user experiences of social media usage, lead to the common use of informal and irregular words in Twitter which make it harder to process. Therefore, besides researches have been made in the area of sentiment analysis there are also a large amount of researches have been made especially in the area of Twitter sentiment analysis which require some extra effort for twitter specific properties of the text.

Sentiment analysis tasks include classification of sentiment polarities expressed in text (e.g., positive, negative, neutral), identifying sentiment target/topic, opinion holder identification, and identifying sentiment for various aspects of a topic, product, or organization (Abbasi et al., 2008). Sentiment polarity classification is the most studied task because of its importance in social media analytics. The sentiment polarity classification problem is usually use a binary sentiment class approach as positive and negative classes, or three sentiment class approach as positive or negative or neutral classes. However, Jansen et al., 2009 and Ghiassi et al., 2016 have used five sentiment classes to target strong and weak positive and negative sentiments as a more sensitive approach.

Approaches applied for twitter sentiment analysis can be mainly categorized into two categories.

The first category is lexicon based approach which use a manually or automatically built lexicon or dictionary with sentiment polarities of words, some scoring methods and rules to evaluate sentiment (Turney, 2002; Taboada et al., 2011; Ding X. et al., 2008; Asgharl M.Z. et al., 2014). These methods can be applied on many classification tasks expecially on conventional text such as blogs, forums, and product reviews, but their performances as accurancy are limited on TSA, because they are unable to account for unconventional text, domain specific words, or irregular structures of the sentences or words.

The second category is machine learning approach to obtain a relationship between a number of feature values and sentiment via some machine learning algorithms [Pang et al., 2002; Go et al., 2009].

Finally, it is strongly recommended to have a glance at a really comprehensive article (Giachanou A. & Crestani F., 2016), which includes many resources that have been used in the Twitter sentiment analysis literature.

3. DATA AND METHODOLOGIES

3.1 DATA

In this thesis, five common datasets with different properties were used as explained below. All of these datasets have two columns, one for tweets or texts, and the other for binary sentiment polarity labels of corresponding tweets or texts as 1 (positive) and 0 (negative).

Stanford Twitter Sentiment Dataset (STANDFORD_DATASET): It is a Twitter dataset that contains 1.600.000 tweets from various areas along with their associated binary sentiment polarity labels. It was collected based on specific emoticons by searching the Twitter API by Go A. et al. 2009. It consists of 800.000 tweets with positive emoticons (such as :), :-), =), :D etc.), and 800.000 tweets with negative emoticons (such as :(, :-(etc.). Because of technical restrictions a sub set of this dataset was used as 80.000 positive and 80.000 negative tweets.

Sentiment Strength Twitter Dataset (SS_DATASET): It is a Twitter dataset that contains 4.242 tweets manually labelled with their positive and negative sentiment strengths which means that the value of both positive and negative labels can be between 1 (weak) and 5 (strong). The dataset was constructed by Thelwall M. et al. to evaluate a lexicon based method for sentiment strength detection. After re-labeling tweets in this dataset with sentiment labels of positive (Positive Sentiment Strength > 1.5 x Negative Sentiment Strength) and negative (Negative Sentiment Strength > 1.5 x Positive Sentiment Strength) to make it appropriate for binary classification rather than sentiment strengths, the dataset contains of 947 negative and 1.336 positive tweets finally.

U.S. Airline Twitter Dataset (AIRLINE_DATASET): It is a Twitter dataset about some major U.S. airline brands that scraped from February of 2015, in which contributors were asked to classify positive, negative, and neutral tweets (Data For

Everyone, 2015). After neutral tweets are eliminated to make it appropriate for binary classification, the dataset contains of 9.178 negative and 2.363 positive tweets finally.

Self-Driving Car Twitter Dataset (SD_CAR_DATASET): It is a Twitter dataset about self-driving cars of Google that scraped from June of 2015, in which contributors read tweets and classified them as very positive, slightly positive, neutral, slightly negative, or very negative (Data For Everyone, 2015). After neutral tweets are eliminated, slightly positive labels converted to positive labels and slightly negative labels converted to negative labels to make it appropriate for binary classification, the dataset contains of 795 negative and 1.902 positive tweets finally.

IMDB Movies Review Dataset (IMDB_DATASET): This dataset contains 50.000 movie reviews along with their associated binary sentiment polarity labels. It is intended to serve as a benchmark for sentiment classification by Maas A.L. et al. 2011. It consists of 25.000 tweets with positive polarity and 25.000 tweets with negative polarity.

3.1.1 Data Preprocesing

Preprocessing is the cleaning and preparation process of each individual text in the dataset before the classification process. It is a very important step for sentiment analysis, especially for TSA because tweets as unstructured texts contain significant amount of noise, or words that do not contain any meaning in terms of sentiment. According to Fayyad U.M. et al. 2013, the total percentage of noise in a dataset can reach 40 percent, which causes confusion in machine learning algorithms. Usage of informal language, abbreviations, slang words and punctuation signs to emphasize emotions are also preferred by twitter users intensely. As a result, it is not necessary to take all of the initial text into consideration in sentiment analysis and some words or terms of the initial text must be removed, replaced, or merged with others. This process contributes to the quality of data and enable algorithms to proceed faster and more accurately.

Let’s take a look at the preprocessing techniques used in this thesis with an example.

Table 3.1: Tweets before preprocessing

And obviously not forgetting the fact I'm gonna feel 122356325x worse :| Oh well...
RT @ihatequotes: True happiness comes from the effort of making others happy. Give and share your love everyday <3 #ihatequotes
OMG !!!!! THIS IS THE WEEEEEEK TO BUY SELECT MACBETH SHOES at 50 % OFF @MidwestTrader92 get in there for a crazy deal !!!!!!!!!!!!

3.1.1.1 Initial Preprocessing

This step of preprocessing phase includes converting all of the text to lowercase, repleacing any character followed by one or more times from the same character by twice of this character (e.g. haaaappppppy to haappy) and normalization of space, tab or new lines as a single white space.

Table 3.2: Tweets after initial preprocessing

and obviously not forgetting the fact i'm gonna feel 122356325x worse :| oh well..
rt @ihatequotes: true happiness comes from the effort of making others happy. give and share your love everyday <3 #ihatequotes
omg !! this is the week to buy select macbeth shoes at 50 % off @midwesttrader92 get in there for a crazy deal !!

3.1.1.2 Twitter Specific Preprocessing

This step of preprocessing phase includes removal of some twitter specific features, such as URLs, hashtags, mentions, retweet(RT) and modified tweet(MT) keywords etc.

Table 3.3: Tweets after twitter specific preprocessing

and obviously not forgetting the fact i'm gonna feel 122356325x worse :| oh well..
true happiness comes from the effo of making others happy. give and share your love everyday <3
omg !! this is the week to buy select macbeth shoes at 50 % off get in there for a crazy deal !!

3.1.1.3 Emoticon Handling

This step of preprocessing phase includes reducing the number of emoticons to only two categories: “EMOPOSITIVE” and “EMONEGATIVE” according to wikipedia list of emoticon topic.

Table 3.4: Tweets after emoticon handling

and obviously not forgetting the fact i'm gonna feel 122356325x worse EMONEGATIVE oh well..
true happiness comes from the effo of making others happy. give and share your love everyday EMOPOSITIVE
omg !! this is the week to buy select macbeth shoes at 50 % off get in there for a crazy deal !!

3.1.1.4 Slang Mapping

This step of preprocessing phase includes conversion of slang words, such as "thx" to "thank you", "omg" to "oh my god", "gonna" to "going to" etc. according to a slang dictionary scraped from noslang web site.

Table 3.5: Tweets after slang mapping

and obviously not forgetting the fact i'm going to feel 122356325x worse EMONEGATIVE oh well..
true happiness comes from the effo of making others happy. give and share your love everyday EMOPOSITIVE
oh my god !! this is the week to buy select macbeth shoes at 50 % off get in there for a crazy deal !!

3.1.1.5 Contraction Mapping

This step of preprocessing phase includes conversion of contraction words, such as "aren't" to "are not", "can't" to "can not", "they'll" to "they will", "would've" to "would have" etc. according to a contraction dictionary from “Yet Another Twitter Sentiment Analysis” Article, 2018.

Table 3.6: Tweets after contraction mapping

and obviously not forgetting the fact i am going to feel 122356325x worse EMONEGATIVE oh well..
true happiness comes from the effo of making others happy. give and share your love everyday EMOPOSITIVE
oh my god !! this is the week to buy select macbeth shoes at 50 % off get in there for a crazy deal !!

3.1.1.6 Parts-of-Speech Tagging (POS Tagging)

This step of preprocessing phase includes part-of-speech (POS) tagging of the words of the sentences. In other words, it is the process of tagging each word of the sentences in terms of what part of speech it belongs to, such as adjective, noun, verb, adverb etc.

In order to implement POS tagging GATE Twitter-POS-Tagger, 2013, which is customized for English tweets was used. Only for the IMDB_DATASET dataset which is not a twitter dataset, the pos tagger of NLTK library was used.

Table 3.7: Tweets after pos tagging

and_CC obviously_RB not_RB forgetting_VBG the_DT fact_NN i_PRP am_VBP going_VBG to_TO feel_VB x_UH worse_JJR EMONEGATIVE oh_UH well_UH .._:
true_JJ happiness_NN comes_VBZ from_IN the_DT effo_NN of_IN making_VBG others_NNS happy_JJ ._. give_VB and_CC share_VB your_PRP$ love_NN everyday_NN EMOPOSITIVE
oh_UH my_PRP$ god_NN !!_. this_DT is_VBZ the_DT week_NN to_TO buy_VB select_JJ macbeth_NNP shoes_NNS at_IN off_IN get_VB in_IN there_RB for_IN a_DT crazy_JJ deal_NN !!_.

3.1.1.7 Lemmatizing and Stemming

This step of preprocessing phase includes lemmatizing and stemming techniques which eliminate variations of base words, such as “go”, “going” and “goes” all are reduced to “go”. This step effectively decreasing entropy and increasing the relevance of the concept of the word.

Lemmatizing is a dictionary based technique which needs the POS tag of the words to proceed. On the other hand stemming is a rule based technique which do not require any POS tag. Therefore, lemmatizing technique was used if POS tagging included in preprocessing phase, or stemming otherwise.

In order to implement this techniques Porter Stemmer and WordNet Lemmatizer from NLTK library were used.

Table 3.8: Tweets after stemming

and obvious not forget the fact i am go to feel x wors EMONEGATIVE oh well ..
true happi come from the effo of make other happi . give and share your love everyday EMOPOSITIVE
oh my god ! ! thi is the week to buy select macbeth shoe at off get in there for a crazi deal ! !

Table 3.9: Tweets after lemmatizing

and_CC obviously_RB not_RB forget_VBG the_DT fact_NN i_PRP be_VBP go_VBG to_TO feel_VB x_UH bad_JJR EMONEGATIVE oh_UH well_UH .._:
true_JJ happiness_NN come_VBZ from_IN the_DT effo_NN of_IN make_VBG others_NNS happy_JJ ._. give_VB and_CC share_VB your_PRP$ love_NN everyday_NN EMOPOSITIVE
oh_UH my_PRP$ god_NN !!_. this_DT be_VBZ the_DT week_NN to_TO buy_VB select_JJ macbeth_NNP shoe_NNS at_IN off_IN get_VB in_IN there_RB for_IN a_DT crazy_JJ deal_NN !!_.

3.1.1.8 Negation Handling

This step of preprocessing phase deals with negations (e.g. “not good”). It is a critical step in sentiment analysis, becasuse negation words can influence most of the words around it, and ignoring negations is one of the main causes of misclassification.

To handle negation, the “not_” phrase was simply added to the begining of the first three words that take place just after the “not”, “n’t”, “no” and “never” words. If POS tagging is included in preprocessing phase it was also checked if these three words are verb, adjective, adverb or noun to add “not_” phrase.

Table 3.10: Tweets after negation handling

and obvious not not_forget not_the not_fact i am go to feel x wors EMONEGATIVE oh well ..
true happi come from the effo of make other happi . give and share your love everyday EMOPOSITIVE
oh my god ! ! thi is the week to buy select macbeth shoe at off get in there for a crazi deal ! !

Table 3.11: Tweets after negation handling (POS)

and_CC obviously_RB not_RB not_forget_VBG the_DT not_fact_NN i_PRP be_VBP go_VBG to_TO feel_VB x_UH bad_JJR EMONEGATIVE oh_UH well_UH .._:
true_JJ happiness_NN come_VBZ from_IN the_DT effo_NN of_IN make_VBG others_NNS happy_JJ ._. give_VB and_CC share_VB your_PRP$ love_NN everyday_NN EMOPOSITIVE
oh_UH my_PRP$ god_NN !!_. this_DT be_VBZ the_DT week_NN to_TO buy_VB select_JJ macbeth_NNP shoe_NNS at_IN off_IN get_VB in_IN there_RB for_IN a_DT crazy_JJ deal_NN !!_.

3.1.1.9 Punctation Handling

This step of preprocessing phase includes the removal of punctation marks, such as “.”, “,”, “?”, “!”, “_”, “)”, “(” etc.

Table 3.12: Tweets after punctation handling (before pos tagging)

and obviously not forgetting the fact i am going to feel x worse EMONEGATIVE oh well ..
true happiness comes from the effo of making others happy . give and share your love everyday EMOPOSITIVE
oh my god !! this is the week to buy select macbeth shoes at off get in there for a crazy deal !!

Table 3.13: Tweets after punctation handling (after pos tagging)

and obvious not forget the fact i am go to feel x wors EMONEGATIVE oh well
true happi come from the effo of make other happi give and share your love everyday EMOPOSITIVE
oh my god thi is the week to buy select macbeth shoe at off get in there for a crazi deal

Table 3.14: Tweets after punctation handling (NEG)

and obvious not not_forget not_the not_fact i am go to feel x wors EMONEGATIVE oh well
true happi come from the effo of make other happi give and share your love everyday EMOPOSITIVE
oh my god thi is the week to buy select macbeth shoe at off get in there for a crazi deal

Table 3.15: Tweets after punctation handling (POS)

and_CC obviously_RB not_RB forget_VBG the_DT fact_NN i_PRP be_VBP go_VBG to_TO feel_VB x_UH bad_JJR EMONEGATIVE oh_UH well_UH
true_JJ happiness_NN come_VBZ from_IN the_DT effo_NN of_IN make_VBG others_NNS happy_JJ give_VB and_CC share_VB your_PRP$ love_NN everyday_NN EMOPOSITIVE
oh_UH my_PRP$ god_NN this_DT be_VBZ the_DT week_NN to_TO buy_VB select_JJ macbeth_NNP shoe_NNS at_IN off_IN get_VB in_IN there_RB for_IN a_DT crazy_JJ deal_NN

Table 3.16: Tweets after punctation handling (POS + NEG)

and_CC obviously_RB not_RB not_forget_VBG the_DT not_fact_NN i_PRP be_VBP go_VBG to_TO feel_VB x_UH bad_JJR EMONEGATIVE oh_UH well_UH
true_JJ happiness_NN come_VBZ from_IN the_DT effo_NN of_IN make_VBG others_NNS happy_JJ give_VB and_CC share_VB your_PRP$ love_NN everyday_NN EMOPOSITIVE
oh_UH my_PRP$ god_NN this_DT be_VBZ the_DT week_NN to_TO buy_VB select_JJ macbeth_NNP shoe_NNS at_IN off_IN get_VB in_IN there_RB for_IN a_DT crazy_JJ deal_NN

3.1.1.10 Stopword Removal

This step of preprocessing phase includes the removal of stopwords are, such as “the”, “to”, “in”, “from”, “he”, “her”, “is”, “how”etc.

In order to implement this technique stopwords from NLTK library was used.

Table 3.17: Tweets after stopword removal

obvious not forget fact go feel x wors EMONEGATIVE oh well
true happi come effo make happi give share love everyday EMOPOSITIVE
oh god thi week buy select macbeth shoe get crazi deal

Table 3.18: Tweets after stopword removal (NEG)

obvious not not_forget not_the not_fact go feel x wors EMONEGATIVE oh well
true happi come effo make happi give share love everyday EMOPOSITIVE
oh god thi week buy select macbeth shoe get crazi deal

Table 3.19: Tweets after stopword removal (POS)

obviously_RB not_RB forget_VBG fact_NN go_VBG feel_VB x_UH bad_JJR EMONEGATIVE oh_UH well_UH
true_JJ happiness_NN come_VBZ effo_NN make_VBG others_NNS happy_JJ give_VB share_VB love_NN everyday_NN EMOPOSITIVE
oh_UH god_NN week_NN buy_VB select_JJ macbeth_NNP shoe_NNS get_VB crazy_JJ deal_NN

Table 3.20: Tweets after stopword removal (POS + NEG)

obviously_RB not_RB not_forget_VBG not_fact_NN go_VBG feel_VB x_UH bad_JJR EMONEGATIVE oh_UH well_UH
true_JJ happiness_NN come_VBZ effo_NN make_VBG others_NNS happy_JJ give_VB share_VB love_NN everyday_NN EMOPOSITIVE
oh_UH god_NN week_NN buy_VB select_JJ macbeth_NNP shoe_NNS get_VB crazy_JJ deal_NN

3.1.1.11 Final Preprocessing

This step of preprocessing phase includes removal of one length characters and normalization of space, tab or new lines as a single white space.

Table 3.21: Tweets after final preprocessing

obvious not forget fact go feel wors EMONEGATIVE oh well
true happi come effo make happi give share love everyday EMOPOSITIVE
oh god thi week buy select macbeth shoe get crazi deal

Table 3.22: Tweets after final preprocessing (NEG)

obvious not not_forget not_the not_fact go feel wors EMONEGATIVE oh well
true happi come effo make happi give share love everyday EMOPOSITIVE
oh god thi week buy select macbeth shoe get crazi deal

Table 3.23: Tweets after final preprocessing (POS)

obviously_RB not_RB forget_VBG fact_NN go_VBG feel_VB bad_JJR EMONEGATIVE oh_UH well_UH
true_JJ happiness_NN come_VBZ effo_NN make_VBG others_NNS happy_JJ give_VB share_VB love_NN everyday_NN EMOPOSITIVE
oh_UH god_NN week_NN buy_VB select_JJ macbeth_NNP shoe_NNS get_VB crazy_JJ deal_NN

Table 3.24: Tweets after final preprocessing (POS + NEG)

obviously_RB not_RB not_forget_VBG not_fact_NN go_VBG feel_VB bad_JJR EMONEGATIVE oh_UH well_UH
true_JJ happiness_NN come_VBZ effo_NN make_VBG others_NNS happy_JJ give_VB share_VB love_NN everyday_NN EMOPOSITIVE
oh_UH god_NN week_NN buy_VB select_JJ macbeth_NNP shoe_NNS get_VB crazy_JJ deal_NN

3.2 METHODOLOGIES

In this part, some general information about sentiment analysis and general approaches used to apply sentiment analyis will be explained.

3.2.1 Sentiment Analysis

Sentiment analysis, also called opinion mining, is a computational task of automatically determining what feelings an author is expressing in text. The sentiment of a text is usually represented by binary classes as positive and negative. It can also be represented in a more detailed way using multi classes, such as very positive, positive, neutral, negative, very negative etc.

Sentiment analysis is not a perfect solution, but when there are so much text to analyze it is very useful to give a general idea about these texts very quickly.

There are many approches to do sentiment analysis. Most of them use the same general idea, as follows:

i. Creating a list of words with their corresponding polarities, such as positive and negative.
ii. Counting the number of positive and negative words in the text.
iii. Analyzing. If there are more positive words then negative ones it indicates that the text has positive sentiment, negative sentiment otherwise.

Sentiment analysis process generally consists of following steps:

i. Data collection step includes the collection of data from desired resources, such as social media, blogs, forums etc.
ii. Data preprocessing step is the cleaning and preparation process of each individual text in the dataset before the classification process.
iii. Sentiment detection step is the elimination of objective information and only leaving the subjective information in the text.
iv. Sentiment classification step is the classification of subjective information as positive, negative etc.
v. Evalution of output step is the evalution of results by some graphs, charts, or some values such as accurancy, precision, recall, f-score, time etc.

Sentiment analysis can be mainly performed by three approaches:

i. Machine learning approach (especially supervised machine learning approach) is a way in which computers firstly learn or train from the given data using some machine learning algorithms. Then, the sentiment of a given text is predicted by using this knowledge.
ii. Lexicon based approach is a way in which the sentiment of a given text is calculated via a lexicon which includes pre-defined words and their corresponding polarity scores.
iii. Hybrid approach is a way in which the combination of the first two approach is used to predict the sentiment of a given text.

3.2.2 Machine Learning

As a sub field of artificial intelligence (AI), Machine Learning (ML) is the science of getting computers to learn and act like humans do, and improve their learning capabilities over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions (Faggella D., 2018).

Machine learning algorithms and traditional computational algorithms are both a part of computer science. While traditional computational algorithms use a set of instructions to do their job, machine learning algorithms instead additionally use some data as input and construct a model as output via some statistical analysis. Then, this model can be used for decision-making processes of some other given data inputs.

There are mainly two type of machine learning algorithms according to the way they learn about data: supervised learning and unsupervised learning.

3.2.2.1 Supervised learning

Supervised learning is a type of machine learning approach in which algorithms first learn from a training dataset. Training dataset is a part of complete dataset in which both the texts and their corresponding classes or labels are provided. After the learning process is completed these algorithms can predict the label of a new text.

Supervised learning can be divided into two sub part as classification and regression.

Classification: Classification is a technique to predict discrete outputs, such as whether a text is negative or positive, whether an email is spam or not etc.

Classification can be used if data can be categorized into specific classes. Some typical applications of classification include medical imaging, text classification, and credit scoring etc.

Commonly used algorithms to perform classification are naive bayes, support vector machines, decision trees, logistic regression, and neural networks algorithms.

Regression: Regression is a technique to predict continuous outputs, such as, changes in temperature, time until a process finish, etc.

Regression can be used if the nature of output is a real number. Some typical applications include forecasting, predicting house prices, and algorithmic trading etc.

Commonly used algorithms to perform regression are linear model, nonlinear model, decision trees and neural networks.

3.2.2.2 Unsupervised learning

Unsupervised learning is a type of machine learning approach in which there is no training phase unlike supervised learning. Unsupervised learning algorithms try to discover some unknown patterns or structures in the data.

Unsupervised learning can be divided into two sub part as association and clustering.

Clustering: Clustering is a technique to divide data into a number of groups or clusters according to their similarities, such as customer segments, type of species etc.

Commonly used algorithms to perform clustering are k-means, k-medoids, hierarchical clustering, Gaussian mixture models, hidden Markov models etc.

Association Rule: Association rule is a technique to discover interesting relations or rules of a given data, such as “Customers that buy bread also tend to buy cheese” etc.

Because the aim of this thesis is to make some prediction according to content of the tweets by classifying them as positive or negative, some common supervised learning algotithms in the area of text mining such as Multinomial Naive Bayes algorithm, Support Vector Machine algorithm, Random Forest algorithm and Logistic Regression algorithm will be used.

3.2.2.2.1 Multinomial Naive Bayes Algorithm

Naive Bayes algorithm is a simple probabilistic algorithm that uses Bayesian theorem with a strong, or naive assumption, that every feature is independent of the others, in order to predict the category of a given sample. Multinomial Naive Bayes algorithm is a variation of Naive Bayes algorithm which is usually used in text classification. It uses word counts rather than presence or absence of the words.

Let's take a look at how Multinomial Naive Bayes algorithm works and what kind of mathematical view there is behind the algorithm.

Multinomial Naive Bayes algorithm also uses Bayesian theorem to compute probability of features belong to a label as shown in Equation 3.1:

$$P(label|features) = \frac{P(label) * P\,(features|label)}{P\,(features)} \quad \textbf{(3.1)}$$

P(label | features): Posterior probability of features belong to a label

P(label): Prior probability of a label

P(features | label): Conditional probability of a feature belong to a label

P(features): Prior probability of a feature.

Because features are independent of the others, following Equation 3.2 can be used:

$$P\,(features|label) \approx P(f_1|label) * P(f_2|label) * \ldots * P(f_n|label) = \prod_{i=1}^{n} P(f_i|label)$$

$$P(label|features) = \frac{P(label) * \prod_{i=1}^{n} P(f_i|label)}{P\,(features)} \quad \textbf{(3.2)}$$

f_i: i^{th} individual feature.

Naive Bayes classifier uses the maximum a posteriori (MAP) estimation to define the most probable label $label_{map}$ as shown in Equation 3.3 (Manning C. D. et al. 2008):

$$label_{map} = arg \max_{label \in L} \left[\frac{\hat{P}(label) * \prod_{i=1}^{n} \hat{P}(f_i|label)}{\hat{P}(features)} \right] \quad \textbf{(3.3)}$$

After omitting denominator which is same for both positive and negative labels, equation is then can be written as in Equation 3.4.

$$label_{map} = arg \max_{label \in L} \left[\hat{P}(label) * \prod_{i=1}^{n} \hat{P}(f_i|label) \right] \quad \textbf{(3.4)}$$

By using logarithm property (log(xy) = log x + log y) multiplication of probabilities can be represented as the sum of logarithms, so the equation is now as in Equation 3.5.

$$label_{map} = arg \max_{label \in L} \left[\log \hat{P}(label) + \sum_{i=1}^{n} \hat{P}(f_i|label) \right] \quad \textbf{(3.5)}$$

If prior probability $\hat{P}$ is defined as in Equation 3.6 and conditional probability $\hat{P}(f_i \mid label)$ is defined as in Equation 3.7, final formula can be written as in Equation 3.8.

$$\hat{P}(label) = \frac{N_{label}}{N} \quad \textbf{(3.6)}$$

N_{label}: Number of features belong to a specific label in training dataset.

N: Total number of features in training dataset.

$$\hat{P}(f_i|label) = \frac{F_{i\,label}}{\sum_{i' \in V} F_{i'\,label}} \quad \textbf{(3.7)}$$

$F_{i\ label}$: Number of times the ith feature belong to a specific label occurs in the training dataset.

V: Dictionary of all unique features of corresponding label.

$$label_{map} = arg \max_{label \in L} \left[\log \frac{N_{label}}{N} + \sum_{i=1}^{n} \log \frac{F_{i\ label}}{\sum_{i' \in V} F_{i' label}} \right] \quad \textbf{(3.8)}$$

The psudeo code in Figure 3.1 shows how Multinomial Naive Bayes algorithm works during the training and testing phase.

Figure 3.1: Multinomial naive bayes algorithm – training and testing

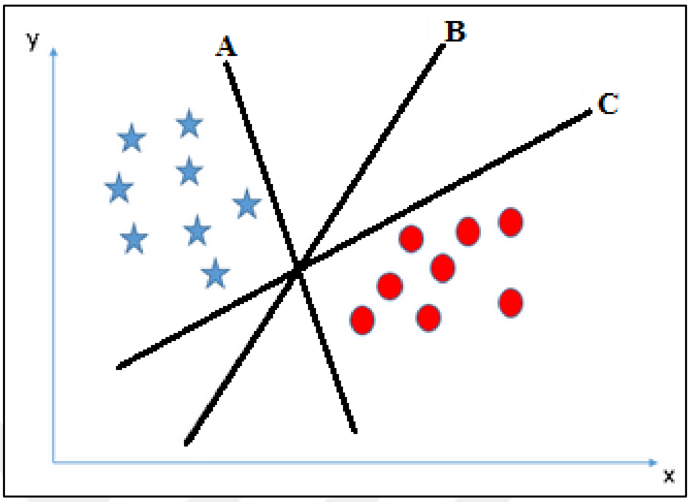
```
Algorithm 1 MUTLINOMIALNB
procedure TRAINMULTINOMIALNB(L, F)
    V ← ExtractVocabulary(F)
    N ← CountFeatures(F)
    for each label ∈ L do
        N_label ← CountFeatures(F)
        prior[c] ← N_label/N
        text_label ← ConcatTextOfAllFeatInLabel(F,label)
        for each f ∈ V do
            F_ilabel ← CountFrequencyOfFeat(text_label, f)
            for each f ∈ V do
                condprob[f][label] ← F_ilabel / Σ_{i'∈V} F_{i'label}
            end for
        end for
    end for
    return V, prior, condprob
end procedure

procedure APPLYMULTINOMIALNB(L, V, prior, condprob)
    W ← ExtractFeatures(V)
    for each label ∈ L do
        score_label ← logprior[label]
        for each f ∈ W do
            score_label+ = logcondprob[f][label]
        end for
    end for
    return argmax_{label∈L} score[label]
end procedure
```

3.2.2.2.2 Support Vector Machine Algorithm

Support Vector Machine algorithm is a supervised classification algorithm that can be used for both classification and regression problems. Basically, this algorithm tries to find a hyperplane that best divides a dataset into binary classes while performing classification.

Datasets that is linearly separable to binary classes, there are lots of possible linear separators as shown in Figure 3.2.

Figure 3.2: Hyperplanes that separate two classes



The linear seperator (B) drawn exactly in the middle of the data items of two classes is called as decision boundary. SVM is simply try to find a decision boundary that is most far away from any data point. The distance from the decision boundary to the closest data points determines the margin of the classifier. These closest data points are called as support vectors and they are the only deterministic points while specifying the decision boundary.

Figure 3.3: Illustration of linear SVM

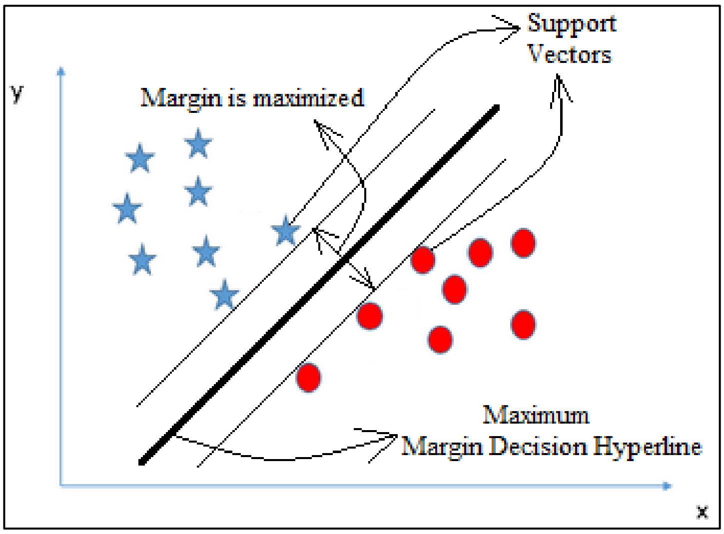
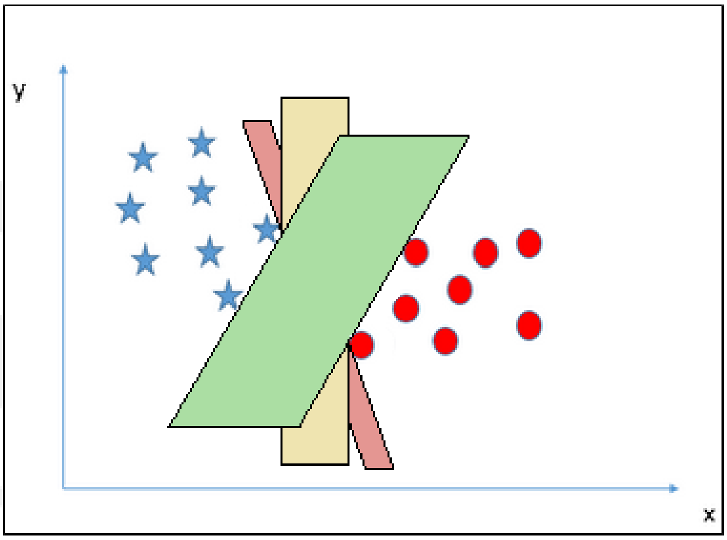


Figure 3.3 is a simple illustration of the margin and support vectors. Other data points have no effect in determining the decision boundary.

As a result SVM construct some hyperplanes while trying to maximize the margin and finally choose fattest one to make better classification decisions as shown in Figure 3.4.

Figure 3.4: A visualization for large-margin classification



Let's take a look at how Support Vector Machine algorithm works and what kind of mathematical view there is behind the algorithm.

A decision hyperplane is defined by an intercept term *b* and a weight vector *w* which is perpendicular to the hyperplane as shown in Equation 3.9.

$$\vec{w}^T\vec{x} = -b \qquad \textbf{(3.9)}$$

Because this algorithm handles classes as "+1" and "-1" rather than "1" and "0" the equation is now as shown in Equation 3.10:

$$f(\vec{x}) = \text{sign}(\vec{w}^T\vec{x} + b) \qquad \textbf{(3.10)}$$

By defining the functional margin of the ith example of x_i according to a hyperplane and constraining the value of this functional margin following equation can be written as shown in Equation 3.11.

$$y_i(\vec{w}^T \vec{x}_i + b) \geq 1 \quad \textbf{(3.11)}$$

Now we have an optimization problem called as quadratic problem (Burges C., 1998., Cristianini N. and Taylor J.S., 2000). By solving this problem the final formula is then as shown in Equation 3.12:

$$f(\vec{x}) = \mathrm{sign}(\textstyle\sum_i \alpha_i y_i \vec{x}_i^T \vec{x} + b) \quad \textbf{(3.12)}$$

3.2.2.2.3 Random Forest Algorithm

Random forest algorithm, as a combination of decision trees, is an ensemble method that can be used for classification and regression problems. Because decision trees are basic building blocks of a random forest algorithm, let’s take a look at decision trees first.

Decision tree algorithm is a supervised learning algorithm that can be used for classification and regression problems. In this algorithm, the dataset is broken down into smaller and smaller homogeneous subsets using most significant attributes, while at the same time a decision tree is built top-down from a root node. As a result a decision tree obtained with leaf nodes, or classes, or labels.

Choosing most significant attributes or in another words choosing the attributes which split dataset into two most homogenious or pure subset is the most important part of building decision trees. There are two commonly used techniques for this purpose:

Information Gain: The information gain is a technique based on the decrease in entropy that is used to construct a decision tree by finding an attribute returns the highest information gain value.

$$IG(A,S) = H(S) - \sum_{t \in T} p(t)H(t) \tag{3.13}$$

H(S): Entropy of dataset S

T: The subsets formed by splitting dataset S by attribute A.

P(t): Number of elements belong to subset t / Number of elements in dataset S

H(t): Entropy of subset t

Entropy is a measure which tell us if a subset of a dataset is homogeneous or not. If the subset is completely homogeneous the entropy value is zero. If the dataset is equally divided according to classes the entropy value has its maximum value, one. Entropy value is calculated as:

$$H(S) = \sum_{c \in C} -p(c) \log_2 p(c) \tag{3.14}$$

S: Dataset

C: Classes

P(c): Number of elements belong to class c / Number of elements in dataset S

Gini Gain: The gini gain is a technique based on the decrease in gini index that is used to construct a decision tree by finding the attribute returns the lowest gini gain value.

$$GG(A,S) = G(S) - \sum_{t \in T} p(t)\,G(t) \tag{3.15}$$

G(S): Gini Index value of dataset S

T: The subsets formed by splitting dataset S by attribute A.

P(t): Number of elements belong to subset t / Number of elements in dataset S

H(t): Gini Index value of subset t

Gini Index is a measure which tell us how often a randomly chosen element can be incorrectly classified according to class labels from the dataset. If a dataset has only one class, its gini index is zero, which means the dataset is completely homogeneous. If a dataset is equally divided according to classes, gini index has its maximum value. Gini Index value is calculated as:

$$G(S) = 1 - \sum_{c \in C} p(c)^2 \tag{3.16}$$

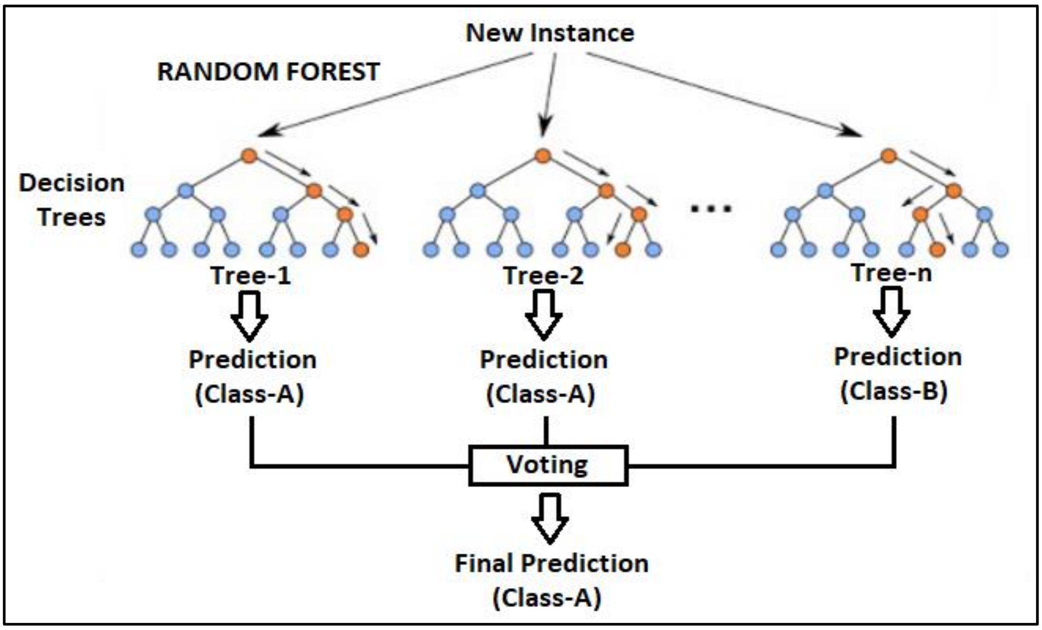
S: Dataset

C: Classes

P(c): Number of elements belong to class c / Number of elements in dataset S

After explaining desicion trees, let’s turn back to the Random forest algorithm. As I mention before random forest algorithm is a combination of decision trees. It selects random samples from the dataset, then construct a decision tree for each sample and get a prediction result from each decision tree. Finally, perform a vote for each predicted result and select the prediction result with the most votes as the final prediction as shown in Figure 3.5.

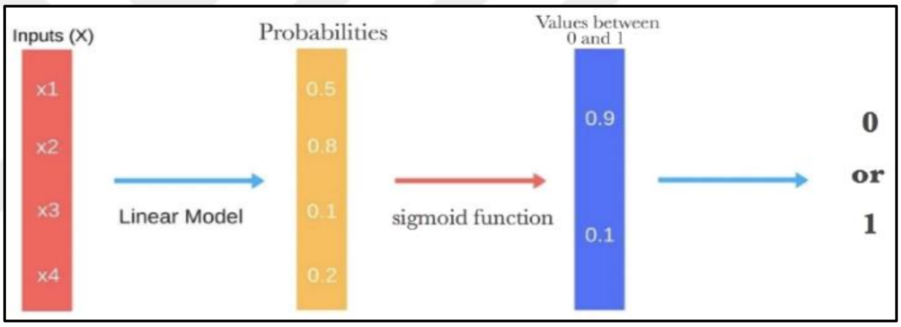
Figure 3.5: Illustration of random forest



3.2.2.2.4 Logistic Regression

Logistic regression is a linear model for classification rather than regression and expecially used for binary classification. It measures the relationship between some independent variables (or features) and a dependent variable (or label), by estimating probabilities. Then estimated probabilities is transformed into binary values to make a prediction. This transformation is done by the logistic function, also called the sigmoid function. The sigmoid function is an S-shaped curve which takes any real-valued number and map it into a number between the range of 0 and 1. Finally, these values between 0 and 1 is transformed into either 0 or 1 using a threshold classifier. Figure 3.6 shows a simple illustration of logistic regression algorithm.

Figure 3.6: Illustration of logistic regression



3.2.3 Lexicon-Based Approach

Lexicon-based approach relies on a lexicon (or a dictionary) which consist of a set of pairs of word and its pre-calculated polarity score. This approach calculates the sentiment of a given text from the polarity score of the words or phrases in that text as discussed in Turney P. D., 2002.

The general idea behind the lexicon-based approach is as follows:

i. Lexicon is constructed via words and their corresponding polarity scores.
ii. Preprocessing steps, such as lowercase convertion of text, stopwords removal, stemming or lemmatizing, negations handling etc. is applied to a new text.

iii. 1-gram or bag-of-words representation of new text is created to analize sentiment of the text.

iv. Sentiment score of new text is calculated by adding the polarity scores of the words to the total sentiment score of the new text if the words are found in the lexicon.

Lexicon construction is the most important part of this approach. There are some different ways of the lexicon construction process that explained in Kolchyna O. et al. and described below:

Hand-Tagged Lexicon Construction: This is a method that manually construct a lexicon and tags words in it as positive or negative. It is the most accurate but also the most time-consuming method. One of the commonly used hand-tagged lexicons is Multi-Perspective-Question-Answering (MPQA) opinion corpus constructed by Wiebe J. et al. consists of 4,850 words. Another one is SentiWordNet created by Esuli A. & Sebastiani F., consists of words extracted from WordNet, a large lexical database of English.

Lexicon Creation From Training Data: This is a supervised method that need a training dataset of labelled sentences. According to this method the labeled sentences from the training dataset are tokenized as words and polarity scores of these words are calculated according to the occurrence of each word in positive and negative sentences. This method can perform well in terms of accuracy because the words and their polarities are belong to a specific domain with a particular properties of text.

Extending a Small Lexicon Using Bootstrapping Techniques: This is a method in which firstly a lexicon created with a small set of opinion words manually. Then, as a dictionary-based method explained in Kim S. & Hovy E., 2004, this set is grown by searching for synonyms and antonyms of the words in a well known corpora. If such a word is founded it is added to the seed list then the next iteration starts. The iterative process stops when no word is remaining to add. As a corpus based method explained in Hatzivassiloglou V. & McKeown K., 1997, the initial set of opinion words are created from adjectives. Then, to extend initial lexicon new adjectives which were conjoined with the words (by “and”, “or”, “but”, “either-or” etc.) are added from the original lexicon.

3.2.3.1 Valence Aware Dictionary and Sentiment Reasoner (VADER):

VADER is a lexicon-based and rule-based sentiment analysis tool that is specifically developed to predict sentiments expressed in social media text (VADER, 2014). Hutto, C.J. & Gilbert, E.E., 2014 construct a generalizable and valance-based gold-standard sentiment lexicon for social microblogging platforms by using a combination of some qualitative and quantitative methods. Then, they combine these lexical features with five general rules that humans use when expressing or emphasizing sentiment intensity.

In this approach, each of the words in the lexicon have valence scores that indicate both the sentiment polarity as positive or negative, and the sentiment intensity on a scale from –4 to +4.

VADER analyses a text by checking if any of the words in the text are present in the lexicon. It also checks some rules that already discuseed above. Then, it produces four sentiment metrics from word ratings, as positive, neutral, negative, and compund scores. The first three, positive, neutral and negative, represent the proportion of the text that falls into those categories. The last one, compound score, is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most negative) and +1 (most positive).

3.2.3.2 Domain Based Lexicon Approach

Domain Based Lexicon (DBL) approach is my own implementation in which a lexicon is constructed via the labeled dataset in probabilistic manner.

Construction of the lexicon includes following steps:

i. Preprocessing steps, such as lowercase convertion of text, stopwords removal, stemming or lemmatizing, negations handling etc. is applied to texts.

ii. Tokenization is applied to texts to split them into words.
iii. The number of occurrences of each word in positive and negative sentences is calculated from the training dataset.
iv. The positive and negative polarity scores of each word are calculated according to the following formulas:

$$\text{Positive Polarity Score} = \frac{\text{\# of Occurrences in Positive Sentences}}{(\text{\# of Occurrences in Positive Sentences}) + (\text{\# of Occurrences in Negative Sentences})} \quad \textbf{(3.18)}$$

$$\text{Negative Polarity Score} = \frac{\text{\# of Occurrences in Negative Sentences}}{(\text{\# of Occurrences in Positive Sentences}) + (\text{\# of Occurrences in Negative Sentences})} \quad \textbf{(3.19)}$$

After the creation of the lexicon, for a given new sentence to predict its polarity includes following steps:

i. Preprocessing steps, such as lowercase convertion of text, stopwords removal, stemming or lemmatizing, negations handling etc. is applied to new sentence.
ii. Tokenization is applied to new sentence to split it into words.
iii. Positive and negative polarity scores are calculated by adding the correspoinding scores of individual words.
iv. Positive and negative polarity scores are normalized to have a value between 0 and 1.
v. If the normalized positive polarity score minus normalized negative polarity score is bigger than a given threshold it means that the new sentence is positive. Conversely, if the normalized negative polarity score minus normalized positive polarity score is bigger than a given threshold it means that the new sentence is negative.

3.2.4 Twitter

Twitter is a micro-blogging site that allows users to write textual entries of up to 140 characters, commonly referred to as tweets. Hitz, L. & Blackburn, B., 2017 states that Twitter has over 328 million monthly active users, 100 million daily active users and 500 million tweets sent per day on 2017.

Sentiment analysis of twitter data when take the structure of tweets into account is a harder process than the conventional text such as review documents. Limitation of length in tweets and the user experience of social media usage, lead to the common use of informal and irregular words such as abbreviations, slang words, misspelled words and emoticons.

Twitter has also some specific usages as stated below and shown in Figure 3.7:

i. Mention (@username): It is used for mentioning about a user or tagging a user.
ii. Reply (@username): It is used for replying to a tweet of another user.
iii. Retweet (RT): It is used for sharing the content of another tweet.
iv. Hashtag (#hashtag): It is used for specifying the content of the tweet is belog to a specific topic.

Figure 3.7: Structure of a typical tweet

Tweets published by public accounts can be retrieved using following Twitter APIs:

i. Twitter REST API: It is a rest search api which allows to search on historical data.
ii. The streaming API: It is a streaming api which allows to search in real-time for live data.

These Twitter APIs give the opportunity to retrieve tweets according to some specific criterias, such as hashtags, geographical locations, or time periods.

In Table 3.3, some attributes that each tweet contains can be seen:

Table 3.25: Tweet attributes overview

Attribute Name	**Attribute Description**
id	Unique tweet ID
created at	Creation date of tweet
favourite count	Number of favourite count of tweet
lang	Language
retweet count	Number of retweet count of tweet
entities	URLs, hashtags, user mentions
user: followers_count, friends_count	Information about user
text	Text of the tweet

By using these attributes many different data mining applications can be produced, but for the sentiment analyses text attribute matter the most. Based on the tweet text many algorithms or techniques can be performed for sentiment analysis if it contains some positive or negative opinions.

4. EXPERIMENTS AND RESULTS

This section describes execution of machine learning algorithms and lexicon-based techniques discussed in section 3.2 on five datasets with different properties as explained in section 3.1.

Programing languages, libraries and tools used on implementation phase are as follows:

Python: Python is an interpreted, object-oriented and high-level programming language especially prefered by data scientists.

Scikit-Learn: Scikit-learn is a machine learning library that provides some supervised and unsupervised learning algorithms for the Python programming language.

The Natural Language Toolkit (NLTK): NLTK is a machine learning and NLP library that provides some supervised and unsupervised learning algorithms, text processing facilities such as tokenization, stemming, lemmatizing, tagging etc. and many lexical resources and corporas for the Python programming language.

Pandas: Pandas is a library used to hold and analysis data for the Python programming language.

NumPy: NumPy is a library used for array and matrice operations for the Python programming language.

PyCharm: PyCharm is a Python integrated development environment (IDE).

Computer's hardware specifications used for execution of algorithms are as follows:

i. **CPU:** Intel Core i7-4700HQ CPU @ 2.40Ghz
ii. **RAM:** 16GB DDR3L 1600 MHz SDRAM
iii. **HDD:** 1TB HDD 5400 RPM

For the purposes of this research, experiments were performed on each of the datasets discussed in section 3.2 independetly. Datasets were seperated as 80 percent for training and 20 percent for testing. Every dataset has four combinations as follows:

i. BASIC: Dataset after basic preprocessing operations.
ii. BASIC + NEG: Dataset after basic preprocessings and negation handling operations.
iii. BASIC + POS: Dataset after basic preprocessings and pos tagging operations.
iv. BASIC + NEG + POS: Dataset after basic preprocessings, negation handling and pos tagging operations.

A vector-space model based on binary and tf-idf weighted word level N-Gram combinations such as;

i. 1-Gram (Unigram),
ii. 1-Gram (Unigram) + 2-Gram (Bigram),
iii. 1-Gram (Unigram) + 2-Gram (Bigram) + 3-Gram (Trigram)

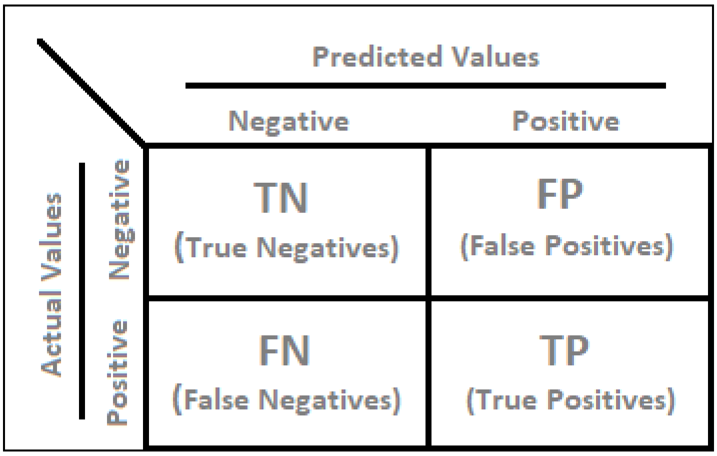
were used to represent dataset.

For machine learning approach, Scikit-Learn implementations of machine learning algorithms that are commonly used for text classification, such as Multinomial Naive Bayes (MNB), Support Vector Machines (SVM), Random Forest (RF) and Logistic Regression (LR) were applied.

For lexicon-based approach, a tweeter specific lexicon and rule-based sentiment analysis tool VADER and my own implementation of Domain Based Lexicon (DBL) were used.

For the evaluation of classification results, some well-known information retrieval measures are used. All of these evaluation measures described in this section use some basic counts of actual test values and predicted test values, such as counts of true positives (TP), counts of true negatives (TN), counts of false positives (FP) and counts of false negatives (FN) according to the class C. These measures show the relation between actual classes and predicted classes of the texts and can be summerized by a confisuaon matrix as shown in Figure 4.1.

Figure 4.1: Confusion matrix



Accuracy: Accuracy (ACC) is a measure that can be calculated by the division of the number of all correct predictions by the total number of the dataset. The accurancy value ranges between 0.0 (worst) and 1.0 (best). Formulation of accurancy is shown below in Equation 4.1.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} \tag{4.1}$$

Recall (Sensitivity or True positive rate): Recall (REC) is a measure that can be calculated by the division of the number of correct positive predictions by the total number of positives. The recall value ranges between 0.0 (worst) and 1.0 (best). Formulation of recall is shown below in Equation 4.2.

$$\mathrm{REC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = \frac{\mathrm{TP}}{\mathrm{P}} \tag{4.2}$$

Precision (Positive predictive value): Precision (PREC) is a measure that can be calculated by the division of the number of correct positive predictions by the total number of positive predictions. The precision value ranges between 0.0 (worst) and 1.0 (best). Formulation of precision is shown below in Equation 4.3.

$$\mathrm{PREC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4.3}$$

F-Measure: F-Measure is a harmonic mean of precision and recall. For the evaluation of experiments balanced F-Measure, or F1-Score is used, i.e. precision and recall are weighted equally. Formulation of F1-Score is shown below in Equation 4.4.

$$\text{F1-SCORE} = \frac{2 * \mathrm{PREC} * \mathrm{REC}}{\mathrm{PREC} + \mathrm{REC}} \tag{4.4}$$

At the following pages you can find the eveluation results.

Table 4.1: Evaluation results of MNB classifier on STANDFORD_DATASET

MNB	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7641	4.36	0.7739	0.7403	0.7567	15787	0.7554	0.7876	0.7712	16077
		(1+2)-Gram	0.7747	13.46	0.8004	0.7266	0.7617	15787	0.7538	0.8221	0.7865	16077
		(1-2-3)-Gram	0.7748	23.07	0.8039	0.7217	0.7606	15787	0.7516	0.8271	0.7876	16077
	TF-IDF	(1)-Gram	0.7553	5.20	0.7613	0.7374	0.7491	15787	0.7498	0.7730	0.7612	16077
		(1+2)-Gram	0.7758	13.84	0.8001	0.7298	0.7634	15787	0.7558	0.8210	0.7870	16077
		(1-2-3)-Gram	0.7782	25.81	0.7994	0.7374	0.7672	15787	0.7604	0.8183	0.7883	16077
BASIC+NEG	BINARY	(1)-Gram	0.7748	5.30	0.7754	0.7681	0.7717	15787	0.7744	0.7815	0.7779	16077
		(1+2)-Gram	0.7811	13.76	0.7975	0.7485	0.7722	15787	0.7671	0.8133	0.7895	16077
		(1-2-3)-Gram	0.7805	23.34	0.7995	0.7435	0.7705	15787	0.7644	0.8169	0.7898	16077
	TF-IDF	(1)-Gram	0.7700	5.65	0.7714	0.7615	0.7664	15787	0.7687	0.7784	0.7735	16077
		(1+2)-Gram	0.7816	13.79	0.7980	0.7488	0.7726	15787	0.7674	0.8138	0.7900	16077
		(1-2-3)-Gram	0.7838	26.16	0.7988	0.7535	0.7755	15787	0.7707	0.8136	0.7916	16077
BASIC+POS	BINARY	(1)-Gram	0.7686	5.17	0.7766	0.7493	0.7627	15804	0.7614	0.7877	0.7744	16051
		(1+2)-Gram	0.7771	14.78	0.7980	0.7375	0.7666	15804	0.7595	0.8161	0.7868	16051
		(1-2-3)-Gram	0.7759	23.72	0.7980	0.7343	0.7648	15804	0.7575	0.8170	0.7861	16051
	TF-IDF	(1)-Gram	0.7619	6.41	0.7680	0.7455	0.7566	15804	0.7564	0.7782	0.7672	16051
		(1+2)-Gram	0.7791	14.38	0.7977	0.7432	0.7695	15804	0.7631	0.8145	0.7880	16051
		(1-2-3)-Gram	0.7804	24.79	0.7961	0.7494	0.7720	15804	0.7667	0.8110	0.7882	16051
BASIC+NEG+POS	BINARY	(1)-Gram	0.7783	4.87	0.7817	0.7676	0.7746	15804	0.7752	0.7889	0.7820	16051
		(1+2)-Gram	0.7807	15.75	0.7946	0.7527	0.7731	15804	0.7685	0.8084	0.7880	16051
		(1-2-3)-Gram	0.7811	27.74	0.7969	0.7502	0.7728	15804	0.7675	0.8117	0.7890	16051
	TF-IDF	(1)-Gram	0.7741	6.41	0.7755	0.7665	0.7710	15804	0.7727	0.7816	0.7771	16051
		(1+2)-Gram	0.7845	15.28	0.7963	0.7603	0.7779	15804	0.7741	0.8085	0.7909	16051
		(1-2-3)-Gram	0.7863	28.26	0.7962	0.7654	0.7805	15804	0.7775	0.8071	0.7920	16051

Table 4.2: Summary of MNB classifier results on STANDFORD_DATASET

MNB SUMMARY	Max Value	Min Value	Average
Accuracy	0.7863	0.7553	0.7759
Time	28.26	4.36	15.05
Contribution of 2-Gram	2.0500%	0.2400%	1.0938%
Contribution of 3-Gram	0.2400%	-0.1200%	0.0800%
Contribution of TF-IDF	0.5200%	-0.8800%	-0.0058%
Contribution of Negation	1.4700%	0.5600%	0.8150%
Contribution of POS-Tagging	0.6600%	0.1100%	0.3350%
Contribution of NEG + POS	1.8800%	0.6000%	1.0350%

Evaluation results of Multinomial Naive Bayes classifier on different combinations of STANDFORD_DATASET are shown in Table 4.1. The maximum accuracy obtained is 78.63 percent but it takes 28.26 seconds to finish, and the minimum time obtained is 4.36 seconds but its accuracy is 76.41 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.2.

Table 4.3: Evaluation results of SVM classifier on STANDFORD_DATASET

SVM	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7592	21.70	0.7508	0.7696	0.7601	15787	0.7681	0.7491	0.7585	16077
		(1+2)-Gram	0.7768	49.26	0.7676	0.7883	0.7778	15787	0.7865	0.7656	0.7759	16077
		(1-2-3)-Gram	0.7799	89.05	0.7703	0.7919	0.7809	15787	0.7898	0.7682	0.7789	16077
	TF-IDF	(1)-Gram	0.7640	8.20	0.7558	0.7738	0.7647	15787	0.7726	0.7546	0.7635	16077
		(1+2)-Gram	0.7854	18.15	0.7906	0.7710	0.7807	15787	0.7805	0.7995	0.7899	16077
		(1-2-3)-Gram	0.7890	36.01	0.8024	0.7618	0.7816	15787	0.7772	0.8158	0.7960	16077
BASIC+NEG	BINARY	(1)-Gram	0.7687	18.28	0.7587	0.7819	0.7701	15787	0.7792	0.7557	0.7673	16077
		(1+2)-Gram	0.7803	44.22	0.7715	0.7910	0.7811	15787	0.7895	0.7700	0.7796	16077
		(1-2-3)-Gram	0.7842	86.79	0.7748	0.7957	0.7851	15787	0.7939	0.7729	0.7833	16077
	TF-IDF	(1)-Gram	0.7761	5.89	0.7693	0.7829	0.7760	15787	0.7831	0.7694	0.7762	16077
		(1+2)-Gram	0.7886	15.40	0.7922	0.7774	0.7847	15787	0.7854	0.7997	0.7925	16077
		(1-2-3)-Gram	0.7923	27.47	0.8027	0.7701	0.7861	15787	0.7829	0.8141	0.7982	16077
BASIC+POS	BINARY	(1)-Gram	0.7617	16.94	0.7531	0.7734	0.7631	15804	0.7708	0.7504	0.7604	16051
		(1+2)-Gram	0.7774	41.78	0.7675	0.7911	0.7791	15804	0.7879	0.7640	0.7757	16051
		(1-2-3)-Gram	0.7787	77.91	0.7692	0.7916	0.7802	15804	0.7887	0.7662	0.7773	16051
	TF-IDF	(1)-Gram	0.7703	5.96	0.7639	0.7775	0.7706	15804	0.7770	0.7634	0.7701	16051
		(1+2)-Gram	0.7873	18.83	0.7898	0.7787	0.7842	15804	0.7851	0.7959	0.7905	16051
		(1-2-3)-Gram	0.7895	31.28	0.7978	0.7711	0.7843	15804	0.7818	0.8076	0.7945	16051
BASIC+NEG+POS	BINARY	(1)-Gram	0.7692	22.64	0.7597	0.7825	0.7709	15804	0.7793	0.7563	0.7676	16051
		(1+2)-Gram	0.7807	48.53	0.7712	0.7936	0.7822	15804	0.7908	0.7682	0.7793	16051
		(1-2-3)-Gram	0.7824	83.45	0.7732	0.7945	0.7837	15804	0.7920	0.7706	0.7812	16051
	TF-IDF	(1)-Gram	0.7785	9.20	0.7711	0.7875	0.7792	15804	0.7863	0.7699	0.7780	16051
		(1+2)-Gram	0.7903	20.20	0.7920	0.7830	0.7875	15804	0.7887	0.7976	0.7931	16051
		(1-2-3)-Gram	0.7919	31.06	0.7978	0.7777	0.7876	15804	0.7864	0.8060	0.7961	16051

Table 4.4: Summary of SVM classifier results on STANDFORD_DATASET

SVM SUMMARY	Max Value	Min Value	Average
Accuracy	0.7923	0.7592	0.7793
Time	89.05	5.89	34.51
Contribution of 2-Gram	2.1400%	1.1500%	1.4888%
Contribution of 3-Gram	0.3900%	0.1300%	0.2638%
Contribution of TF-IDF	1.0800%	0.4800%	0.8667%
Contribution of Negation	1.2100%	0.3200%	0.5983%
Contribution of POS-Tagging	0.6300%	-0.1200%	0.1767%
Contribution of NEG + POS	1.4500%	0.2500%	0.6450%

Evaluation results of Support Vector Machine classifier on different combinations of STANDFORD_DATASET are shown in Table 4.3. The maximum accuracy obtained is 79.23 percent but it takes 27.47 seconds to finish, and the minimum time obtained is 5.89 seconds but its accuracy is 77.61 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.4.

Table 4.5: Evaluation results of RF classifier on STANDFORD_DATASET

RF	DATA REPRESENTATION			ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
						Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
	BASIC	BINARY	(1)-Gram	0.7456	370.04	0.7660	0.7006	0.7318	15787	0.7287	0.7899	0.7581	16077
			(1+2)-Gram	0.7497	2025.38	0.7615	0.7205	0.7404	15787	0.7393	0.7784	0.7584	16077
			(1-2-3)-Gram	0.7502	3809.25	0.7569	0.7305	0.7435	15787	0.7442	0.7696	0.7567	16077
		TF-IDF	(1)-Gram	0.7529	271.30	0.7771	0.7030	0.7382	15787	0.7333	0.8020	0.7661	16077
			(1+2)-Gram	0.7507	1680.67	0.7688	0.7107	0.7386	15787	0.7356	0.7901	0.7619	16077
			(1-2-3)-Gram	0.7449	4034.12	0.7583	0.7120	0.7345	15787	0.7332	0.7772	0.7546	16077
	BASIC+NEG	BINARY	(1)-Gram	0.7531	336.30	0.7685	0.7180	0.7424	15787	0.7399	0.7876	0.7630	16077
			(1+2)-Gram	0.7556	2039.92	0.7683	0.7256	0.7464	15787	0.7445	0.7852	0.7643	16077
			(1-2-3)-Gram	0.7547	3946.27	0.7641	0.7303	0.7469	15787	0.7462	0.7786	0.7621	16077
		TF-IDF	(1)-Gram	0.7541	244.64	0.7756	0.7088	0.7407	15787	0.7364	0.7986	0.7662	16077
			(1+2)-Gram	0.7517	1852.84	0.7630	0.7237	0.7428	15787	0.7418	0.7793	0.7601	16077
			(1-2-3)-Gram	0.7457	3884.54	0.7526	0.7251	0.7386	15787	0.7394	0.7660	0.7525	16077
	BASIC+POS	BINARY	(1)-Gram	0.7480	452.75	0.7671	0.7067	0.7357	15804	0.7320	0.7887	0.7593	16051
			(1+2)-Gram	0.7538	2408.25	0.7627	0.7313	0.7467	15804	0.7457	0.7760	0.7606	16051
			(1-2-3)-Gram	0.7506	3652.98	0.7525	0.7411	0.7468	15804	0.7489	0.7600	0.7544	16051
		TF-IDF	(1)-Gram	0.7531	387.76	0.7726	0.7120	0.7411	15804	0.7368	0.7937	0.7642	16051
			(1+2)-Gram	0.7523	1994.30	0.7666	0.7199	0.7425	15804	0.7398	0.7842	0.7614	16051
			(1-2-3)-Gram	0.7479	3415.41	0.7558	0.7267	0.7410	15804	0.7407	0.7689	0.7545	16051
	BASIC+NEG+POS	BINARY	(1)-Gram	0.7480	326.31	0.7595	0.7204	0.7394	15804	0.7380	0.7753	0.7562	16051
			(1+2)-Gram	0.7546	2079.43	0.7603	0.7383	0.7491	15804	0.7495	0.7708	0.7600	16051
			(1-2-3)-Gram	0.7522	4133.57	0.7471	0.7568	0.7519	15804	0.7574	0.7478	0.7526	16051
		TF-IDF	(1)-Gram	0.7559	370.62	0.7724	0.7205	0.7455	15804	0.7418	0.7909	0.7656	16051
			(1+2)-Gram	0.7526	2346.92	0.7668	0.7206	0.7430	15804	0.7403	0.7842	0.7616	16051
			(1-2-3)-Gram	0.7478	4239.52	0.7562	0.7258	0.7407	15804	0.7403	0.7697	0.7547	16051

Table 4.6: Summary of RF classifier results on STANDFORD_DATASET

RF SUMMARY	Max Value	Min Value	Average
Accuracy	0.7559	0.7449	0.7511
Time	4239.52	244.64	2095.96
Contribution of 2-Gram	0.6600%	-0.3300%	0.1286%
Contribution of 3-Gram	0.0500%	-0.6000%	-0.3375%
Contribution of TF-IDF	0.7900%	-0.9000%	-0.0541%
Contribution of Negation	0.7500%	0.0800%	0.3482%
Contribution of POS-Tagging	0.4100%	0.0190%	0.1948%
Contribution of NEG+POS	0.4900%	0.1900%	0.2848%

Evaluation results of Random Forest classifier on different combinations of STANDFORD_DATASET are shown in Table 4.5. The maximum accuracy obtained is 75.59 percent but it takes 370.62 seconds to finish, and the minimum time obtained is 244.64 seconds but its accuracy is 75.41 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.6.

Table 4.7: Evaluation results of LR classifier on STANDFORD_DATASET

LR	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7435	416.78	0.7662	0.6942	0.7284	15787	0.7251	0.7919	0.7570	16077
		(1+2)-Gram	0.7486	2559.02	0.7622	0.7161	0.7384	15787	0.7368	0.7806	0.7581	16077
		(1-2-3)-Gram	0.7500	4747.00	0.7621	0.7205	0.7407	15787	0.7395	0.7791	0.7588	16077
	TF-IDF	(1)-Gram	0.7479	303.60	0.7734	0.6948	0.7320	15787	0.7275	0.8001	0.7621	16077
		(1+2)-Gram	0.7506	2138.38	0.7687	0.7107	0.7385	15787	0.7355	0.7899	0.7618	16077
		(1-2-3)-Gram	0.7387	4342.02	0.7575	0.6951	0.7250	15787	0.7230	0.7815	0.7511	16077
BASIC + NEG	BINARY	(1)-Gram	0.7474	391.41	0.7657	0.7065	0.7349	15787	0.7321	0.7878	0.7589	16077
		(1+2)-Gram	0.7549	2657.41	0.7676	0.7250	0.7457	15787	0.7439	0.7844	0.7636	16077
		(1-2-3)-Gram	0.7534	4627.92	0.7607	0.7331	0.7466	15787	0.7469	0.7735	0.7600	16077
	TF-IDF	(1)-Gram	0.7576	384.41	0.7799	0.7117	0.7443	15787	0.7393	0.8028	0.7697	16077
		(1+2)-Gram	0.7514	2263.79	0.7680	0.7139	0.7400	15787	0.7373	0.7883	0.7619	16077
		(1-2-3)-Gram	0.7524	4438.37	0.7636	0.7248	0.7437	15787	0.7426	0.7796	0.7607	16077
BASIC + POS	BINARY	(1)-Gram	0.7488	519.60	0.7663	0.7105	0.7373	15804	0.7340	0.7866	0.7594	16051
		(1+2)-Gram	0.7521	2891.38	0.7593	0.7325	0.7457	15804	0.7455	0.7714	0.7582	16051
		(1-2-3)-Gram	0.7530	4862.08	0.7606	0.7329	0.7465	15804	0.7461	0.7728	0.7592	16051
	TF-IDF	(1)-Gram	0.7537	418.66	0.7754	0.7089	0.7407	15804	0.7357	0.7978	0.7655	16051
		(1+2)-Gram	0.7520	2497.25	0.7698	0.7136	0.7406	15804	0.7369	0.7899	0.7624	16051
		(1-2-3)-Gram	0.7495	4756.31	0.7633	0.7178	0.7398	15804	0.7375	0.7808	0.7586	16051
BASIC + NEG + POS	BINARY	(1)-Gram	0.7494	549.76	0.7627	0.7186	0.7400	15804	0.7379	0.7799	0.7583	16051
		(1+2)-Gram	0.7549	3122.15	0.7649	0.7306	0.7474	15804	0.7460	0.7788	0.7620	16051
		(1-2-3)-Gram	0.7484	5352.82	0.7503	0.7388	0.7445	15804	0.7467	0.7580	0.7523	16051
	TF-IDF	(1)-Gram	0.7567	447.21	0.7746	0.7189	0.7457	15804	0.7415	0.7940	0.7669	16051
		(1+2)-Gram	0.7541	2875.06	0.7648	0.7283	0.7461	15804	0.7445	0.7795	0.7616	16051
		(1-2-3)-Gram	0.7495	5226.79	0.7574	0.7285	0.7427	15804	0.7424	0.7702	0.7561	16051

Table 4.8: Summary of LR classifier results on STANDFORD_DATASET

LR SUMMARY	Max Value	Min Value	Average
Accuracy	0.7576	0.7387	0.7508
Time	5352.82	303.60	2616.22
Contribution of 2-Gram	0.7500%	-0.6200%	0.1700%
Contribution of 3-Gram	0.1400%	-1.1900%	-0.2963%
Contribution of TF-IDF	1.0200%	-1.1300%	0.0808%
Contribution of Negation	1.3700%	0.0800%	0.6300%
Contribution of POS-Tagging	1.0800%	0.1400%	0.4967%
Contribution of NEG + POS	1.0800%	-0.1600%	0.5617%

Evaluation results of Linear Regression classifier on different combinations of STANDFORD_DATASET are shown in Table 4.7. The maximum accuracy obtained is 75.76 percent but it takes 384.41 seconds to finish, and the minimum time obtained is 303.60 seconds but its accuracy is 74.79 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.8.

Table 4.9: Evaluation results of VADER SA tool on STANDFORD_DATASET

VDR	DATA REPRESENTATION	ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
				Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
NO_PP	(1)-Gram	0.7128	0.00	0.6628	0.8609	0.7490	11507	0.8043	0.5662	0.6646	11619
BASIC	(1)-Gram	0.6856	0.00	0.6298	0.8412	0.7203	10371	0.7860	0.5413	0.6411	11178

Evaluation results of VADER sentiment analysis tool on different combinations of STANDFORD_DATASET are shown in Table 4.9. The best accuracy obtained is 71.28 percent and time value is zero because there is no training step.

Table 4.10: Evaluation results of DBL classifier on STANDFORD_DATASET

DBL	DATA REPRESENTATION	ACCURACY	TRAINING TIME (*)	POSITIVE				NEGATIVE			
				Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	(1)-Gram	0.7562	43.24	0.8037	0.6669	0.7290	14949	0.7236	0.8426	0.7786	15468
BASIC NEG	(1)-Gram	0.7675	46.05	0.8097	0.6901	0.7451	15109	0.7370	0.8427	0.7863	15574
BASIC POS	(1)-Gram	0.7608	41.30	0.7923	0.6988	0.7426	15083	0.7366	0.8214	0.7767	15468
BASIC POSNEG	(1)-Gram	0.7704	44.70	0.7974	0.7189	0.7562	15239	0.7488	0.8210	0.7832	15547

Table 4.11: Summary of DBL classifier results on STANDFORD_DATASET

DBL SUMMARY	Max Value	Min Value	Average
Accuracy	0.7704	0.7562	0.7637
Time	46.05	41.30	43.82
Contribution of Negation	1.1300%	1.1300%	1.1300%
Contribution of POS-Tagging	0.4600%	0.4600%	0.4600%
Contribution of NEG + POS	1.4200%	1.4200%	1.4200%

Evaluation results of my own Domain Based Lexicon implementation on different combinations of STANDFORD_DATASET are shown in Table 4.10. The maximum accuracy obtained is 77.04 percent but it takes 44.70 seconds to finish, and the minimum time obtained is 41.30 seconds but its accuracy is 76.08 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.11.

Table 4.12: Evaluation results of MNB classifier on SS_DATASET

MNB	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7615	0.15	0.7815	0.8459	0.8124	279	0.7226	0.6292	0.6727	178
		(1+2)-Gram	0.7637	0.31	0.7697	0.8746	0.8188	279	0.7500	0.5899	0.6604	178
		(1-2-3)-Gram	0.7637	0.45	0.7631	0.8889	0.8212	279	0.7652	0.5674	0.6516	178
	TF-IDF	(1)-Gram	0.7309	0.13	0.7143	0.9319	0.8087	279	0.7957	0.4157	0.5461	178
		(1+2)-Gram	0.7090	0.27	0.6891	0.9534	0.8000	279	0.8169	0.3258	0.4659	178
		(1-2-3)-Gram	0.6937	0.47	0.6733	0.9677	0.7941	279	0.8393	0.2640	0.4017	178
BASIC + NEG	BINARY	(1)-Gram	0.7746	0.11	0.7876	0.8638	0.8239	279	0.7483	0.6348	0.6869	178
		(1+2)-Gram	0.7702	0.30	0.7843	0.8602	0.8205	279	0.7417	0.6292	0.6809	178
		(1-2-3)-Gram	0.7681	0.49	0.7781	0.8674	0.8203	279	0.7466	0.6124	0.6728	178
	TF-IDF	(1)-Gram	0.7287	0.12	0.7171	0.9176	0.8050	279	0.7700	0.4326	0.5540	178
		(1+2)-Gram	0.7090	0.28	0.6891	0.9534	0.8000	279	0.8169	0.3258	0.4659	178
		(1-2-3)-Gram	0.7155	0.48	0.6886	0.9749	0.8071	279	0.8871	0.3090	0.4583	178
BASIC + POS	BINARY	(1)-Gram	0.7374	0.11	0.7573	0.8387	0.7959	279	0.6959	0.5787	0.6319	178
		(1+2)-Gram	0.7352	0.30	0.7484	0.8530	0.7973	279	0.7050	0.5506	0.6183	178
		(1-2-3)-Gram	0.7243	0.45	0.7368	0.8530	0.7907	279	0.6940	0.5225	0.5962	178
	TF-IDF	(1)-Gram	0.7046	0.11	0.6957	0.9176	0.7913	279	0.7416	0.3708	0.4944	178
		(1+2)-Gram	0.7068	0.25	0.6835	0.9677	0.8012	279	0.8548	0.2978	0.4417	178
		(1-2-3)-Gram	0.6980	0.33	0.6741	0.9785	0.7982	279	0.8846	0.2584	0.4000	178
BASIC + NEG + POS	BINARY	(1)-Gram	0.7352	0.11	0.7565	0.8351	0.7939	279	0.6913	0.5787	0.6300	178
		(1+2)-Gram	0.7243	0.29	0.7398	0.8459	0.7893	279	0.6884	0.5337	0.6013	178
		(1-2-3)-Gram	0.7243	0.50	0.7368	0.8530	0.7907	279	0.6940	0.5225	0.5962	178
	TF-IDF	(1)-Gram	0.7068	0.11	0.6975	0.9176	0.7926	279	0.7444	0.3764	0.5000	178
		(1+2)-Gram	0.7221	0.25	0.6939	0.9749	0.8107	279	0.8923	0.3258	0.4774	178
		(1-2-3)-Gram	0.7068	0.39	0.6808	0.9785	0.8029	279	0.8929	0.2809	0.4274	178

Table 4.13: Summary of MNB classifier results on SS_DATASET

MNB SUMMARY	Max Value	Min Value	Average
Accuracy	0.7746	0.6937	0.7298
Time	0.50	0.11	0.28
Contribution of 2-Gram	1.5317%	-2.1882%	-0.4923%
Contribution of 3-Gram	0.6565%	-1.5317%	-0.5744%
Contribution of TF-IDF	-0.2188%	-7.0022%	-3.7564%
Contribution of Negation	2.1882%	-0.2188%	0.7294%
Contribution of POS-Tagging	0.4376%	-3.9387%	-1.9329%
Contribution of NEG + POS	1.3129%	-3.9387%	-1.7141%

Evaluation results of Multinomial Naive Bayes classifier on different combinations of SS_DATASET are shown in Table 4.12. The maximum accuracy obtained is 77.46 percent but it takes 0.11 seconds to finish, and the minimum time obtained is 0.11 seconds but its accuracy is 73.52 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.13.

Table 4.14: Evaluation results of SVM classifier on SS_DATASET

SVM	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7330	0.16	0.7698	0.8029	0.7860	279	0.6687	0.6236	0.6453	178
		(1+2)-Gram	0.7571	0.31	0.7727	0.8530	0.8109	279	0.7248	0.6067	0.6606	178
		(1-2-3)-Gram	0.7593	0.48	0.7666	0.8710	0.8154	279	0.7429	0.5843	0.6541	178
	TF-IDF	(1)-Gram	0.7484	0.14	0.7828	0.8136	0.7979	279	0.6886	0.6461	0.6667	178
		(1+2)-Gram	0.7702	0.36	0.7979	0.8351	0.8161	279	0.7212	0.6685	0.6939	178
		(1-2-3)-Gram	0.7702	0.53	0.7979	0.8351	0.8161	279	0.7212	0.6685	0.6939	178
BASIC+NEG	BINARY	(1)-Gram	0.7396	0.16	0.7759	0.8065	0.7909	279	0.6766	0.6348	0.6551	178
		(1+2)-Gram	0.7637	0.33	0.7785	0.8566	0.8157	279	0.7333	0.6180	0.6707	178
		(1-2-3)-Gram	0.7549	0.50	0.7651	0.8638	0.8114	279	0.7324	0.5843	0.6500	178
	TF-IDF	(1)-Gram	0.7571	0.14	0.7937	0.8136	0.8035	279	0.6959	0.6685	0.6819	178
		(1+2)-Gram	0.7768	0.39	0.8062	0.8351	0.8204	279	0.7262	0.6854	0.7052	178
		(1-2-3)-Gram	0.7746	0.47	0.7973	0.8459	0.8209	279	0.7329	0.6629	0.6962	178
BASIC+POS	BINARY	(1)-Gram	0.7309	0.17	0.7583	0.8208	0.7883	279	0.6774	0.5899	0.6306	178
		(1+2)-Gram	0.7527	0.33	0.7660	0.8566	0.8088	279	0.7241	0.5899	0.6502	178
		(1-2-3)-Gram	0.7615	0.50	0.7623	0.8853	0.8192	279	0.7594	0.5674	0.6495	178
	TF-IDF	(1)-Gram	0.7374	0.18	0.7695	0.8136	0.7909	279	0.6790	0.6180	0.6471	178
		(1+2)-Gram	0.7659	0.34	0.7945	0.8315	0.8126	279	0.7152	0.6629	0.6880	178
		(1-2-3)-Gram	0.7659	0.52	0.7905	0.8387	0.8139	279	0.7205	0.6517	0.6844	178
BASIC+NEG+POS	BINARY	(1)-Gram	0.7440	0.17	0.7755	0.8172	0.7958	279	0.6871	0.6292	0.6569	178
		(1+2)-Gram	0.7637	0.31	0.7785	0.8566	0.8157	279	0.7333	0.6180	0.6707	178
		(1-2-3)-Gram	0.7593	0.41	0.7666	0.8710	0.8154	279	0.7429	0.5843	0.6541	178
	TF-IDF	(1)-Gram	0.7462	0.13	0.7840	0.8065	0.7951	279	0.6824	0.6517	0.6667	178
		(1+2)-Gram	0.7615	0.32	0.7891	0.8315	0.8098	279	0.7117	0.6517	0.6804	178
		(1-2-3)-Gram	0.7702	0.55	0.7900	0.8495	0.8187	279	0.7325	0.6461	0.6866	178

Table 4.15: Summary of SVM classifier results on SS_DATASET

SVM SUMMARY	Max Value	Min Value	Average
Accuracy	0.7768	0.7309	0.7568
Time	0.55	0.13	0.33
Contribution of 2-Gram	2.8446%	1.5317%	2.1882%
Contribution of 3-Gram	0.8753%	-0.8753%	0.0547%
Contribution of TF-IDF	1.9694%	-0.2188%	1.0394%
Contribution of Negation	0.8753%	-0.4376%	0.4741%
Contribution of POS-Tagging	0.2188%	-1.0941%	-0.4012%
Contribution of NEG+POS	1.0941%	-0.8753%	0.1094%

Evaluation results of Support Vector Machine classifier on different combinations of SS_DATASET are shown in Table 4.14. The maximum accuracy obtained is 77.68 percent but it takes 0.39 seconds to finish, and the minimum time obtained is 0.13 seconds but its accuracy is 74.62 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.15.

Table 4.16: Evaluation results of RF classifier on SS_DATASET

RF	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7287	0.65	0.7541	0.8244	0.7877	279	0.6776	0.5787	0.6242	178
		(1+2)-Gram	0.7418	1.24	0.7347	0.9032	0.8103	279	0.7632	0.4888	0.5959	178
		(1-2-3)-Gram	0.7090	1.58	0.7147	0.8710	0.7851	279	0.6923	0.4551	0.5492	178
	TF-IDF	(1)-Gram	0.7199	0.61	0.7649	0.7814	0.7730	279	0.6453	0.6236	0.6343	178
		(1+2)-Gram	0.7243	1.14	0.7476	0.8280	0.7857	279	0.6757	0.5618	0.6135	178
		(1-2-3)-Gram	0.7221	1.63	0.7500	0.8172	0.7822	279	0.6667	0.5730	0.6163	178
BASIC + NEG	BINARY	(1)-Gram	0.7309	0.71	0.7566	0.8244	0.7890	279	0.6797	0.5843	0.6284	178
		(1+2)-Gram	0.7155	1.27	0.7198	0.8746	0.7896	279	0.7034	0.4663	0.5608	178
		(1-2-3)-Gram	0.7133	1.59	0.7079	0.9032	0.7937	279	0.7327	0.4157	0.5305	178
	TF-IDF	(1)-Gram	0.7090	0.66	0.7450	0.7957	0.7695	279	0.6415	0.5730	0.6053	178
		(1+2)-Gram	0.7243	1.20	0.7476	0.8280	0.7857	279	0.6757	0.5618	0.6135	178
		(1-2-3)-Gram	0.6958	1.61	0.7229	0.8136	0.7656	279	0.6364	0.5112	0.5670	178
BASIC + POS	BINARY	(1)-Gram	0.7309	0.78	0.7349	0.8746	0.7987	279	0.7200	0.5056	0.5941	178
		(1+2)-Gram	0.7243	1.25	0.7257	0.8817	0.7961	279	0.7203	0.4775	0.5743	178
		(1-2-3)-Gram	0.7090	1.63	0.7186	0.8602	0.7830	279	0.6829	0.4719	0.5581	178
	TF-IDF	(1)-Gram	0.7133	0.63	0.7357	0.8280	0.7791	279	0.6643	0.5337	0.5919	178
		(1+2)-Gram	0.7221	1.25	0.7568	0.8029	0.7791	279	0.6584	0.5955	0.6254	178
		(1-2-3)-Gram	0.7046	1.60	0.7130	0.8638	0.7812	279	0.6807	0.4551	0.5455	178
BASIC + NEG + POS	BINARY	(1)-Gram	0.7309	0.71	0.7378	0.8674	0.7974	279	0.7132	0.5169	0.5993	178
		(1+2)-Gram	0.7418	1.19	0.7389	0.8925	0.8084	279	0.7500	0.5056	0.6040	178
		(1-2-3)-Gram	0.7177	1.57	0.7131	0.8996	0.7956	279	0.7333	0.4326	0.5442	178
	TF-IDF	(1)-Gram	0.7440	0.71	0.7630	0.8423	0.8007	279	0.7047	0.5899	0.6422	178
		(1+2)-Gram	0.7046	1.10	0.7236	0.8351	0.7754	279	0.6593	0.5000	0.5687	178
		(1-2-3)-Gram	0.7199	1.49	0.7323	0.8530	0.7881	279	0.6894	0.5112	0.5871	178

Table 4.17: Summary of RF classifier results on SS_DATASET

RF SUMMARY	Max Value	Min Value	Average
Accuracy	0.7440	0.6958	0.7207
Time	1.63	0.61	1.16
Contribution of 2-Gram	1.5317%	-3.9387%	-0.1094%
Contribution of 3-Gram	1.5317%	-3.2823%	-1.3403%
Contribution of TF-IDF	1.3129%	-3.7199%	-0.7476%
Contribution of Negation	0.4376%	-2.6258%	-0.9482%
Contribution of POS-Tagging	0.2188%	-1.7505%	-0.6929%
Contribution of NEG + POS	2.4070%	-1.9694%	0.2188%

Evaluation results of Random Forest classifier on different combinations of SS_DATASET are shown in Table 4.16. The maximum accuracy obtained is 74.40 percent but it takes 0.71 seconds to finish, and the minimum time obtained is 0.61 seconds but its accuracy is 71.99 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.17.

Table 4.18: Evaluation results of LR classifier on SS_DATASET

LR	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7330	0.68	0.7476	0.8495	0.7953	279	0.7000	0.5506	0.6164	178
		(1+2)-Gram	0.7068	1.23	0.7302	0.8244	0.7744	279	0.6549	0.5225	0.5813	178
		(1-2-3)-Gram	0.7330	1.47	0.7343	0.8817	0.8013	279	0.7295	0.5000	0.5933	178
	TF-IDF	(1)-Gram	0.7243	0.65	0.7593	0.8029	0.7805	279	0.6605	0.6011	0.6294	178
		(1+2)-Gram	0.7330	1.26	0.7661	0.8100	0.7875	279	0.6728	0.6124	0.6412	178
		(1-2-3)-Gram	0.6958	1.77	0.7215	0.8172	0.7664	279	0.6383	0.5056	0.5643	178
BASIC + NEG	BINARY	(1)-Gram	0.7243	0.57	0.7429	0.8387	0.7879	279	0.6831	0.5449	0.6063	178
		(1+2)-Gram	0.7112	1.22	0.7130	0.8817	0.7885	279	0.7054	0.4438	0.5448	178
		(1-2-3)-Gram	0.7155	1.29	0.7147	0.8889	0.7923	279	0.7182	0.4438	0.5486	178
	TF-IDF	(1)-Gram	0.7133	0.54	0.7552	0.7849	0.7698	279	0.6407	0.6011	0.6203	178
		(1+2)-Gram	0.7155	1.17	0.7350	0.8351	0.7819	279	0.6714	0.5281	0.5912	178
		(1-2-3)-Gram	0.7046	1.66	0.7278	0.8244	0.7731	279	0.6525	0.5169	0.5768	178
BASIC + POS	BINARY	(1)-Gram	0.7265	0.66	0.7362	0.8602	0.7934	279	0.7023	0.5169	0.5955	178
		(1+2)-Gram	0.7265	1.24	0.7188	0.9068	0.8019	279	0.7524	0.4438	0.5583	178
		(1-2-3)-Gram	0.7024	1.41	0.7014	0.8925	0.7855	279	0.7059	0.4045	0.5143	178
	TF-IDF	(1)-Gram	0.7440	0.69	0.7485	0.8746	0.8066	279	0.7328	0.5393	0.6214	178
		(1+2)-Gram	0.7046	1.31	0.7236	0.8351	0.7754	279	0.6593	0.5000	0.5687	178
		(1-2-3)-Gram	0.7221	1.66	0.7331	0.8566	0.7901	279	0.6947	0.5112	0.5890	178
BASIC + NEG + POS	BINARY	(1)-Gram	0.7287	0.59	0.7313	0.8781	0.7980	279	0.7213	0.4944	0.5867	178
		(1+2)-Gram	0.7330	1.05	0.7316	0.8889	0.8026	279	0.7373	0.4888	0.5878	178
		(1-2-3)-Gram	0.7133	1.62	0.7033	0.9176	0.7963	279	0.7527	0.3933	0.5166	178
	TF-IDF	(1)-Gram	0.7199	0.56	0.7382	0.8387	0.7852	279	0.6786	0.5337	0.5975	178
		(1+2)-Gram	0.7352	0.90	0.7532	0.8423	0.7953	279	0.6966	0.5674	0.6254	178
		(1-2-3)-Gram	0.7046	1.55	0.7278	0.8244	0.7731	279	0.6525	0.5169	0.5768	178

Table 4.19: Summary of LR classifier results on SS_DATASET

LR SUMMARY	Max Value	Min Value	Average
Accuracy	0.7440	0.6958	0.7196
Time	1.77	0.54	1.11
Contribution of 2-Gram	1.5317%	-3.9387%	-0.6018%
Contribution of 3-Gram	2.6258%	-3.7199%	-0.9300%
Contribution of TF-IDF	2.6258%	-3.7199%	-0.3100%
Contribution of Negation	0.8753%	-1.7505%	-0.6929%
Contribution of POS-Tagging	2.6258%	-3.0635%	0.0000%
Contribution of NEG + POS	2.6258%	-1.9694%	0.1459%

Evaluation results of Linear Regression classifier on different combinations of SS_DATASET are shown in Table 4.18. The maximum accuracy obtained is 74.40 percent but it takes 0.69 seconds to finish, and the minimum time obtained is 0.54 seconds but its accuracy is 71.33 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.19.

Table 4.20: Evaluation results of VADER SA tool on SS_DATASET

VDR	DATA REPRESENTATION	ACCURACY	TRAINING TIME	POSITIVE Precision	POSITIVE Recall	POSITIVE F1-Score	POSITIVE Support	NEGATIVE Precision	NEGATIVE Recall	NEGATIVE F1-Score	NEGATIVE Support
NO_PP	(1)-Gram	0.7889	0.00	0.7857	0.9009	0.8394	232	0.7965	0.6122	0.6923	147
BASIC	(1)-Gram	0.7500	0.00	0.7645	0.8645	0.8114	214	0.7157	0.5615	0.6293	130

Evaluation results of VADER sentiment analysis tool on different combinations of SS_DATASET are shown in Table 4.20. The best accuracy obtained is 78.89 percent and time value is zero because there is no training step.

Table 4.21: Evaluation results of DBL classifier on SS_DATASET

DBL	DATA REPRESENTATION	ACCURACY	TRAINING TIME (*)	POSITIVE Precision	POSITIVE Recall	POSITIVE F1-Score	POSITIVE Support	NEGATIVE Precision	NEGATIVE Recall	NEGATIVE F1-Score	NEGATIVE Support
BASIC	(1)-Gram	0.7188	0.69	0.7467	0.8235	0.7832	272	0.6596	0.5503	0.6000	169
BASIC NEG	(1)-Gram	0.7238	0.80	0.7500	0.8212	0.7840	274	0.6711	0.5714	0.6173	175
BASIC POS	(1)-Gram	0.6948	0.73	0.7313	0.7963	0.7624	270	0.6207	0.5325	0.5732	169
BASIC POSNEG	(1)-Gram	0.7045	0.78	0.7356	0.8067	0.7695	269	0.6414	0.5439	0.5886	171

Table 4.22: Summary of DBL classifier results on SS_DATASET

DBL SUMMARY	Max Value	Min Value	Average
Accuracy	0.7238	0.6948	0.7105
Time	0.80	0.69	0.75
Contribution of Negation	0.5010%	0.5010%	0.5010%
Contribution of POS-Tagging	-2.4060%	-2.4060%	-2.4060%
Contribution of NEG + POS	-1.4275%	-1.4275%	-1.4275%

Evaluation results of my own Domain Based Lexicon implementation on different combinations of SS_DATASET are shown in Table 4.21. The maximum accuracy obtained is 72.38 percent but it takes 0.80 seconds to finish, and the minimum time obtained is 0.69 seconds but its accuracy is 71.88 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.22.

Table 4.23: Evaluation results of MNB classifier on AIRLINE_DATASET

MNB	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.9077	0.43	0.8939	0.6608	0.7599	510	0.9104	0.9777	0.9428	1797
		(1+2)-Gram	0.8825	1.42	0.9799	0.4784	0.6430	510	0.8707	0.9972	0.9297	1797
		(1-2-3)-Gram	0.8669	2.11	0.9903	0.4020	0.5718	510	0.8548	0.9989	0.9212	1797
	TF-IDF	(1)-Gram	0.8474	0.52	0.9877	0.3137	0.4762	510	0.8368	0.9989	0.9107	1797
		(1+2)-Gram	0.8132	1.37	1.0000	0.1549	0.2683	510	0.8066	1.0000	0.8929	1797
		(1-2-3)-Gram	0.8101	2.30	1.0000	0.1412	0.2474	510	0.8040	1.0000	0.8914	1797
BASIC + NEG	BINARY	(1)-Gram	0.9129	0.53	0.8911	0.6902	0.7779	510	0.9174	0.9761	0.9458	1797
		(1+2)-Gram	0.8877	1.52	0.9736	0.5059	0.6658	510	0.8766	0.9961	0.9325	1797
		(1-2-3)-Gram	0.8743	2.62	0.9867	0.4373	0.6060	510	0.8621	0.9983	0.9252	1797
	TF-IDF	(1)-Gram	0.8500	0.44	0.9881	0.3255	0.4897	510	0.8392	0.9989	0.9121	1797
		(1+2)-Gram	0.8162	1.32	1.0000	0.1686	0.2886	510	0.8091	1.0000	0.8945	1797
		(1-2-3)-Gram	0.8110	2.48	1.0000	0.1451	0.2534	510	0.8047	1.0000	0.8918	1797
BASIC + POS	BINARY	(1)-Gram	0.9020	0.44	0.8778	0.6345	0.7366	498	0.9065	0.9757	0.9398	1809
		(1+2)-Gram	0.8804	1.41	0.9625	0.4639	0.6260	498	0.8708	0.9950	0.9288	1809
		(1-2-3)-Gram	0.8704	2.04	0.9807	0.4076	0.5759	498	0.8595	0.9978	0.9235	1809
	TF-IDF	(1)-Gram	0.8461	0.54	0.9931	0.2892	0.4479	498	0.8363	0.9994	0.9106	1809
		(1+2)-Gram	0.8201	1.40	0.9882	0.1687	0.2882	498	0.8137	0.9994	0.8970	1809
		(1-2-3)-Gram	0.8162	2.33	0.9868	0.1506	0.2613	498	0.8104	0.9994	0.8950	1809
BASIC + NEG + POS	BINARY	(1)-Gram	0.9068	0.50	0.8675	0.6707	0.7565	498	0.9147	0.9718	0.9424	1809
		(1+2)-Gram	0.8877	1.46	0.9650	0.4980	0.6570	498	0.8780	0.9950	0.9329	1809
		(1-2-3)-Gram	0.8734	2.22	0.9682	0.4277	0.5933	498	0.8634	0.9961	0.9251	1809
	TF-IDF	(1)-Gram	0.8483	0.52	0.9933	0.2992	0.4599	498	0.8382	0.9994	0.9117	1809
		(1+2)-Gram	0.8214	1.34	0.9886	0.1747	0.2969	498	0.8148	0.9994	0.8977	1809
		(1-2-3)-Gram	0.8184	2.00	0.9877	0.1606	0.2763	498	0.8122	0.9994	0.8962	1809

Table 4.24: Summary of MNB classifier results on AIRLINE_DATASET

MNB SUMMARY	Max Value	Min Value	Average
Accuracy	0.9129	0.8101	0.8571
Time	2.62	0.43	1.39
Contribution of 2-Gram	-1.9072%	-3.4244%	-2.6495%
Contribution of 3-Gram	-0.3034%	-1.5605%	-0.8561%
Contribution of TF-IDF	-5.4183%	-7.1521%	-6.1191%
Contribution of Negation	0.7369%	0.0867%	0.4046%
Contribution of POS-Tagging	0.6935%	-0.5635%	0.1228%
Contribution of NEG + POS	0.8236%	-0.0867%	0.4696%

Evaluation results of Multinomial Naive Bayes classifier on different combinations of AIRLINE_DATASET are shown in Table 4.23. The maximum accuracy obtained is 91.29 percent but it takes 0.53 seconds to finish, and the minimum time obtained is 0.43 seconds but its accuracy is 90.77 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.24.

Table 4.25: Evaluation results of SVM classifier on AIRLINE_DATASET

SVM	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.9038	0.74	0.7975	0.7569	0.7767	510	0.9320	0.9455	0.9387	1797
		(1+2)-Gram	0.9202	1.43	0.8622	0.7608	0.8083	510	0.9343	0.9655	0.9496	1797
		(1-2-3)-Gram	0.9181	3.19	0.8559	0.7569	0.8033	510	0.9332	0.9638	0.9483	1797
	TF-IDF	(1)-Gram	0.9168	0.62	0.8630	0.7412	0.7975	510	0.9294	0.9666	0.9476	1797
		(1+2)-Gram	0.9224	1.50	0.8950	0.7353	0.8073	510	0.9285	0.9755	0.9514	1797
		(1-2-3)-Gram	0.9211	2.38	0.8905	0.7333	0.8043	510	0.9279	0.9744	0.9506	1797
BASIC + NEG	BINARY	(1)-Gram	0.9090	0.75	0.8074	0.7725	0.7896	510	0.9362	0.9477	0.9419	1797
		(1+2)-Gram	0.9181	2.11	0.8482	0.7667	0.8054	510	0.9355	0.9610	0.9481	1797
		(1-2-3)-Gram	0.9159	3.33	0.8450	0.7588	0.7996	510	0.9335	0.9605	0.9468	1797
	TF-IDF	(1)-Gram	0.9211	0.62	0.8710	0.7549	0.8088	510	0.9330	0.9683	0.9503	1797
		(1+2)-Gram	0.9272	1.45	0.9014	0.7529	0.8205	510	0.9330	0.9766	0.9543	1797
		(1-2-3)-Gram	0.9254	2.55	0.8912	0.7549	0.8174	510	0.9333	0.9738	0.9532	1797
BASIC + POS	BINARY	(1)-Gram	0.9103	0.65	0.7975	0.7831	0.7903	498	0.9406	0.9453	0.9429	1809
		(1+2)-Gram	0.9142	1.97	0.8219	0.7691	0.7946	498	0.9375	0.9541	0.9458	1809
		(1-2-3)-Gram	0.9133	3.13	0.8239	0.7610	0.7912	498	0.9356	0.9552	0.9453	1809
	TF-IDF	(1)-Gram	0.9194	0.61	0.8662	0.7410	0.7987	498	0.9314	0.9685	0.9496	1809
		(1+2)-Gram	0.9202	1.06	0.8720	0.7390	0.8000	498	0.9310	0.9701	0.9502	1809
		(1-2-3)-Gram	0.9202	2.35	0.8703	0.7410	0.8004	498	0.9315	0.9696	0.9502	1809
BASIC + NEG + POS	BINARY	(1)-Gram	0.9064	0.66	0.7809	0.7871	0.7840	498	0.9413	0.9392	0.9402	1809
		(1+2)-Gram	0.9146	1.61	0.8209	0.7731	0.7963	498	0.9385	0.9536	0.9460	1809
		(1-2-3)-Gram	0.9098	2.81	0.8112	0.7590	0.7842	498	0.9348	0.9514	0.9430	1809
	TF-IDF	(1)-Gram	0.9176	0.32	0.8581	0.7410	0.7953	498	0.9313	0.9663	0.9485	1809
		(1+2)-Gram	0.9202	1.60	0.8738	0.7369	0.7996	498	0.9306	0.9707	0.9502	1809
		(1-2-3)-Gram	0.9202	2.69	0.8634	0.7490	0.8022	498	0.9333	0.9674	0.9501	1809

Table 4.26: Summary of SVM classifier results on AIRLINE_DATASET

SVM SUMMARY	Max Value	Min Value	Average
Accuracy	0.9272	0.9038	0.9169
Time	3.33	0.32	1.67
Contribution of 2-Gram	1.6472%	0.0867%	0.6610%
Contribution of 3-Gram	0.0000%	-0.4768%	-0.1625%
Contribution of TF-IDF	1.3004%	0.2167%	0.8200%
Contribution of Negation	0.5202%	-0.2167%	0.2384%
Contribution of POS-Tagging	0.6502%	-0.6068%	-0.0795%
Contribution of NEG + POS	0.2601%	-0.8236%	-0.2240%

Evaluation results of Support Vector Machine classifier on different combinations of AIRLINE_DATASET are shown in Table 4.25. The maximum accuracy obtained is 92.72 percent but it takes 1.45 seconds to finish, and the minimum time obtained is 0.32 seconds but its accuracy is 91.76 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.26.

Table 4.27: Evaluation results of RF classifier on AIRLINE_DATASET

RF	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.9033	2.80	0.8747	0.6569	0.7503	510	0.9090	0.9733	0.9401	1797
		(1+2)-Gram	0.8847	10.44	0.8765	0.5569	0.6811	510	0.8860	0.9777	0.9296	1797
		(1-2-3)-Gram	0.8873	16.05	0.8834	0.5647	0.6890	510	0.8879	0.9789	0.9312	1797
	TF-IDF	(1)-Gram	0.8929	2.73	0.8725	0.6039	0.7138	510	0.8966	0.9750	0.9342	1797
		(1+2)-Gram	0.8999	7.44	0.8886	0.6255	0.7342	510	0.9020	0.9777	0.9383	1797
		(1-2-3)-Gram	0.8877	13.79	0.8294	0.6196	0.7093	510	0.8993	0.9638	0.9304	1797
BASIC+NEG	BINARY	(1)-Gram	0.8947	3.19	0.8579	0.6275	0.7248	510	0.9018	0.9705	0.9349	1797
		(1+2)-Gram	0.8877	9.66	0.9035	0.5510	0.6845	510	0.8853	0.9833	0.9317	1797
		(1-2-3)-Gram	0.8942	15.86	0.8889	0.5961	0.7136	510	0.8952	0.9789	0.9351	1797
	TF-IDF	(1)-Gram	0.8938	2.60	0.8863	0.5961	0.7128	510	0.8951	0.9783	0.9349	1797
		(1+2)-Gram	0.8921	9.34	0.8615	0.6098	0.7141	510	0.8977	0.9722	0.9335	1797
		(1-2-3)-Gram	0.8856	15.49	0.7929	0.6529	0.7161	510	0.9062	0.9516	0.9283	1797
BASIC+POS	BINARY	(1)-Gram	0.8938	3.65	0.8355	0.6325	0.7200	498	0.9052	0.9657	0.9345	1809
		(1+2)-Gram	0.8929	10.22	0.8420	0.6205	0.7145	498	0.9026	0.9679	0.9341	1809
		(1-2-3)-Gram	0.8890	15.68	0.8477	0.5924	0.6974	498	0.8964	0.9707	0.9321	1809
	TF-IDF	(1)-Gram	0.8916	3.14	0.8758	0.5803	0.6981	498	0.8943	0.9773	0.9340	1809
		(1+2)-Gram	0.8877	9.14	0.7965	0.6446	0.7125	498	0.9070	0.9547	0.9302	1809
		(1-2-3)-Gram	0.8925	13.41	0.8019	0.6667	0.7281	498	0.9123	0.9547	0.9330	1809
BASIC+NEG+POS	BINARY	(1)-Gram	0.8977	3.97	0.8075	0.6908	0.7446	498	0.9181	0.9547	0.9360	1809
		(1+2)-Gram	0.8942	10.96	0.8396	0.6305	0.7202	498	0.9048	0.9668	0.9348	1809
		(1-2-3)-Gram	0.8851	15.29	0.8123	0.6084	0.6958	498	0.8992	0.9613	0.9292	1809
	TF-IDF	(1)-Gram	0.8895	3.21	0.8403	0.6024	0.7018	498	0.8985	0.9685	0.9322	1809
		(1+2)-Gram	0.8960	10.05	0.8162	0.6687	0.7351	498	0.9131	0.9585	0.9353	1809
		(1-2-3)-Gram	0.8847	14.47	0.7437	0.7108	0.7269	498	0.9214	0.9326	0.9269	1809

Table 4.28: Summary of RF classifier results on AIRLINE_DATASET

RF SUMMARY	Max Value	Min Value	Average
Accuracy	0.9033	0.8847	0.8916
Time	16.05	2.60	9.27
Contribution of 2-Gram	0.6935%	-1.8639%	-0.2763%
Contribution of 3-Gram	0.6502%	-1.2137%	-0.3630%
Contribution of TF-IDF	1.5171%	-1.0403%	-0.0903%
Contribution of Negation	0.6935%	-0.8669%	-0.1300%
Contribution of POS-Tagging	0.8236%	-1.2137%	-0.1373%
Contribution of NEG + POS	0.9536%	-0.5635%	-0.1445%

Evaluation results of Random Forest classifier on different combinations of AIRLINE_DATASET are shown in Table 4.27. The maximum accuracy obtained is 90.33 percent but it takes 2.80 seconds to finish, and the minimum time obtained is 2.60 seconds but its accuracy is 89.38 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.28.

Table 4.29: Evaluation results of LR classifier on AIRLINE_DATASET

LR	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.8916	3.25	0.8736	0.5961	0.7086	510	0.8948	0.9755	0.9334	1797
		(1+2)-Gram	0.8955	9.62	0.9325	0.5686	0.7065	510	0.8898	0.9883	0.9365	1797
		(1-2-3)-Gram	0.8947	15.48	0.9159	0.5765	0.7076	510	0.8912	0.9850	0.9358	1797
	TF-IDF	(1)-Gram	0.8847	2.57	0.8813	0.5529	0.6795	510	0.8853	0.9789	0.9297	1797
		(1+2)-Gram	0.8955	8.11	0.8876	0.6039	0.7188	510	0.8969	0.9783	0.9359	1797
		(1-2-3)-Gram	0.8908	15.70	0.8644	0.6000	0.7083	510	0.8955	0.9733	0.9328	1797
BASIC + NEG	BINARY	(1)-Gram	0.9029	3.20	0.8763	0.6529	0.7483	510	0.9081	0.9738	0.9398	1797
		(1+2)-Gram	0.8968	9.64	0.8908	0.6078	0.7226	510	0.8979	0.9789	0.9366	1797
		(1-2-3)-Gram	0.8934	16.44	0.9074	0.5765	0.7050	510	0.8911	0.9833	0.9349	1797
	TF-IDF	(1)-Gram	0.8968	2.65	0.8842	0.6137	0.7245	510	0.8991	0.9772	0.9365	1797
		(1+2)-Gram	0.8882	9.17	0.8351	0.6157	0.7088	510	0.8985	0.9655	0.9308	1797
		(1-2-3)-Gram	0.8886	14.69	0.7935	0.6706	0.7269	510	0.9104	0.9505	0.9300	1797
BASIC + POS	BINARY	(1)-Gram	0.8968	3.89	0.8110	0.6807	0.7402	498	0.9158	0.9563	0.9356	1809
		(1+2)-Gram	0.9020	10.19	0.8617	0.6506	0.7414	498	0.9099	0.9713	0.9396	1809
		(1-2-3)-Gram	0.8886	14.28	0.8453	0.5924	0.6966	498	0.8963	0.9701	0.9318	1809
	TF-IDF	(1)-Gram	0.8925	3.10	0.8571	0.6024	0.7075	498	0.8988	0.9724	0.9341	1809
		(1+2)-Gram	0.8942	9.76	0.8290	0.6426	0.7240	498	0.9073	0.9635	0.9346	1809
		(1-2-3)-Gram	0.8851	12.90	0.8074	0.6145	0.6978	498	0.9004	0.9596	0.9291	1809
BASIC + NEG + POS	BINARY	(1)-Gram	0.8916	4.32	0.8039	0.6586	0.7241	498	0.9105	0.9558	0.9326	1809
		(1+2)-Gram	0.8890	10.66	0.8361	0.6044	0.7016	498	0.8988	0.9674	0.9318	1809
		(1-2-3)-Gram	0.8951	16.14	0.8478	0.6265	0.7206	498	0.9041	0.9690	0.9354	1809
	TF-IDF	(1)-Gram	0.8942	3.54	0.8528	0.6165	0.7156	498	0.9019	0.9707	0.9350	1809
		(1+2)-Gram	0.8916	9.81	0.8039	0.6586	0.7241	498	0.9105	0.9558	0.9326	1809
		(1-2-3)-Gram	0.8791	13.68	0.7731	0.6225	0.6897	498	0.9014	0.9497	0.9249	1809

Table 4.30: Summary of LR classifier results on AIRLINE_DATASET

LR SUMMARY	Max Value	Min Value	Average
Accuracy	0.9029	0.8791	0.8925
Time	16.44	2.57	9.28
Contribution of 2-Gram	1.0837%	-0.8669%	0.0217%
Contribution of 3-Gram	0.6068%	-1.3437%	-0.4714%
Contribution of TF-IDF	0.2601%	-1.6038%	-0.4732%
Contribution of Negation	1.2137%	-0.7369%	0.2312%
Contribution of POS-Tagging	0.7802%	-0.6068%	0.1084%
Contribution of NEG + POS	0.9536%	-1.1704%	-0.2023%

Evaluation results of Linear Regression classifier on different combinations of AIRLINE_DATASET are shown in Table 4.29. The maximum accuracy obtained is 90.29 percent but it takes 3.20 seconds to finish, and the minimum time obtained is 2.57 seconds but its accuracy is 88.47 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.30.

Table 4.31: Evaluation results of VADER SA tool on AIRLINE_DATASET

VDR	DATA REPRESENTATION	ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
				Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
NO_PP	(1)-Gram	0.7044	0.00	0.4345	0.9441	0.5952	429	0.9742	0.6328	0.7672	1435
BASIC	(1)-Gram	0.6973	0.00	0.4452	0.9393	0.6041	428	0.9690	0.6184	0.7550	1313

Evaluation results of VADER sentiment analysis tool on different combinations of AIRLINE_DATASET are shown in Table 4.31. The best accuracy obtained is 70.44 percent and time value is zero because there is no training step.

Table 4.32: Evaluation results of DBL classifier on AIRLINE_DATASET

DBL	DATA REPRESENTATION	ACCURACY	TRAINING TIME (*)	POSITIVE				NEGATIVE			
				Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	(1)-Gram	0.8119	4.07	0.9620	0.1502	0.2598	506	0.8066	0.9983	0.8923	1796
BASIC NEG	(1)-Gram	0.8149	4.02	0.9545	0.1660	0.2828	506	0.8093	0.9978	0.8937	1795
BASIC POS	(1)-Gram	0.8275	4.46	0.9712	0.2040	0.3372	495	0.8207	0.9983	0.9009	1807
BASIC POSNEG	(1)-Gram	0.8292	4.09	0.9720	0.2105	0.3461	494	0.8222	0.9983	0.9018	1807

Table 4.33: Summary of DBL classifier results on AIRLINE_DATASET

DBL SUMMARY	Max Value	Min Value	Average
Accuracy	0.8292	0.8119	0.8209
Time	4.46	4.02	4.16
Contribution of Negation	0.2960%	0.2960%	0.2960%
Contribution of POS-Tagging	1.5639%	1.5639%	1.5639%
Contribution of NEG + POS	1.7302%	1.7302%	1.7302%

Evaluation results of my own Domain Based Lexicon implementation on different combinations of AIRLINE_DATASET are shown in Table 4.32. The maximum accuracy obtained is 82.92 percent but it takes 4.09 seconds to finish, and the minimum time obtained is 4.02 seconds but its accuracy is 81.49 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.33.

Table 4.34: Evaluation results of MNB classifier on SD_CAR_DATASET

MNB	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7852	0.16	0.7780	0.9593	0.8592	369	0.8235	0.4094	0.5469	171
		(1+2)-Gram	0.7704	0.25	0.7495	0.9973	0.8558	369	0.9796	0.2807	0.4364	171
		(1-2-3)-Gram	0.7556	0.37	0.7365	1.0000	0.8483	369	1.0000	0.2281	0.3714	171
	TF-IDF	(1)-Gram	0.7185	0.11	0.7083	1.0000	0.8292	369	1.0000	0.1111	0.2000	171
		(1+2)-Gram	0.7037	0.23	0.6975	1.0000	0.8218	369	1.0000	0.0643	0.1209	171
		(1-2-3)-Gram	0.7056	0.31	0.6989	1.0000	0.8227	369	1.0000	0.0702	0.1311	171
BASIC + NEG	BINARY	(1)-Gram	0.7907	0.08	0.7832	0.9593	0.8624	369	0.8295	0.4269	0.5637	171
		(1+2)-Gram	0.7741	0.20	0.7526	0.9973	0.8578	369	0.9804	0.2924	0.4505	171
		(1-2-3)-Gram	0.7704	0.42	0.7485	1.0000	0.8561	369	1.0000	0.2749	0.4312	171
	TF-IDF	(1)-Gram	0.7185	0.12	0.7083	1.0000	0.8292	369	1.0000	0.1111	0.2000	171
		(1+2)-Gram	0.7019	0.25	0.6962	1.0000	0.8209	369	1.0000	0.0585	0.1105	171
		(1-2-3)-Gram	0.7093	0.41	0.7015	1.0000	0.8246	369	1.0000	0.0819	0.1514	171
BASIC + POS	BINARY	(1)-Gram	0.7648	0.12	0.7553	0.9702	0.8493	369	0.8333	0.3216	0.4641	171
		(1+2)-Gram	0.7537	0.24	0.7379	0.9919	0.8462	369	0.9318	0.2398	0.3814	171
		(1-2-3)-Gram	0.7481	0.41	0.7335	0.9919	0.8433	369	0.9268	0.2222	0.3585	171
	TF-IDF	(1)-Gram	0.7093	0.12	0.7015	1.0000	0.8246	369	1.0000	0.0819	0.1514	171
		(1+2)-Gram	0.6944	0.26	0.6910	1.0000	0.8173	369	1.0000	0.0351	0.0678	171
		(1-2-3)-Gram	0.6963	0.42	0.6923	1.0000	0.8182	369	1.0000	0.0409	0.0787	171
BASIC + NEG + POS	BINARY	(1)-Gram	0.7759	0.11	0.7616	0.9783	0.8565	369	0.8788	0.3392	0.4895	171
		(1+2)-Gram	0.7648	0.25	0.7469	0.9919	0.8522	369	0.9400	0.2749	0.4253	171
		(1-2-3)-Gram	0.7574	0.42	0.7419	0.9892	0.8479	369	0.9167	0.2573	0.4018	171
	TF-IDF	(1)-Gram	0.7111	0.11	0.7029	1.0000	0.8255	369	1.0000	0.0877	0.1613	171
		(1+2)-Gram	0.6944	0.22	0.6910	1.0000	0.8173	369	1.0000	0.0351	0.0678	171
		(1-2-3)-Gram	0.6963	0.42	0.6923	1.0000	0.8182	369	1.0000	0.0409	0.0787	171

Table 4.35: Summary of MNB classifier results on SD_CAR_DATASET

MNB SUMMARY	Max Value	Min Value	Average
Accuracy	0.7907	0.6944	0.7363
Time	0.42	0.08	0.25
Contribution of 2-Gram	-1.1111%	-1.6667%	-1.4583%
Contribution of 3-Gram	0.7407%	-1.4815%	-0.2315%
Contribution of TF-IDF	-5.0000%	-7.2222%	-6.2654%
Contribution of Negation	1.4815%	-0.1852%	0.4321%
Contribution of POS-Tagging	-0.7407%	-2.0370%	-1.2037%
Contribution of NEG + POS	0.1852%	-0.9259%	-0.6481%

Evaluation results of Multinomial Naive Bayes classifier on different combinations of SD_CAR_DATASET are shown in Table 4.34. The maximum accuracy obtained is 79.07 percent but it takes 0.08 seconds to finish, and the minimum time obtained is also 0.08 seconds and its accuracy is also 79.07 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.35.

Table 4.36: Evaluation results of SVM classifier on SD_CAR_DATASET

SVM	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7722	0.15	0.8045	0.8808	0.8409	369	0.6765	0.5380	0.5993	171
		(1+2)-Gram	0.7852	0.40	0.7991	0.9160	0.8535	369	0.7350	0.5029	0.5972	171
		(1-2-3)-Gram	0.7796	0.64	0.7880	0.9268	0.8518	369	0.7453	0.4620	0.5704	171
	TF-IDF	(1)-Gram	0.7759	0.13	0.7981	0.8997	0.8459	369	0.7016	0.5088	0.5898	171
		(1+2)-Gram	0.7796	0.28	0.7894	0.9241	0.8514	369	0.7407	0.4678	0.5735	171
		(1-2-3)-Gram	0.7852	0.37	0.7817	0.9512	0.8582	369	0.8022	0.4269	0.5573	171
BASIC+NEG	BINARY	(1)-Gram	0.7759	0.19	0.8010	0.8943	0.8451	369	0.6953	0.5205	0.5953	171
		(1+2)-Gram	0.7926	0.39	0.8024	0.9241	0.8589	369	0.7565	0.5088	0.6084	171
		(1-2-3)-Gram	0.7796	0.56	0.7854	0.9322	0.8525	369	0.7549	0.4503	0.5641	171
	TF-IDF	(1)-Gram	0.7963	0.12	0.8091	0.9187	0.8604	369	0.7521	0.5322	0.6233	171
		(1+2)-Gram	0.7963	0.27	0.7950	0.9458	0.8639	369	0.8020	0.4737	0.5956	171
		(1-2-3)-Gram	0.7833	0.36	0.7813	0.9485	0.8568	369	0.7935	0.4269	0.5551	171
BASIC+POS	BINARY	(1)-Gram	0.7778	0.16	0.7943	0.9106	0.8485	369	0.7179	0.4912	0.5833	171
		(1+2)-Gram	0.7759	0.33	0.7857	0.9241	0.8493	369	0.7358	0.4561	0.5632	171
		(1-2-3)-Gram	0.7685	0.43	0.7711	0.9404	0.8474	369	0.7556	0.3977	0.5211	171
	TF-IDF	(1)-Gram	0.7889	0.09	0.7918	0.9377	0.8586	369	0.7767	0.4678	0.5839	171
		(1+2)-Gram	0.7796	0.23	0.7778	0.9485	0.8547	369	0.7889	0.4152	0.5441	171
		(1-2-3)-Gram	0.7741	0.44	0.7702	0.9539	0.8523	369	0.7952	0.3860	0.5197	171
BASIC+NEG+POS	BINARY	(1)-Gram	0.7926	0.19	0.8096	0.9106	0.8571	369	0.7360	0.5380	0.6216	171
		(1+2)-Gram	0.7741	0.41	0.7813	0.9295	0.8490	369	0.7426	0.4386	0.5515	171
		(1-2-3)-Gram	0.7630	0.58	0.7648	0.9431	0.8447	369	0.7529	0.3743	0.5000	171
	TF-IDF	(1)-Gram	0.7907	0.16	0.7977	0.9295	0.8586	369	0.7636	0.4912	0.5979	171
		(1+2)-Gram	0.7685	0.31	0.7687	0.9458	0.8481	369	0.7674	0.3860	0.5136	171
		(1-2-3)-Gram	0.7685	0.45	0.7641	0.9566	0.8496	369	0.7949	0.3626	0.4980	171

Table 4.37: Summary of SVM classifier results on SD_CAR_DATASET

SVM SUMMARY	Max Value	Min Value	Average
Accuracy	0.7963	0.7630	0.7802
Time	0.64	0.09	0.32
Contribution of 2-Gram	1.6667%	-2.2222%	-0.2315%
Contribution of 3-Gram	0.5556%	-1.2963%	-0.6250%
Contribution of TF-IDF	2.0370%	-0.5556%	0.4167%
Contribution of Negation	2.0370%	-0.1852%	0.7716%
Contribution of POS-Tagging	1.2963%	-1.1111%	-0.2160%
Contribution of NEG + POS	2.0370%	-1.6667%	-0.3395%

Evaluation results of Support Vector Machine classifier on different combinations of SD_CAR_DATASET are shown in Table 4.36. The maximum accuracy obtained is 79.63 percent but it takes 0.12 seconds to finish, and the minimum time obtained is 0.09 seconds but its accuracy is 78.89 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.37.

Table 4.38: Evaluation results of RF classifier on SD_CAR_DATASET

RF		DATA REPRESENTATION	ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7556	0.59	0.7801	0.8943	0.8333	369	0.6667	0.4561	0.5417	171
		(1+2)-Gram	0.7556	1.12	0.7593	0.9404	0.8402	369	0.7349	0.3567	0.4803	171
		(1-2-3)-Gram	0.7296	1.29	0.7217	0.9837	0.8326	369	0.8378	0.1813	0.2981	171
	TF-IDF	(1)-Gram	0.7519	0.38	0.7594	0.9322	0.8370	369	0.7126	0.3626	0.4806	171
		(1+2)-Gram	0.7333	0.95	0.7574	0.8970	0.8213	369	0.6311	0.3801	0.4745	171
		(1-2-3)-Gram	0.7241	1.44	0.7331	0.9377	0.8228	369	0.6618	0.2632	0.3766	171
BASIC + NEG	BINARY	(1)-Gram	0.7741	0.66	0.7852	0.9214	0.8479	369	0.7290	0.4561	0.5612	171
		(1+2)-Gram	0.7574	1.16	0.7565	0.9512	0.8427	369	0.7632	0.3392	0.4696	171
		(1-2-3)-Gram	0.7370	1.47	0.7331	0.9675	0.8341	369	0.7736	0.2398	0.3661	171
	TF-IDF	(1)-Gram	0.7537	0.52	0.7658	0.9214	0.8364	369	0.6979	0.3918	0.5019	171
		(1+2)-Gram	0.7648	0.90	0.7750	0.9241	0.8430	369	0.7200	0.4211	0.5314	171
		(1-2-3)-Gram	0.7241	1.39	0.7321	0.9404	0.8233	369	0.6667	0.2573	0.3713	171
BASIC + POS	BINARY	(1)-Gram	0.7426	0.59	0.7556	0.9214	0.8303	369	0.6778	0.3567	0.4674	171
		(1+2)-Gram	0.7370	1.07	0.7350	0.9621	0.8333	369	0.7544	0.2515	0.3772	171
		(1-2-3)-Gram	0.7426	1.61	0.7366	0.9702	0.8374	369	0.7963	0.2515	0.3822	171
	TF-IDF	(1)-Gram	0.7537	0.67	0.7521	0.9539	0.8411	369	0.7639	0.3216	0.4527	171
		(1+2)-Gram	0.7259	1.11	0.7346	0.9377	0.8238	369	0.6667	0.2690	0.3833	171
		(1-2-3)-Gram	0.7093	1.37	0.7190	0.9431	0.8159	369	0.6250	0.2047	0.3084	171
BASIC + NEG + POS	BINARY	(1)-Gram	0.7556	0.70	0.7548	0.9512	0.8417	369	0.7600	0.3333	0.4634	171
		(1+2)-Gram	0.7370	1.22	0.7331	0.9675	0.8341	369	0.7736	0.2398	0.3661	171
		(1-2-3)-Gram	0.7333	1.75	0.7255	0.9810	0.8341	369	0.8293	0.1988	0.3208	171
	TF-IDF	(1)-Gram	0.7593	0.70	0.7615	0.9431	0.8426	369	0.7470	0.3626	0.4882	171
		(1+2)-Gram	0.7315	1.28	0.7489	0.9133	0.8230	369	0.6444	0.3392	0.4444	171
		(1-2-3)-Gram	0.7130	1.58	0.7220	0.9431	0.8179	369	0.6379	0.2164	0.3231	171

Table 4.39: Summary of RF classifier results on SD_CAR_DATASET

RF SUMMARY	Max Value	Min Value	Average
Accuracy	0.7741	0.7093	0.7417
Time	1.75	0.38	1.06
Contribution of 2-Gram	1.1111%	-2.7778%	-1.2963%
Contribution of 3-Gram	0.5556%	-4.0741%	-1.6204%
Contribution of TF-IDF	1.1111%	-3.3333%	-0.9414%
Contribution of Negation	3.1481%	0.0000%	1.0185%
Contribution of POS-Tagging	1.2963%	-1.8519%	-0.6481%
Contribution of NEG + POS	0.7407%	-1.8519%	-0.3395%

Evaluation results of Random Forest classifier on different combinations of SD_CAR_DATASET are shown in Table 4.38. The maximum accuracy obtained is 77.41 percent but it takes 0.66 seconds to finish, and the minimum time obtained is 0.38 seconds but its accuracy is 75.19 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.39.

Table 4.40: Evaluation results of LR classifier on SD_CAR_DATASET

LR	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7667	0.65	0.7886	0.8997	0.8405	369	0.6891	0.4795	0.5655	171
		(1+2)-Gram	0.7426	1.07	0.7376	0.9675	0.8370	369	0.7857	0.2573	0.3877	171
		(1-2-3)-Gram	0.7481	1.57	0.7412	0.9702	0.8404	369	0.8070	0.2690	0.4035	171
	TF-IDF	(1)-Gram	0.7185	0.57	0.7395	0.9079	0.8151	369	0.6092	0.3099	0.4109	171
		(1+2)-Gram	0.7519	0.87	0.7516	0.9512	0.8397	369	0.7534	0.3216	0.4508	171
		(1-2-3)-Gram	0.7556	1.26	0.7516	0.9593	0.8429	369	0.7826	0.3158	0.4500	171
BASIC+NEG	BINARY	(1)-Gram	0.7722	0.59	0.7860	0.9160	0.8461	369	0.7182	0.4620	0.5623	171
		(1+2)-Gram	0.7481	1.15	0.7495	0.9485	0.8373	369	0.7397	0.3158	0.4426	171
		(1-2-3)-Gram	0.7500	1.42	0.7407	0.9756	0.8421	369	0.8333	0.2632	0.4000	171
	TF-IDF	(1)-Gram	0.7426	0.45	0.7602	0.9106	0.8286	369	0.6633	0.3801	0.4833	171
		(1+2)-Gram	0.7259	0.71	0.7356	0.9350	0.8234	369	0.6620	0.2749	0.3884	171
		(1-2-3)-Gram	0.7204	1.26	0.7319	0.9322	0.8200	369	0.6429	0.2632	0.3734	171
BASIC+POS	BINARY	(1)-Gram	0.7315	0.64	0.7545	0.8997	0.8208	369	0.6300	0.3684	0.4649	171
		(1+2)-Gram	0.7444	1.05	0.7442	0.9539	0.8361	369	0.7463	0.2924	0.4202	171
		(1-2-3)-Gram	0.7370	1.28	0.7312	0.9729	0.8349	369	0.7959	0.2281	0.3545	171
	TF-IDF	(1)-Gram	0.7611	0.45	0.7532	0.9675	0.8470	369	0.8182	0.3158	0.4557	171
		(1+2)-Gram	0.7259	0.71	0.7387	0.9268	0.8221	369	0.6494	0.2924	0.4032	171
		(1-2-3)-Gram	0.7370	1.34	0.7302	0.9756	0.8353	369	0.8085	0.2222	0.3486	171
BASIC+NEG+POS	BINARY	(1)-Gram	0.7593	0.66	0.7698	0.9241	0.8399	369	0.7113	0.4035	0.5149	171
		(1+2)-Gram	0.7537	1.04	0.7489	0.9621	0.8422	369	0.7879	0.3041	0.4388	171
		(1-2-3)-Gram	0.7463	1.38	0.7367	0.9783	0.8405	369	0.8400	0.2456	0.3801	171
	TF-IDF	(1)-Gram	0.7519	0.63	0.7549	0.9431	0.8386	369	0.7342	0.3392	0.4640	171
		(1+2)-Gram	0.7259	1.03	0.7326	0.9431	0.8246	369	0.6769	0.2573	0.3729	171
		(1-2-3)-Gram	0.7037	1.33	0.7219	0.9214	0.8095	369	0.5797	0.2339	0.3333	171

Table 4.41: Summary of LR classifier results on SD_CAR_DATASET

LR SUMMARY	Max Value	Min Value	Average
Accuracy	0.7722	0.7037	0.7425
Time	1.57	0.45	0.96
Contribution of 2-Gram	3.3333%	-3.5185%	-1.0648%
Contribution of 3-Gram	1.1111%	-2.2222%	-0.2546%
Contribution of TF-IDF	2.9630%	-4.8148%	-1.4969%
Contribution of Negation	2.4074%	-3.5185%	-0.4012%
Contribution of POS-Tagging	4.2593%	-3.5185%	-0.7716%
Contribution of NEG+POS	3.3333%	-5.1852%	-0.7099%

Evaluation results of Linear Regression classifier on different combinations of SD_CAR_DATASET are shown in Table 4.40. The maximum accuracy obtained is 77.22 percent but it takes 0.59 seconds to finish, and the minimum time obtained is 0.45 seconds but its accuracy is 74.26 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.41.

Table 4.42: Evaluation results of VADER SA tool on SD_CAR_DATASET

VDR		DATA REPRESENTATION	ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
NO_PP		(1)-Gram	0.7792	0.00	0.8392	0.8168	0.8279	262	0.6757	0.7092	0.6920	141
BASIC		(1)-Gram	0.7275	0.00	0.7709	0.7955	0.7830	220	0.6512	0.6176	0.6340	136

Evaluation results of VADER sentiment analysis tool on different combinations of SD_CAR_DATASET are shown in Table 4.42. The best accuracy obtained is 77.92 percent and time value is zero because there is no training step.

Table 4.43: Evaluation results of DBL classifier on SD_CAR_DATASET

DBL		DATA REPRESENTATION	ACCURACY	TRAINING TIME (*)	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC		(1)-Gram	0.7114	0.83	0.7044	0.9973	0.8256	368	0.9375	0.0888	0.1622	169
BASIC NEG		(1)-Gram	0.7087	0.85	0.7017	0.9973	0.8238	368	0.9375	0.0877	0.1604	171
BASIC POS		(1)-Gram	0.7249	0.66	0.7146	0.9973	0.8326	369	0.9565	0.1302	0.2292	169
BASIC POSNEG		(1)-Gram	0.7249	0.77	0.7154	0.9946	0.8322	369	0.9200	0.1361	0.2371	169

Table 4.44: Summary of DBL classifier results on SD_CAR_DATASET

DBL SUMMARY	Max Value	Min Value	Average
Accuracy	0.7249	0.7087	0.7175
Time	0.85	0.66	0.78
Contribution of Negation	-0.2640%	-0.2640%	-0.2640%
Contribution of POS-Tagging	1.3548%	1.3548%	1.3548%
Contribution of NEG + POS	1.3548%	1.3548%	1.3548%

Evaluation results of my own Domain Based Lexicon implementation on different combinations of SD_CAR_DATASET are shown in Table 4.43. The maximum accuracy obtained is 72.49 percent and it takes 0.66 seconds to finish, and the minimum time obtained is also 0.66 seconds but its accuracy is 72.49 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.44.

Table 4.45: Evaluation results of MNB classifier on IMDB_DATASET

MNB	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.8661	22.01	0.8826	0.8471	0.8645	5041	0.8506	0.8855	0.8677	4959
		(1+2)-Gram	0.8897	80.13	0.9121	0.8645	0.8877	5041	0.8692	0.9153	0.8917	4959
		(1-2-3)-Gram	0.8936	118.67	0.9187	0.8655	0.8913	5041	0.8709	0.9222	0.8958	4959
	TF-IDF	(1)-Gram	0.8676	21.64	0.8787	0.8554	0.8669	5041	0.8569	0.8800	0.8683	4959
		(1+2)-Gram	0.8928	80.80	0.9207	0.8615	0.8901	5041	0.8679	0.9246	0.8953	4959
		(1-2-3)-Gram	0.8930	121.55	0.9211	0.8615	0.8903	5041	0.8679	0.9250	0.8955	4959
BASIC + NEG	BINARY	(1)-Gram	0.8682	21.13	0.8847	0.8492	0.8666	5041	0.8527	0.8875	0.8698	4959
		(1+2)-Gram	0.8937	73.51	0.9177	0.8669	0.8916	5041	0.8719	0.9210	0.8958	4959
		(1-2-3)-Gram	0.8970	129.01	0.9232	0.8679	0.8947	5041	0.8734	0.9266	0.8992	4959
	TF-IDF	(1)-Gram	0.8743	22.92	0.8874	0.8598	0.8734	5041	0.8618	0.8891	0.8752	4959
		(1+2)-Gram	0.8945	85.26	0.9253	0.8601	0.8915	5041	0.8673	0.9294	0.8973	4959
		(1-2-3)-Gram	0.8960	127.83	0.9268	0.8617	0.8931	5041	0.8688	0.9308	0.8988	4959
BASIC + POS	BINARY	(1)-Gram	0.8684	21.24	0.8865	0.8475	0.8665	5041	0.8516	0.8897	0.8702	4959
		(1+2)-Gram	0.8928	84.75	0.9134	0.8699	0.8911	5041	0.8738	0.9161	0.8945	4959
		(1-2-3)-Gram	0.8934	128.56	0.9152	0.8691	0.8915	5041	0.8734	0.9181	0.8952	4959
	TF-IDF	(1)-Gram	0.8761	27.06	0.8918	0.8584	0.8748	5041	0.8613	0.8941	0.8774	4959
		(1+2)-Gram	0.8943	105.31	0.9176	0.8683	0.8923	5041	0.8730	0.9208	0.8963	4959
		(1-2-3)-Gram	0.8927	124.65	0.9158	0.8669	0.8907	5041	0.8717	0.9189	0.8947	4959
BASIC + NEG + POS	BINARY	(1)-Gram	0.8758	21.23	0.8922	0.8572	0.8743	5041	0.8604	0.8947	0.8772	4959
		(1+2)-Gram	0.8962	78.70	0.9152	0.8752	0.8947	5041	0.8785	0.9175	0.8976	4959
		(1-2-3)-Gram	0.8985	123.28	0.9182	0.8768	0.8970	5041	0.8803	0.9205	0.9000	4959
	TF-IDF	(1)-Gram	0.8810	21.43	0.8958	0.8645	0.8799	5041	0.8670	0.8978	0.8821	4959
		(1+2)-Gram	0.8951	98.40	0.9174	0.8703	0.8932	5041	0.8747	0.9203	0.8969	4959
		(1-2-3)-Gram	0.8955	122.82	0.9169	0.8717	0.8937	5041	0.8758	0.9197	0.8972	4959

Table 4.46: Summary of MNB classifier results on IMDB_DATASET

MNB SUMMARY	Max Value	Min Value	Average
Accuracy	0.8985	0.8661	0.8869
Time	129.01	21.13	77.58
Contribution of 2-Gram	2.5500%	1.4100%	2.1450%
Contribution of 3-Gram	0.3900%	-0.1600%	0.1325%
Contribution of TF-IDF	0.7700%	-0.3000%	0.1625%
Contribution of Negation	0.6700%	0.1700%	0.3483%
Contribution of POS-Tagging	0.8500%	-0.0300%	0.2483%
Contribution of NEG + POS	1.3400%	0.2300%	0.6550%

Evaluation results of Multinomial Naive Bayes classifier on different combinations of IMDB_DATASET are shown in Table 4.45. The maximum accuracy obtained is 89.85 percent but it takes 123.28 seconds to finish, and the minimum time obtained is 21.13 seconds but its accuracy is 86.82 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.46.

Table 4.47: Evaluation results of SVM classifier on IMDB_DATASET

SVM	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.8676	34.32	0.8645	0.8744	0.8694	5041	0.8708	0.8607	0.8657	4959
		(1+2)-Gram	0.9070	165.84	0.9038	0.9127	0.9082	5041	0.9104	0.9012	0.9058	4959
		(1-2-3)-Gram	0.9083	206.21	0.9042	0.9151	0.9096	5041	0.9126	0.9014	0.9070	4959
	TF-IDF	(1)-Gram	0.8990	23.50	0.8896	0.9129	0.9011	5041	0.9091	0.8849	0.8968	4959
		(1+2)-Gram	0.9167	96.16	0.9081	0.9288	0.9183	5041	0.9259	0.9044	0.9150	4959
		(1-2-3)-Gram	0.9127	112.74	0.9059	0.9226	0.9142	5041	0.9199	0.9026	0.9111	4959
BASIC+NEG	BINARY	(1)-Gram	0.8758	43.66	0.8714	0.8841	0.8777	5041	0.8805	0.8673	0.8738	4959
		(1+2)-Gram	0.9080	185.66	0.9063	0.9117	0.9090	5041	0.9097	0.9042	0.9070	4959
		(1-2-3)-Gram	0.9076	217.82	0.9051	0.9123	0.9087	5041	0.9101	0.9028	0.9065	4959
	TF-IDF	(1)-Gram	0.9040	25.17	0.8969	0.9147	0.9057	5041	0.9115	0.8931	0.9022	4959
		(1+2)-Gram	0.9207	93.21	0.9139	0.9304	0.9220	5041	0.9279	0.9109	0.9193	4959
		(1-2-3)-Gram	0.9163	124.35	0.9105	0.9248	0.9176	5041	0.9223	0.9076	0.9149	4959
BASIC+POS	BINARY	(1)-Gram	0.8702	47.62	0.8691	0.8742	0.8716	5041	0.8714	0.8661	0.8687	4959
		(1+2)-Gram	0.9036	196.54	0.8988	0.9113	0.9050	5041	0.9086	0.8957	0.9021	4959
		(1-2-3)-Gram	0.9022	220.57	0.8976	0.9097	0.9036	5041	0.9070	0.8945	0.9007	4959
	TF-IDF	(1)-Gram	0.9011	24.94	0.8951	0.9105	0.9027	5041	0.9074	0.8915	0.8994	4959
		(1+2)-Gram	0.9106	93.31	0.9001	0.9254	0.9126	5041	0.9219	0.8955	0.9086	4959
		(1-2-3)-Gram	0.9039	129.15	0.8937	0.9187	0.9060	5041	0.9149	0.8889	0.9017	4959
BASIC+NEG+POS	BINARY	(1)-Gram	0.8784	49.05	0.8764	0.8834	0.8799	5041	0.8805	0.8734	0.8769	4959
		(1+2)-Gram	0.9032	203.16	0.9011	0.9076	0.9043	5041	0.9053	0.8988	0.9020	4959
		(1-2-3)-Gram	0.9051	215.99	0.9016	0.9111	0.9064	5041	0.9087	0.8990	0.9038	4959
	TF-IDF	(1)-Gram	0.9050	27.06	0.9002	0.9127	0.9064	5041	0.9100	0.8972	0.9035	4959
		(1+2)-Gram	0.9120	106.36	0.9025	0.9254	0.9138	5041	0.9222	0.8984	0.9101	4959
		(1-2-3)-Gram	0.9067	151.21	0.8985	0.9187	0.9085	5041	0.9154	0.8945	0.9048	4959

Table 4.48: Summary of SVM classifier results on IMDB_DATASET

SVM SUMMARY	Max Value	Min Value	Average
Accuracy	0.9207	0.8676	0.9019
Time	220.57	23.50	116.40
Contribution of 2-Gram	3.9400%	0.7000%	2.2588%
Contribution of 3-Gram	0.1900%	-0.6700%	-0.2375%
Contribution of TF-IDF	3.1400%	0.1600%	1.4308%
Contribution of Negation	0.8200%	-0.0700%	0.3517%
Contribution of POS-Tagging	0.2600%	-0.8800%	-0.3283%
Contribution of NEG + POS	1.0800%	-0.6000%	-0.0150%

Evaluation results of Support Vector Machine classifier on different combinations of IMDB_DATASET are shown in Table 4.47. The maximum accuracy obtained is 92.07 percent but it takes 93.21 seconds to finish, and the minimum time obtained is 23.50 seconds but its accuracy is 89.90 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.48.

Table 4.49: Evaluation results of RF classifier on IMDB_DATASET

RF	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7567	84.45	0.8071	0.6798	0.7380	5041	0.7195	0.8348	0.7729	4959
		(1+2)-Gram	0.7688	453.86	0.8117	0.7048	0.7545	5041	0.7354	0.8338	0.7815	4959
		(1-2-3)-Gram	0.7752	686.85	0.8039	0.7328	0.7667	5041	0.7508	0.8183	0.7831	4959
	TF-IDF	(1)-Gram	0.7595	58.53	0.8104	0.6826	0.7410	5041	0.7219	0.8377	0.7755	4959
		(1+2)-Gram	0.7594	329.54	0.8045	0.6905	0.7432	5041	0.7250	0.8294	0.7737	4959
		(1-2-3)-Gram	0.7466	649.27	0.8054	0.6558	0.7229	5041	0.7057	0.8389	0.7665	4959
BASIC+NEG	BINARY	(1)-Gram	0.7654	81.79	0.8173	0.6886	0.7474	5041	0.7271	0.8435	0.7810	4959
		(1+2)-Gram	0.7783	485.38	0.8152	0.7245	0.7671	5041	0.7484	0.8330	0.7884	4959
		(1-2-3)-Gram	0.7715	693.33	0.8021	0.7258	0.7621	5041	0.7459	0.8179	0.7802	4959
	TF-IDF	(1)-Gram	0.7577	64.02	0.8111	0.6770	0.7380	5041	0.7189	0.8397	0.7746	4959
		(1+2)-Gram	0.7511	465.98	0.8143	0.6558	0.7265	5041	0.7079	0.8480	0.7716	4959
		(1-2-3)-Gram	0.7384	629.30	0.7968	0.6457	0.7133	5041	0.6981	0.8326	0.7594	4959
BASIC+POS	BINARY	(1)-Gram	0.7660	100.23	0.8130	0.6959	0.7499	5041	0.7303	0.8373	0.7802	4959
		(1+2)-Gram	0.7718	383.52	0.8069	0.7195	0.7607	5041	0.7431	0.8250	0.7819	4959
		(1-2-3)-Gram	0.7680	716.13	0.8059	0.7110	0.7555	5041	0.7376	0.8260	0.7793	4959
	TF-IDF	(1)-Gram	0.7558	84.16	0.8054	0.6798	0.7373	5041	0.7191	0.8330	0.7719	4959
		(1+2)-Gram	0.7513	366.36	0.8178	0.6519	0.7255	5041	0.7066	0.8524	0.7727	4959
		(1-2-3)-Gram	0.7403	642.05	0.7936	0.6552	0.7178	5041	0.7023	0.8268	0.7595	4959
BASIC+NEG+POS	BINARY	(1)-Gram	0.7689	103.04	0.8147	0.7011	0.7536	5041	0.7338	0.8379	0.7824	4959
		(1+2)-Gram	0.7824	452.44	0.8231	0.7239	0.7703	5041	0.7500	0.8419	0.7933	4959
		(1-2-3)-Gram	0.7777	650.14	0.8101	0.7302	0.7681	5041	0.7507	0.8260	0.7866	4959
	TF-IDF	(1)-Gram	0.7755	87.66	0.8265	0.7020	0.7592	5041	0.7373	0.8502	0.7897	4959
		(1+2)-Gram	0.7600	438.72	0.8145	0.6784	0.7403	5041	0.7206	0.8429	0.7770	4959
		(1-2-3)-Gram	0.7492	700.27	0.8122	0.6536	0.7243	5041	0.7062	0.8463	0.7700	4959

Table 4.50: Summary of RF classifier results on IMDB_DATASET

RF SUMMARY	Max Value	Min Value	Average
Accuracy	0.7824	0.7384	0.7623
Time	716.13	58.53	391.96
Contribution of 2-Gram	1.3500%	-1.5500%	0.2200%
Contribution of 3-Gram	0.6400%	-1.2800%	-0.7025%
Contribution of TF-IDF	0.6600%	-3.3100%	-1.7158%
Contribution of Negation	0.9500%	-0.8300%	-0.0633%
Contribution of POS-Tagging	0.9300%	-0.8100%	-0.2167%
Contribution of NEG + POS	1.6000%	0.0600%	0.7917%

Evaluation results of Random Forest classifier on different combinations of IMDB_DATASET are shown in Table 4.49. The maximum accuracy obtained is 78.24 percent but it takes 452.44 seconds to finish, and the minimum time obtained is 58.53 seconds but its accuracy is 75.95 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.50.

Table 4.51: Evaluation results of LR classifier on IMDB_DATASET

LR	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC	BINARY	(1)-Gram	0.7572	87.06	0.8035	0.6862	0.7402	5041	0.7222	0.8294	0.7721	4959
		(1+2)-Gram	0.7756	490.47	0.8074	0.7286	0.7660	4959	0.7490	0.8234	0.7844	5041
		(1-2-3)-Gram	0.7717	658.67	0.8078	0.7179	0.7602	5041	0.7424	0.8264	0.7821	4959
	TF-IDF	(1)-Gram	0.7581	77.64	0.8025	0.6899	0.7420	5041	0.7241	0.8274	0.7723	4959
		(1+2)-Gram	0.7629	438.35	0.8186	0.6804	0.7431	5041	0.7227	0.8467	0.7798	4959
		(1-2-3)-Gram	0.7639	613.99	0.8086	0.6965	0.7484	5041	0.7296	0.8324	0.7776	4959
BASIC + NEG	BINARY	(1)-Gram	0.7746	83.36	0.8254	0.7012	0.7583	5041	0.7366	0.8492	0.7889	4959
		(1+2)-Gram	0.7770	478.03	0.8096	0.7290	0.7672	5041	0.7499	0.8258	0.7860	4959
		(1-2-3)-Gram	0.7902	596.48	0.8160	0.7538	0.7837	5041	0.7677	0.8272	0.7964	4959
	TF-IDF	(1)-Gram	0.7710	72.49	0.8160	0.7046	0.7562	5041	0.7363	0.8385	0.7841	4959
		(1+2)-Gram	0.7503	443.91	0.8009	0.6717	0.7306	5041	0.7133	0.8302	0.7673	4959
		(1-2-3)-Gram	0.7462	639.99	0.7910	0.6749	0.7283	5041	0.7124	0.8187	0.7619	4959
BASIC + POS	BINARY	(1)-Gram	0.7642	113.20	0.8118	0.6929	0.7476	5041	0.7283	0.8367	0.7787	4959
		(1+2)-Gram	0.7675	613.84	0.8004	0.7177	0.7568	5041	0.7403	0.8181	0.7773	4959
		(1-2-3)-Gram	0.7709	724.83	0.8077	0.7159	0.7591	5041	0.7411	0.8268	0.7816	4959
	TF-IDF	(1)-Gram	0.7726	98.51	0.8220	0.7007	0.7565	5041	0.7354	0.8457	0.7867	4959
		(1+2)-Gram	0.7474	513.32	0.8072	0.6554	0.7235	5041	0.7059	0.8409	0.7675	4959
		(1-2-3)-Gram	0.7491	730.02	0.8118	0.6538	0.7243	5041	0.7062	0.8459	0.7698	4959
BASIC + NEG + POS	BINARY	(1)-Gram	0.7745	117.70	0.8136	0.7169	0.7622	5041	0.7433	0.8330	0.7856	4959
		(1+2)-Gram	0.7708	634.46	0.8011	0.7255	0.7614	5041	0.7454	0.8169	0.7795	4959
		(1-2-3)-Gram	0.7813	674.60	0.8208	0.7243	0.7695	5041	0.7496	0.8393	0.7919	4959
	TF-IDF	(1)-Gram	0.7730	103.86	0.8217	0.7020	0.7572	5041	0.7362	0.8451	0.7869	4959
		(1+2)-Gram	0.7521	582.96	0.8130	0.6600	0.7286	5041	0.7099	0.8457	0.7719	4959
		(1-2-3)-Gram	0.7511	723.62	0.8084	0.6636	0.7288	5041	0.7107	0.8401	0.7700	4959

Table 4.52: Summary of LR classifier results on IMDB_DATASET

LR SUMMARY	Max Value	Min Value	Average
Accuracy	0.7902	0.7462	0.7656
Time	730.02	72.49	429.64
Contribution of 2-Gram	1.8400%	-2.5200%	-0.5200%
Contribution of 3-Gram	1.3200%	-0.4100%	0.2600%
Contribution of TF-IDF	0.8400%	-4.4000%	-1.4817%
Contribution of Negation	1.8500%	-1.7700%	0.3317%
Contribution of POS-Tagging	1.4500%	-1.5500%	-0.2950%
Contribution of NEG + POS	1.7300%	-1.2800%	0.2233%

Evaluation results of Linear Regression classifier on different combinations of IMDB_DATASET are shown in Table 4.51. The maximum accuracy obtained is 79.02 percent but it takes 596.48 seconds to finish, and the minimum time obtained is 72.49 seconds but its accuracy is 77.10 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.52.

Table 4.53: Evaluation results of VADER SA tool on IMDB_DATASET

VDR	DATA REPRESENTATION		ACCURACY	TRAINING TIME	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
NO_PP		(1)-Gram	0.7006	0.00	0.6539	0.8647	0.7447	5019	0.7946	0.5334	0.6383	4923
BASIC		(1)-Gram	0.6608	0.00	0.6241	0.8278	0.7117	4989	0.7355	0.4900	0.5881	4876

Evaluation results of VADER sentiment analysis tool on different combinations of IMDB_DATASET are shown in Table 4.53. The best accuracy obtained is 70.06 percent and time value is zero because there is no training step.

Table 4.54: Evaluation results of DBL classifier on IMDB_DATASET

DBL	DATA REPRESENTATION		ACCURACY	TRAINING TIME (*)	POSITIVE				NEGATIVE			
					Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
BASIC		(1)-Gram	0.8850	227.94	0.8891	0.8826	0.8858	4616	0.8809	0.8876	0.8842	4518
BASIC NEG		(1)-Gram	0.8887	244.95	0.9019	0.8741	0.8878	4616	0.8763	0.9036	0.8897	4553
BASIC POS		(1)-Gram	0.8889	218.10	0.8872	0.8944	0.8908	4706	0.8908	0.8835	0.8871	4591
BASIC POSNEG		(1)-Gram	0.8943	232.28	0.8920	0.9001	0.8960	4724	0.8969	0.8886	0.8927	4621

Table 4.55: Summary of DBL classifier results on IMDB_DATASET

DBL SUMMARY	Max Value	Min Value	Average
Accuracy	0.8943	0.8850	0.8892
Time	244.95	218.10	230.82
Contribution of Negation	0.3700%	0.3700%	0.3700%
Contribution of POS-Tagging	0.3900%	0.3900%	0.3900%
Contribution of NEG + POS	0.9300%	0.9300%	0.9300%

Evaluation results of my own Domain Based Lexicon implementation on different combinations of IMDB_DATASET are shown in Table 4.54. The maximum accuracy obtained is 89.43 percent but it takes 232.28 seconds to finish, and the minimum time obtained is 218.10 seconds but its accuracy is 88.89 percent. Summary of evaluation results and contribution of the techniques to the accuracy can be seen from Table 4.55.

5. CONCLUSION

The aim of this thesis was to compare some well-known machine learning approaches and lexicon-based approaches applied to sentiment analysis on twitter datasets.

Another aim was to observe the effect of preprocessing techniques and to compare some well-known data representations on twitter datasets.

When evaluation results are analysed it is obvious that the machine learning approach beats the lexicon based approach. The main reason of this result is that in machine learning approach classifiers learn directly from the given dataset while in lexicon based approach there is usually a predefined lexicon independent of the given dataset that can not always cover the dataset very well. However, in lexicon based approach there is no training phase and so no training time (accept DBL approach which spends a small time to construct a domain based lexicon) to spend and their evalution results are also acceptable.

It can also be said that SVM classifier performs best in machine learning approach. The MNB classifier follows it. Both of these classifiers are also performed very fast especially on bigger datasets according to RF and LR classifiers which are quite slow. RF and LR classifiers share the 3th place together in terms of accuracy. In lexicon-based approaches the DBL approach is slightly better than the VADER tool in terms of accurancy. However, DBL approach spends a small time to construct a domain based lexicon while VADER do not need such a process because it uses its own predefined lexicon and rules. In fact, which makes DBL perform slightly better than VADER is the construction and usage of this domain based lexicon.

In terms of data representation, it can be said that using 2-Gram in addition to 1-Gram increase the accurancy especially in bigger datasets because the usage of two word phrases are also very common. However, using 3-Gram in addition to 1-Gram and 2-Gram do not contribute to the results very much because the usage of three word phrases are not so common. Both of these usages also cause the enlargement of feature spaces and so increase in training times.

Because the preprocessing step was already performed and many features that can cause noise were already eliminated, results show that tf-idf representation does not have an important contribution comparing with binary representation.

It is also observed that negation contributes to the results becasuse negation words can influence most of the words around it and so the meaning of the text. Although it is important to know what part of speech all individual words belong to in sentences, POS-Tagging can not contribute to the results as negation. The contribution of negation and POS-Tagging together is slightly decrease the contribution of negation but increase the contribution of POS-Tagging.

These are general results that are observed in this thesis by using some specific datasets to provide a point of view. They may change according to the dataset used during the classification process, so it is useful to try all of them and then decide which techniques to use.

REFERENCES

Books

Manning C. D., Raghavan P. & Schütze H., 2008. *Introduction to information retrieval*. 1st Edition. Cambridge: Cambridge university press.

Periodical

Abbasi A., Chen H., & Salem A., 2008. Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. *ACM Transactions on Information Systems, Article 12*. pp. 1-34

Adrover C., Bodnar T., & Salathe M., 2015. Targeting HIV-Related Medication Side Effects and Sentiment Using Twitter Data. *JMIR Public Health and Surveillance.*

Asgharl M.Z., Ullah R., Ahmad S., Kundi F.M., Nawaz I.U., 2014. Lexicon based approach for sentiment classification of user reviews. *Life Science Journal*. pp. 468-473.

Bollen J., Mao H. & Pepe A., 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *In Fifth International AAAI Conference on Weblogs and Social Media*. pp. 450-453

Das S. & Chen M., 2007. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*. pp. 1375-1388.

Ding X., Liu B., & Yu P.S., 2008. Aholistic lexicon-based approach to opinion mining. *In Proceedings of the 2008 International Confer-ence on Web Search and Data Mining*. pp. 231–240

Fayyad U.M., Piatetsky-Shapiro G. & Uthurusamy R., 2003. Summary from the KDD-03 panel: data mining: the next 10 years. *In Proceedings of the Ninth International Conference on Data Mining and Knowledge Discovery: KDD-03*. pp. 191-196

Forman C., Ghose A., & Wiesenfeld B., 2008. Examining the Relationships Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research, The Interplay Between Digital and Social Networks.* pp. 291-313

Ghiassi M., Zimbra D., & Lee S., 2016. Targeted Twitter Sentiment Analysis for Brands using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. *Journal of Management Information Systems*. pp. 1034-1058.

Gautam G. & Yadav D., 2014. Sentiment analysis of twitter data using machine learning approaches and semantic analysis. *In Proceedings of the Seventh International Conference on Contemporary Computing (IC3)*. pp. 437–442.

Giachanou A. & Crestani F., 2016. Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Computing Surveys*. pp. 1-41.

Hatzivassiloglou V. & McKeown K., 1997. Predicting the semantic orientation of adjectives. *In Proceedings of annual meeting of the Association for Computational Linguistics.* pp. 174-181

Hutto, C.J. & Gilbert, E.E., 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*

Jansen B., Zhang M., Sobel K., & Chowdury A., 2009. Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*. pp. 2169-2188.

Kim S. & Hovy E., 2004, Determining the sentiment of opinions. *In Proceedings of 20th interntional conference on Computational Linguistics.* pp. 1008-1014

Kim D. & Kim J.W., 2014. Public Opinion Mining on Social Media: A Case Study of Twitter Opinion on Nuclear Power. *Advanced Science and Technology Letters (CES-CUBE 2014).* pp. 224-228

Liu Y., 2016. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing.* pp. 74-89.

Maas A.L., Daly R.E., Pham P.T., Huang D., Ng A.Y. & Potts C., 2011. Learning Word Vectors for Sentiment Analysis. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. pp 142–150

O'Connor B., Balasubramanyan R., Routledge B. R. & Smith N. A., 2010. From tweets to polls: Linking text sentiment to public opinion time series. *In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. pp. 122-129

Pang B., Lee L. & Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural*. pp. 79-86.

Taboada M., Brooke J., Tofiloski M., Voll K. & Stede M., 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*. pp. 267-307.

Thelwall M., Buckley K. & Paltoglou G., 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology 63(1)*. pp. 163-173

Wiebe J., Wilson T., and Cardie C., 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*. pp. 164-210

Turney P. D., 2002. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. *In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. pp. 417–424

Tumasjan A., Sprenger T. O., Sandner P. G. & Welpe I. M., 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. pp. 178-185

Zimbra D., Chen H., & Lusch R.F., 2015. Stakeholder Analyses of Firm-Related Web Forums: Applications in Stock Return Prediction. *ACM Transactions on Management Information* Systems. pp. 1-38.

Other Publications

Bütow F., Schultze F. & Strauch L., *Semantic Search: Sentiment Analysis with Machine Learning Algorithms on German News Articles*.

Data For Everyone, 2015. https://www.figure-eight.com/data-for-everyone/

Esuli A. & Sebastiani F., 2006. Sentiwordnet: A publicly available lexical resource for opinion mining.

Faggella D., 2018: https://www.techemergence.com/what-is-machine-learning/

GATE POS-Tagger, 2013: https://gate.ac.uk/wiki/twitter-postagger.html

Go A., Bhayani R. & Huang L., 2009. Twitter sentiment classification using distant supervision.

Kolchyna O., Souza T.T.P., Treleaven P. & Aste T., 2015. *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination.*

O'Reilly, T., 2007. *What is web 2.0: Design patterns and business models for the next generation of software*.

Shepelenko O., 2016. Sentiment analysis on tweets using ClowdFlows platform.

VADER, 2014. https://github.com/cjhutto/vaderSentiment

"Yet Another Twitter Sentiment Analysis" Article, 2018: https://towardsdatascience.com/yet-another-twitter-sentiment-analysis-part-1-tackling-class-imbalance-4d7a7f717d44