

Dissertation Report on

Sentiment Refinement by Extraction of Hidden Information from Customer Comments

Submitted to



Delhi Technological University

In partial fulfillment of the requirements for the award of the degree of

Master of Technology

In

Software Technology

By

Mishu Jain (2K16/SWT/507)

under the guidance of

Professor Dr. Rajni Jindal

Head of Department

Computer Science & Engineering

**Delhi Technological University
(Formerly Delhi College of Engineering)**

DECLARATION

I, Mishu Jain, student of MTech (SWT) hereby declare that the Dissertation **titled “Sentiment Refinement by Extraction of Hidden Information from Customer Comments ”** which is submitted by me to Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfillment of requirement for the award of degree of Master of Technology in “Software Technology“, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Delhi

Date 28.06.2020

Mishu Jain

A handwritten signature in black ink, appearing to read 'Mishu Jain', with a horizontal line extending to the right.

CERTIFICATE

On the basis of Dissertation submitted by Mishu Jain, student of MTech (SWT), I hereby certify that the Dissertation “**Sentiment Refinement by Extraction of Hidden Information from Customer Comments**” which is submitted to Department of Computer Science and Technology, Delhi Technological University, Delhi in partial fulfillment of requirement for the award of the degree of Master of Technology in “SOFTWARE TECHNOLOGY“ is an original contribution with existing knowledge and faithful record of work carried out by her under my guidance and supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Delhi

Date 28.06.2020

Professor Dr. Rajni Jindal

Head of Department

Department of Computer Science and Engineering

(Faculty Guide)

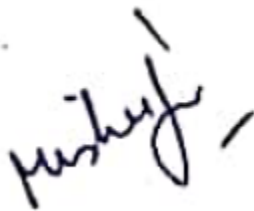
Delhi Technological University,

Delhi

Student Details:

Mishu Jain

2k16/SWT/507



28.06.2020

ACKNOWLEDGEMENT

It is my sincere thanks to all the enlisted people as well as not to enlisted people, whose help and continuous inspiration leads to get this project done in a pleasant way; however, it will be tough to thank them enough. I will nevertheless try....

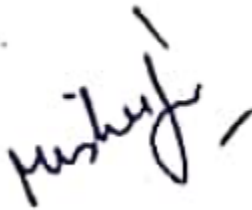
I would like to owe my gratitude and deep sense of respect towards my advisor and my guide Dr. Rajni Jindal, whose timely supervision and guidance has given me the moral boost in doing this work fine and furnished. I would like to thank her for giving me a chance to work in the field of software testing, which is really interesting and wonderful. I have enjoyed a lot in learning, in expressing my thought in free and embraced manner, for which I feel really fortunate to have guide like him.

Last but not the least I would like to thank all the staff for their sympathetic cooperation.

Thanks & Regards

Mishu Jain

28.06.2020

A handwritten signature in black ink, appearing to read 'Mishu Jain', with a horizontal line underneath.

ABSTRACT

An important area that leads to failure or success of a product lies in customer comments posted online at various social media sources. What customer will buy will vary as per pattern of what other old customers write in comments that are available online. These reviews can have different emotional nature varying from positive to negative and sometimes neutral. There are various elements and off-shelf tools that can help us to classify these posts. However, there can be various loopholes that can modify the real emotion of these comments because of Unidentified/Hidden information in neutral comments”, “Wrong keyword extraction while splitting words”, “fake comments based on frequency of duplicate comment or reviewer”. This report is use to evaluate and categorize user feedback reported on Amazon platform for e-commerce and proposes solution in the form of a flow chart and algorithm.

Data may be of many types but as per the globalization of models and techniques nowadays clustering on uncertain data have become one of the essential tasks in mining uncertain data. Earlier approaches faced the various roadblocks in modeling similarity between uncertain areas and bringing up an efficient approach of computation method. The efficient retrieval of meaningful information from the databases has always been remained as a significant issue in Data Mining.

CONTENT

ORGANISATION OF THESIS.....	8
INTRODUCTION.....	11
LIMITATIONS OF THE PREVIOUS WORK.....	12
<i>Problem 1: Extraction of Wrong Keyword and its categorization</i>	12
<i>Problem 2: Are neutral comments really neutral in true sense</i>	13
<i>Problem 3: Duplicate comments or fake comments affecting brand value</i>	13
IDEAS AND CONTRIBUTIONS.....	14
Analysis	15
Word Cloud:	15
LITERATURE REVIEW	17
METHODOLOGY	20
MINIMUM SYSTEM REQUIREMENTS	25
PROJECT DESIGN AND IMPLEMENTATION.....	29
EXPERIMENTAL RESULTS.....	41
CONCLUSION.....	47
FUTURE PROSPECTS	48
REFERENCES	49
PUBLICATION FROM THESIS.....	50

LIST OF FIGURES

- Fig. 1. Polarity and symptom wrongly classified of a split word
- Fig. 2. Neutral review information extraction
- Fig. 3. Polarity change due to meaningless and duplicate usage of words
- Fig. 4. Layout for web scraping technology
- Fig. 5. Method for Pre-Processing text
- Fig. 6. Flow of events
- Fig. 7. Positive Data Dictionary
- Fig. 8. Negative Data Dictionary
- Fig. 9. Code for Data Dictionary processing
- Fig. 10. Tool Launch
- Fig. 11. Tool connecting with website
- Fig. 12. Tool Data Saving
- Fig. 13. Method for Data extraction
- Fig. 14. Data analysis
- Fig. 15. Saved output excel
- Fig. 16. Installation
- Fig. 17. Scoring Work Flow
- Fig. 18. Symptom count
- Fig. 19. A30 Emotional polarity
- Fig. 20. Word cloud
- Fig. 21. Polarity Wise reviews
- Fig. 22. Neutral comments high count from conventional approach
- Fig. 23. Reduced neutral review count
- Fig. 24. A30 Before vs After polarity
- Fig. 25. M30 Before vs After polarity
- Fig. 26. M30 Emotional Polarity
- Fig. 27. Visual Output
- Fig. 28. Trend

ORGANISATION OF THESIS

CHAPTER 1: INTRODUCTION

This chapter focusses on identifying the real usage of data mining and its applied applications in various domains. It discusses the architecture, industrial usage and different tools that can be used to make data mining an efficient scheme.

LIMITATIONS OF THE PREVIOUS WORK

Here, we discuss about how raw data consist of a large number of parameters and how difficult it could be to gather information from raw data which is meaningful and accurate.

This section also covers 3 major problems identified as base ground for this thesis and possible solution to rectify the problems as listed below

Problem 1: Wrong keyword extraction and categorization

Problem 2: Neutral scored comments are not really neutral

Problem 3: Fake or duplicate single word comments

IDEAS AND CONTRIBUTIONS

This section talks about underlying ideas that make the algorithm work like preprocessing of data, emotional analysis, polarity of sentences etc.

CHAPTER 2: LITERATURE REVIEW

This section highlights key observations presented by various authors in the past.

Mainly below mentioned papers have been explained in this chapter.

- A. Vo, Q. Nguyen, C. Ock, Opinion Aspect Relations in Cognizing Customer Feelings via Reviews, 2018 [15]
- S.Bhat, S.Garg, ASSIGNING SENTIMENT SCORE FOR Twitter tweets, 2018 [16]

- Y.Ding, B.Li, Y.Zhao, C.Cheng, Scoring Tourist Attractions Based on Sentiment Lexicon, 2017 [17]
- Yakup yildirim, Senior Member, IEEE, and Turgay, Student Member, IEEE [12]
- Scaling Big Data Mining Infrastructure: - The Twitter Experience by Jimmy lin and Dmitriy ryaboy. [13]
- Mining user data which is large for receiving recommendations by Xavier amatriain .[14]

CHAPTER 3: METHODOLOGY

This section outlines the use of data dictionary for identification of positive and negative and neutral reviews. It also highlights our source data set and the method used to identify and segregate emotions and polarity of reviews. It explains our representative data of 5000 comments from Amazon and scored method used by our tool. The tool focusses on gathering data, removing stop words, reduction of data and preprocessing it. We identify key symptoms which are used by majority of customers and calculate a computation score.

Based on data dictionaries, reviews are classified and scored as positive or negative or neutral.

CHAPTER 4: MINIMUM SYSTEM REQUIREMENTS

This section is a brief description of coding platform used, system requirements from hardware and software side.

CHAPTER 5: PROJECT DESIGN AND IMPLEMENTATION

This chapter proposes Algorithm and Flow of events. It describes key events and method that governs the proposed polarity scoring system [18]. It gives new method for calculation of polarity and explains step by step process of extraction of reviews, preprocessing of data by removal of stop words and identification of high frequency conversation topics which is used to restore polarity shift. This chapter also shows output w.r.t emotion graph, symptom drill down, polarity calculator and word cloud.

CHAPTER 6: EXPERIMENTAL RESULTS

This section shows how neutral comments are not really neutral in nature but they have underlying information. Whenever a review is categorized as neutral, drill down function is called which helps to move earlier marked neutral comments to correct category.

The experimental result on 5000 Amazon comments that shows neutral comments were wrongly placed and have improved by 70% (on comparing neutral review count) is depicted.

CHAPTER 7: CONCLUSION

This section gives a closing statement for the thesis by mentioning that Reviews that were identified as neutral by conventional approach were rescored by proposed algorithm to shift their polarity to positive or negative.

CHAPTER 8: FUTURE PROSPECTS

This chapter mentions about future scope of research in this field. One such area is to drill down comments posted by fake consumers or competitors who might try to destroy a company's brand value by posting repeated negative comments [18].

Further there can be development related to categorizing comments written in slang language that do not really exist in dictionaries [18].

CHAPTER 9: REFERENCES & PUBLICATION FROM THESIS

17 research papers have been used for direct reference for this thesis which are accredited in this section.

Also, a research paper has been published in Nov 2019 derived from this thesis in International Journal of Recent Technology and Engineering. (M.Jain, R.Jindal, Accurate Information Extraction from Customer comments posted Online)

CHAPTER 1

INTRODUCTION

Data mining is known as process of analyzing the data. It analyses the data from different perspectives and then summarizes that data into useful information as per the user's demand. Mining allows user to analyze his data from various different dimensions, then categorize it and then summarize the relationship. Dataset may be any facts, numbers, or texts which are processed with the computer, companies are widening and accumulating vastly and growing huge amounts of data with dynamic formats. Mining is primarily used nowadays by companies with a desperate customer focus like retailing, financials, communicating and marketing organizations, this also enables companies to determine relationship among inner factors like price or product or positioning and the outer factors like competition, customer demographics and economic indicators. While huge scale information technology have been evolved with separate transactions and many analytical system, the mining tool and software analyzes relationship and the patterns which are stored transactions based upon open-ended users and their queries. Fast retrieval for the appropriate information has been an issue as different techniques have been developed for this purpose. In this era where age of available information and the size of processed data is in terabytes or petabytes and day by day this volume of data is on increase so this explosive growth has made the field very dynamic for mining information from data. There is a comprehensive journey beginning from data to spontaneous information extraction age that is coming ahead.

Around the world, various organizations make advantage by using information that is available online in order to understand their customers. These Companies might get information about their products by getting necessary information, vibration of emotional pattern from these online post. Thereby affecting the success or failure of a product because these comment can directly influence the buyers in future who has easy access of this information posted online on Social Media. Hence, this concept of opinion mining that has real information hidden underlying to be extracted from textual comments, provides paths for future businesses success [1]. But their exist a problem to identify accurate information. Various tools such as RapidMiner [2] exist and are proposed in the past but these tools do not take care of underlying problems of distribution of polarity. Each algorithm can categorise reviews or comments in "Positive", "Negative" or "Neutral" category.

However, we are well aware that a positive review can be slightly positive mixed with slightly negative information or vice versa.

Multiple Researches done in different languages for sentiment analysis that have been published, for example, Thai [7], Arabic [4], etc. These researches calculates polarity score from emotion word described in data dictionary and thereby marks them into respective sentiment category [8][7][10][9]. However, if the information processed is really accurate cannot be done manually and thus this problem arises from that a review marked in one category may fall into another if it has hidden emotion which might lie under another category. As we are aware that English language has contextual meanings, a review that is categorized as negative could have an overall positive or vice versa impact due to context-based usage of sentences [18].

LIMITATIONS OF THE PREVIOUS WORK

In earlier studies data analysis is performed with various parameters huge number of extensions have been used some of them were the traditional methods. But not a lot done on uncertain data. An uncertain data needs some similarity metrics for that certain methods and terminologies have been used in this project report. Since the object is certain at some points but when the size increases the size of the data set defines its uncertainty which may be a challenging task for the researchers , even the organizations have to face this due to the versatility of the market and the wideness of the market and customer or the user.

This project report covers the three principals which were not considered in earlier works with such a sharp-edge cut way [18].

Problem 1: Extraction of Wrong Keyword and its categorization

False categorization can happen as many comments that are posted online can fall can fall into false category due to wrong splitting of words or keyword selection. A word that consist of composition of 2 meaning full words can be wrongly understood by the tool. Consider below reviews that were posted online by a customer in Fig. 1.

Symptoms	Reviews	Polarity Old	Polarity New
Network,Battery,Camera,Heat,Display,Sensor	excellent, weight, display quality super excellent battery life and quality: i did a video conference call using mobile hot spot from this phone and the battery was barely discharged, plus the phone did not heat up at all the speed is excellent too i have not used the fingerprint reader, and only used the basic features of the camera so cannot make a proper assessment the price is a little high	Neutral	Positive

Fig. 1. Wrong Polarity and symptom wrongly classified of a split word

This review uses a term “Hotspot” commonly used in mobile technology. However, when a common off-shelf tool will try to split it, it will break it into Hot + Spot. This will imply that the something is getting hot which can make the meaning if this review to be negative however the real polarity of this review lies in different category (neutral polarity in this case).

Problem 2: Are neutral comments really neutral in true sense

Since most of the algorithm design in a language are lexicon based [5][6], it is possible that many comments which fall into neutral category due to their overall score being ‘0’, can have hidden information that might be capable of tilting them towards being either positive or negative.

Consider below mentioned example

Symptoms	Reviews	Polarity Old	Polarity New
Sound,Battery,Camera,Sensor	battery life is good in camera wide angle is most effective sound quality is good but there are no vibration on fingerprint camera in day light is amazing but low light in camera is notgood	Neutral	Positive

Fig. 2. Neutral review information extraction

On analyzing above mentioned example. Here a customer is implying Positive meaning in his/her review however generic Lexicon based algorithm [5] will calculate overall neutral score because of +1 score for positive and -1 score for negative. End result that is calculated comes out to be a zero for a neutral review despite the fact that in actual this review is implying positive meaning.

Problem 3: Duplicate comments or fake comments affecting brand value

There can be single word comments that do not really infer anything but can tilt or tweak the

polarity graph that is plotted. As can be seen below in Fig. 3 some irrelevant comments can be counted as neutral however they should simply be discarded to plot the accurate polarity of a product.

Symptoms	Reviews	Polarity Old	Polarity New
	nothing	Neutral	NA
	nothing	Neutral	NA
	plz call me	Neutral	NA

Fig. 3. Polarity change due to meaningless and duplicate usage of words

The extraction of information from the data have never been an easy task as there are lots of attributes to go through but these principal approaches take them all in confidence. But the approaches previously existed only depended on the semantics but now in this work we have focused on all the instances of the data.

Since the object is certain at some points but when the size increases the size of the data set defines its uncertainty which may be a challenging task for the researchers, even the organizations have to face this due to the versatility of the market and the wideness of the market and customer or the user [18].

IDEAS AND CONTRIBUTIONS

- Extract Feedback from websites to a local file, that can be later processed and analyzed
- SentimentMiner uses selenium for Web Scraping



Fig. 4. Layout for web scraping technology

- **Pre-processing** is a Collection of activities in which Text Documents are pre-processed.
- **Tokenization:** Breaking a stream of text into words.

- **Stop Words Removal:** The words which are not useful in analysis like “and” , “are” , “this” etc. are removed.
- **Stemming:** Removal of same forms of word, grammar etc.

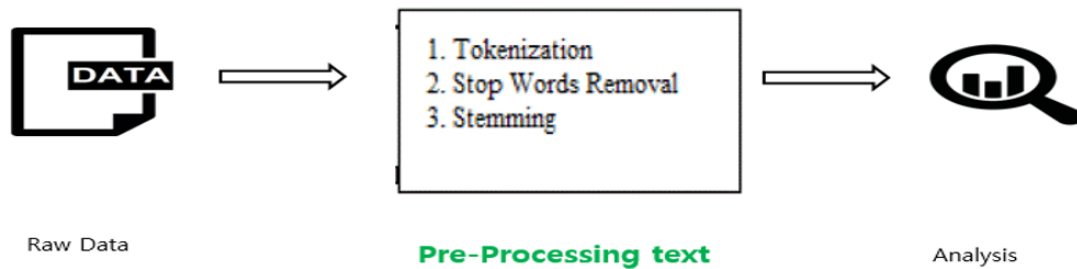


Fig. 5. Method for Pre-Processing text

Analysis

Word Cloud:

- Frequency of words is calculated
- Trending terms are highlighted.
- Word Cloud is formed based on frequency of occurrence (Higher frequency word get bold)
- R Library is used

Polarity:

- Lexical Dictionary are used to categorise Positive and Negative sentiments.
- Sentence scores are calculated based on positive and negative scores
- Polarity identification based on Sentence scores
 - If Score > 0 => Positive
 - If Score < 0 => Negative
 - If Score = 0 => Neutral
- Used Python

Emotion Analysis:

- Analyse emotions of the texts
- Emotions like Joy. Sadness. Surprise. Fear. Disgust. Anger is categorised
- Using NLTK library of R

Drilldown negative reviews:

- Sentences are traversed based on symptoms like Battery, camera, heat etc.
- Lexical Dictionary is used on the symptoms for Negative Drilldown
- Scoring is done based on negative words dictionary.
- Used Java

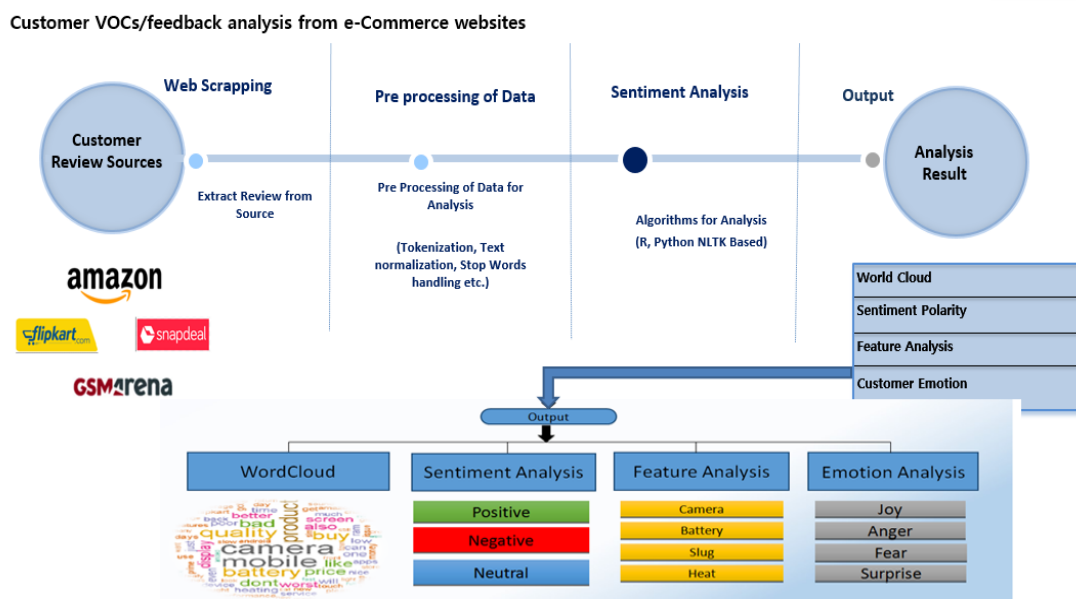


Fig. 6. Flow of events

CHAPTER 2

LITERATURE REVIEW

Amongst many popular research areas that exist in field of mining, one that is becoming increasingly popular is opinion mining. One major road block in the path of opinion mining is the structure of the data. If the data has missing structure or has a lot of outliers, then mining can be a tedious task. Natural Language processing (NLP) is one field of opinion mining that is used for segregation of emotions and sentiments in a sentence and its analysis by a relevant dictionary for language [1],[7],[4],[11],[8].

NLP has been used in various researches to assign sentiment score on comments that are posted on twitter [5] and helps in scoring based on emotion present in a lexicon [6] which explains how words can be scored in correct category by giving them a positive, negative or neutral score thereby calculating score for the entire statement. Both the papers [5],[6] explains the methods to find out polarity based on content by firstly identifying opinion words, assigning positive, negative or neutral score and then summing up these scores to get overall polarity [18].

A. Vo, Q. Nguyen, C. Ock, Opinion Aspect Relations in Cognizing Customer Feelings via Reviews, 2018 [15]

This paper presents a way to get and summarize aspects of a product and related opinions from database containing large number of reviews on a product. The focus is to maximize accuracy by using a 2 stage approach of extraction of knowledge and analysis of sentiment. Extraction of knowledge is performed using natural language processing tools. Aggregate statistics from raw knowleged is used in second stage to do infer aspect relationship.

S.Bhat, S.Garg, Assigning Sentiment score for twitter tweets, 2018 [16]

This paper presents the scoring system used to assign sentiment score to tweets. It uses twitter large amount of data and gives score on the basis of various emotions like anger, jou, happiness, hatred etc.. Positive score is given to positive emotions and negative score is assigned to negative emotions to identify overall polarity of the tweet.

Y.Ding, C.Cheng, Y.Zhao, B.Li, Tourist Attractions Scoring Based on Sentiment Lexicon, 2017 [17]

This paper gives a key application of identifying lexicon based sentiments and on the basis of which tourist attractions are rated. Same concept has been used widely in various other industries to rate products, locations, restaurants, hotels and many more.

Yakup yildirim, Senior Member, IEEE, and Turgay, Student Member, IEEE [12]

This paper presented how the low-level features and raw datasets are not enough for aligning to user requirement but there is also a need to understand the meaning & semantic aspects of any content or information. Manmade techniques, that are not sufficient and can be predictive or subjective along with are sometimes prevent to be expensive with respect to time and its ability to find answers to queries, that are being used as a medium to diminish the distance between low-level feature representation and its corresponding high-level information from content. They also proposed a content extraction method that is based on meanings and semantics which enables the analyst to query information enabling concepts derived from events to be extracted with minimum manual effort, they also introduced a fuzzy content model method based on ontology and video output that uses spatial relations with time and event occurrence for definition of concepts. That ontology definition provided a domain applicable rule construction standard that allows the user to construct an ontology for the prescribed task. Therefore, it is able to analyze complex scenarios which are eventually tested on three domains of different nature. This is done by usage of additional definitions of rules that help in reducing cost of computation and achieve better precision.

Scaling Big Data Mining Infrastructure: - The Twitter Experience by Jimmy lin and Dmitriy ryaboy. [13]

The paper presented insights about Big Data mining infrastructures and the experiences of doing analysis at Twitter type sites. This also depicts that as the real-time status of the data mining tools varies, it is not direct to carry out analysis. Mostly that time is consumed in preparatory work to the application of data mining technologies and turning preliminary models into robust models.

Mining user data which is large for receiving recommendations by Xavier amatriain .[14]

The paper presented information that are processed with the Netix prize and discussed the recommended methods which can be personalized and used. It discussed recently occurred important problems and future research areas. This contained a curious discussion about if one

requires large data or better models to improve learning methods then what all techniques must be followed.

CHAPTER 3

METHODOLOGY

Data was gathered for representation purpose for 5K comments posted on Amazon and scoring of these comments was done by our tool. The tool is used to gather data and here forward it will remove words that are called stop words. Later, data is reduced in order to preprocess it. Major symptoms were identified and stored in data dictionary that was used by tool to compute score. These symptoms were based on most frequently used words to describe an object that were used by customers.

Calculated score review was done and it was observed that with respect to various conventional approaches that were published in earlier researches [5][6], which did not focus on the 3 main loopholes problems i.e. Extraction of Wrong Keyword and its categorization, Are neutral comments really neutral in true sense, Duplicate comments or fake comments affecting brand value.

The solution to these problems was taken care in algorithm proposed by us.

The programs were implemented in python. These experiments were conducted on a computer with an Intel core i5, G570, 4 GHz CPU and 5 GB main memory.

Data Dictionary:

Data Dictionary used here act as a central repository of bag of words both positive and negative. This information stored in these data dictionary mainly contains symptoms of most frequently used words in online reviews about data such as meaning, relationships to other data, origin, usage.

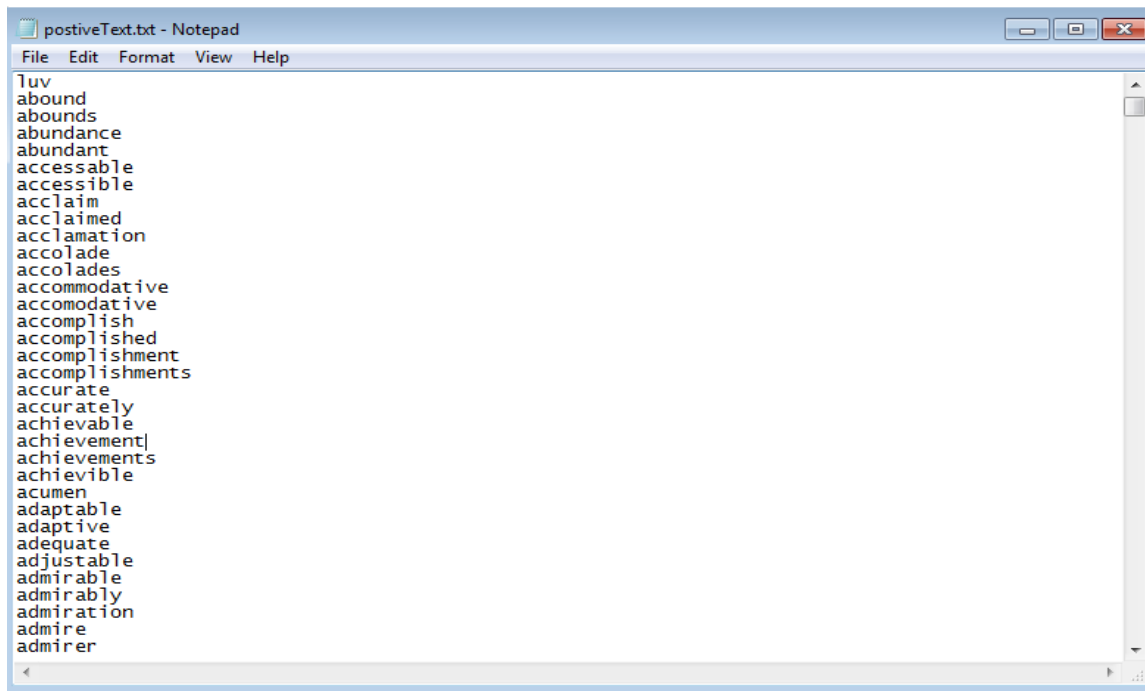


Fig. 7. *Positive Data Dictionary*

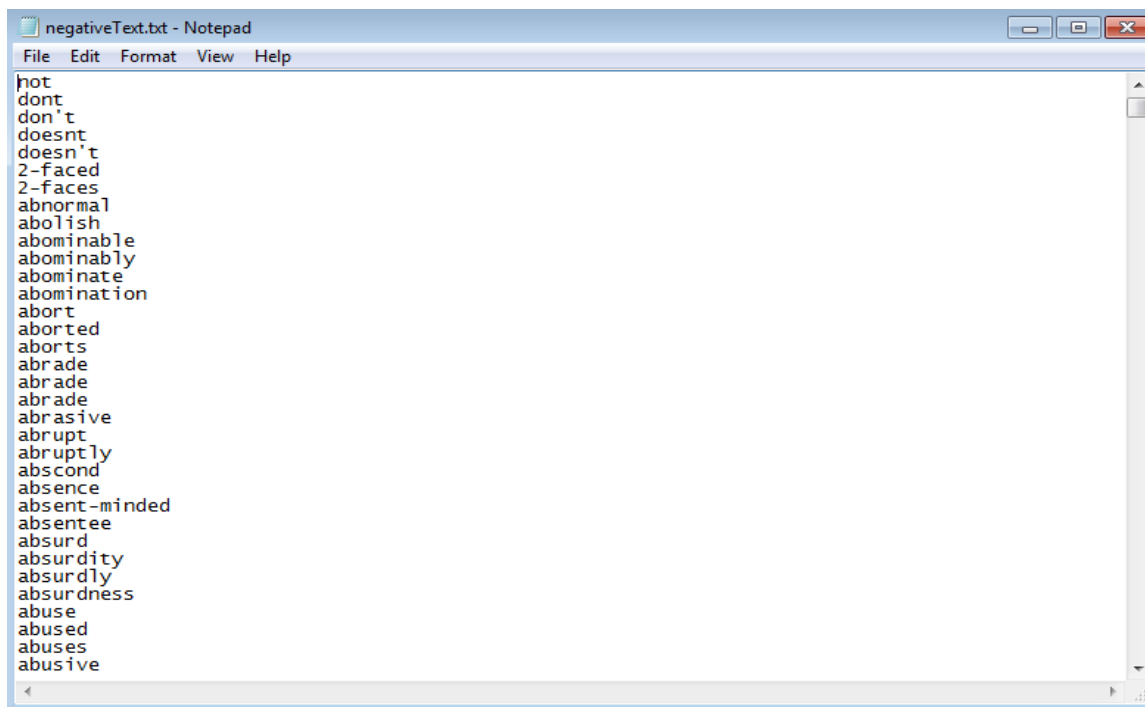


Fig. 8. *Negative Data Dictionary*

Above dictionary are regularly updated based upon customer preferences, product high selling keypoints and based on these dictionaries, reviews are classified and scored as positive or negative or neutral. Scoring code depicted below shows how positive and negative words are appended and sorted after comparing with positive and negative data dictionary.

```
f = open('C:\\Program Files\\SentimentMiner\\RDocs\\EmotionsGrabber\\positiveText.txt')
for line in f:
    if len(line.strip().split(" ")) > 1:
        multWord.append(line.strip())
        posWords.append(line.strip().replace(" ", ""))
f.close()

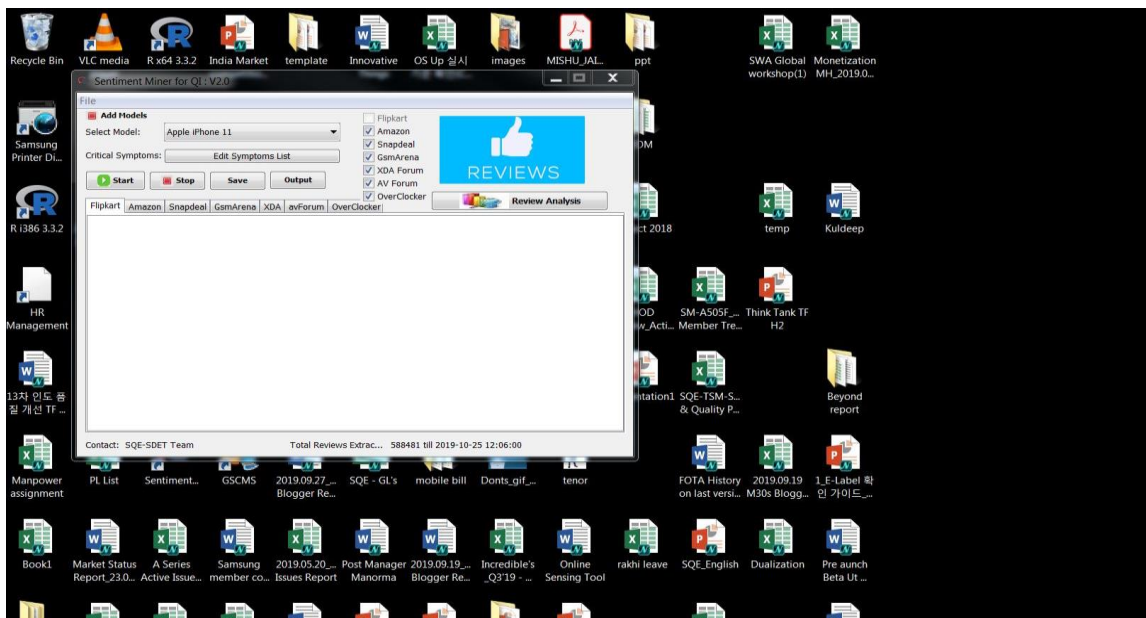
f = open('C:\\Program Files\\SentimentMiner\\RDocs\\EmotionsGrabber\\negativeText.txt')
for line in f:
    if len(line.strip().split(" ")) > 1:
        multWord.append(line.strip())
        negWords.append(line.strip().replace(" ", ""))
f.close()

posWords.sort()
negWords.sort()

feedbacklen = len(feedbacklines)-1
for tempmultWord in multWord:
    for i in range(0,feedbacklen):
        if(feedbacklines[i].find(tempmultWord) != -1):
            feedbacklines[i] = feedbacklines[i].replace(tempmultWord,tempmultWord.replace(" ", ""))
            #print tempmultWord
            #print feedbacklines[i]+"\\n"
```

Fig. 9. Code for Data Dictionary processing

The data set once selected needs to be performed on the tool the tool window look similar as the following.



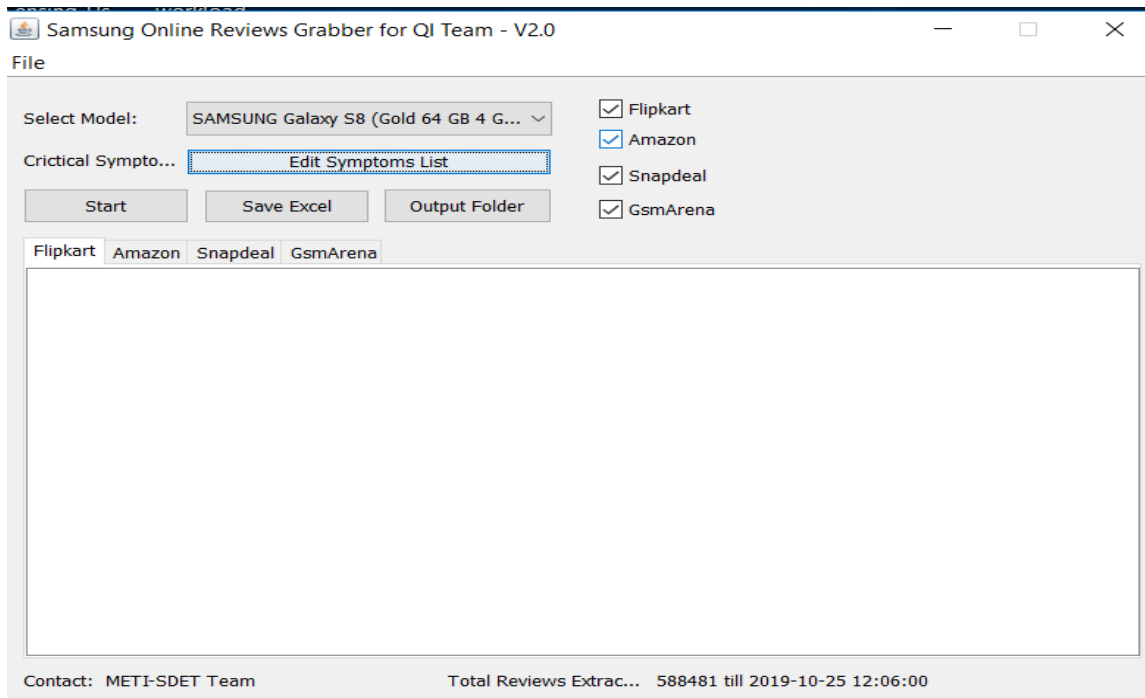


Fig. 10. Tool Launch

Choose the web page where we need to carry on the process of review extraction. We may change the data dictionary based on our need. Now run the algorithm. The following screen shots are taken in regard to show working of tool.

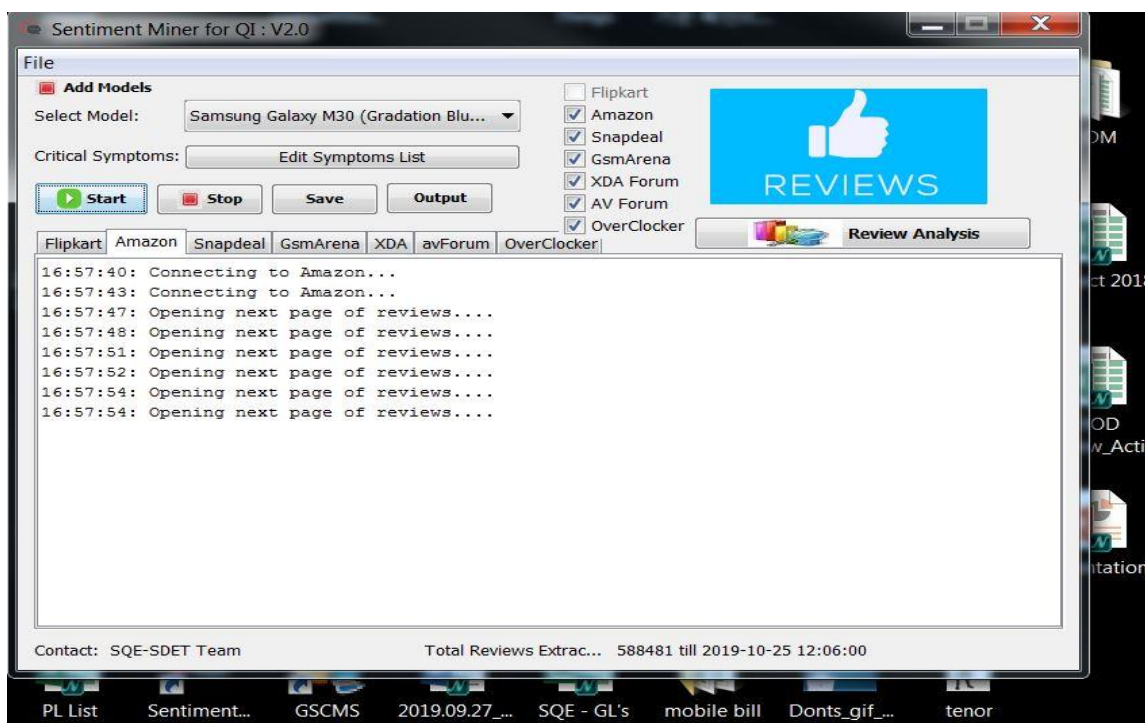


Fig. 11. Tool connecting with website

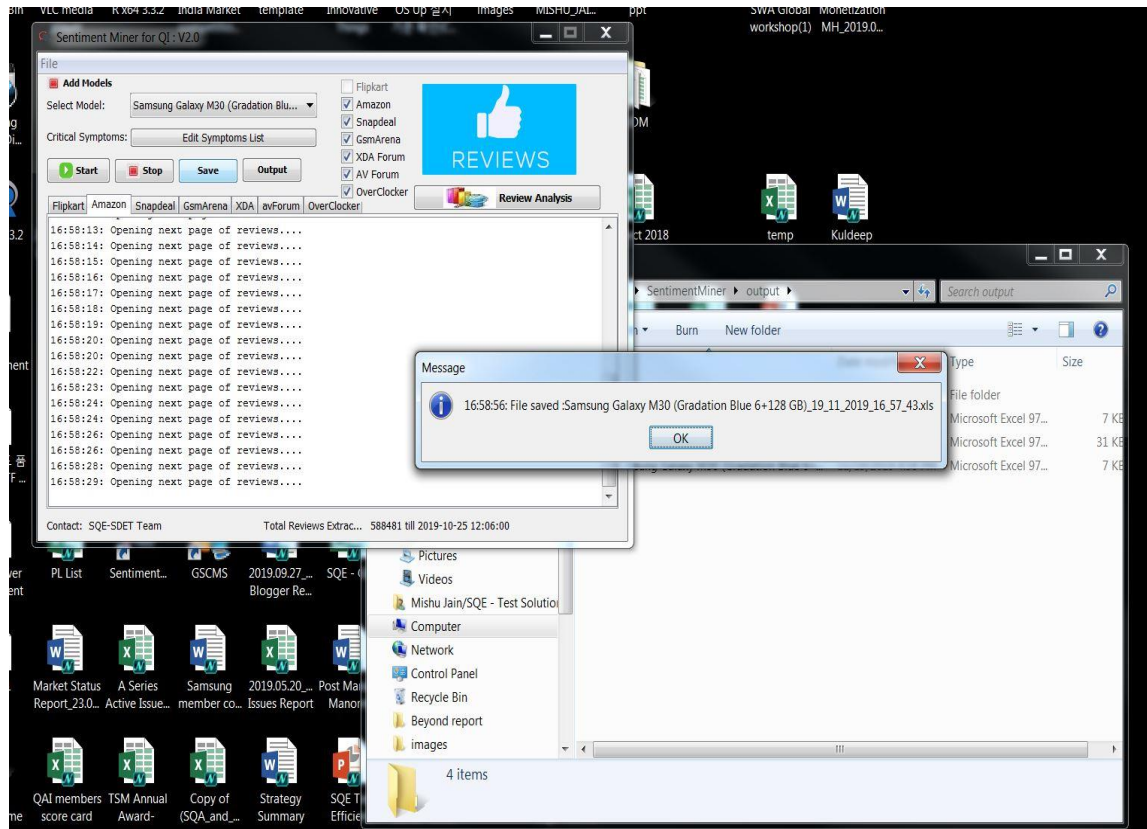


Fig. 12. Tool Data Saving

CHAPTER-4

MINIMUM SYSTEM REQUIREMENTS

SOFTWARE REQUIREMENTS

Prerequisites and Installation

Prerequisite:

Sentiment Miner t requires JDK 8, R, Python and chrome driver. These all dependencies automatically get installed by installing Sentiment Miner Setup.exe

The environmental variables will automatically be set in the path.

Installation:

Steps to follow for installing Sentiment Miner:

- 1) Click on the SentimentMiner.exe and perform the installation process.
- 2) Shortcut of Sentiment Miner Will be created on desktop and window's Start.

Procedure to Execute Sentiment Miner Tool

1. Click on Sentiment Miner Icon on desktop .Tool will be launched

DATA EXTRACTION

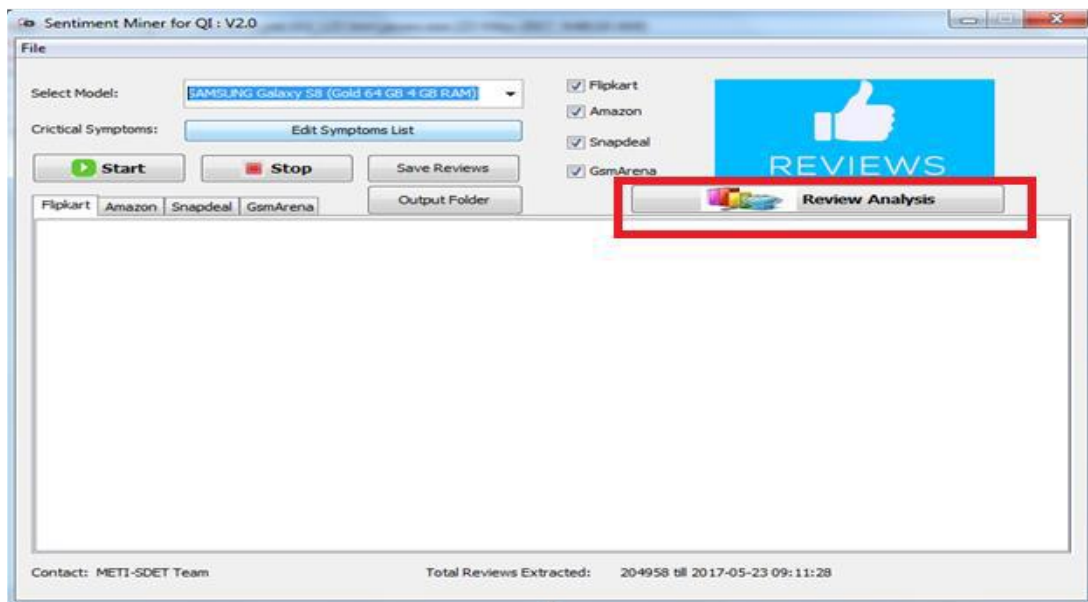


Fig. 13. Method for Data extraction

- a. Select the model from select model dropdown
- b. Select resource (Flipkart, Amazon, Snapdeal, GSMarena)
- c. Click on 'Start' Button for Collection of data from E-Retailers and Mobile Forum.
- d. Click on 'Stop' Button once fetching of reviews is Done.
- e. Click on 'Save Reviews'

After Saving Reviews Extraction of Data is done.

DATA ANALYSIS

- 1) Click on Review Analysis a window will pop up for sentiment analysis.

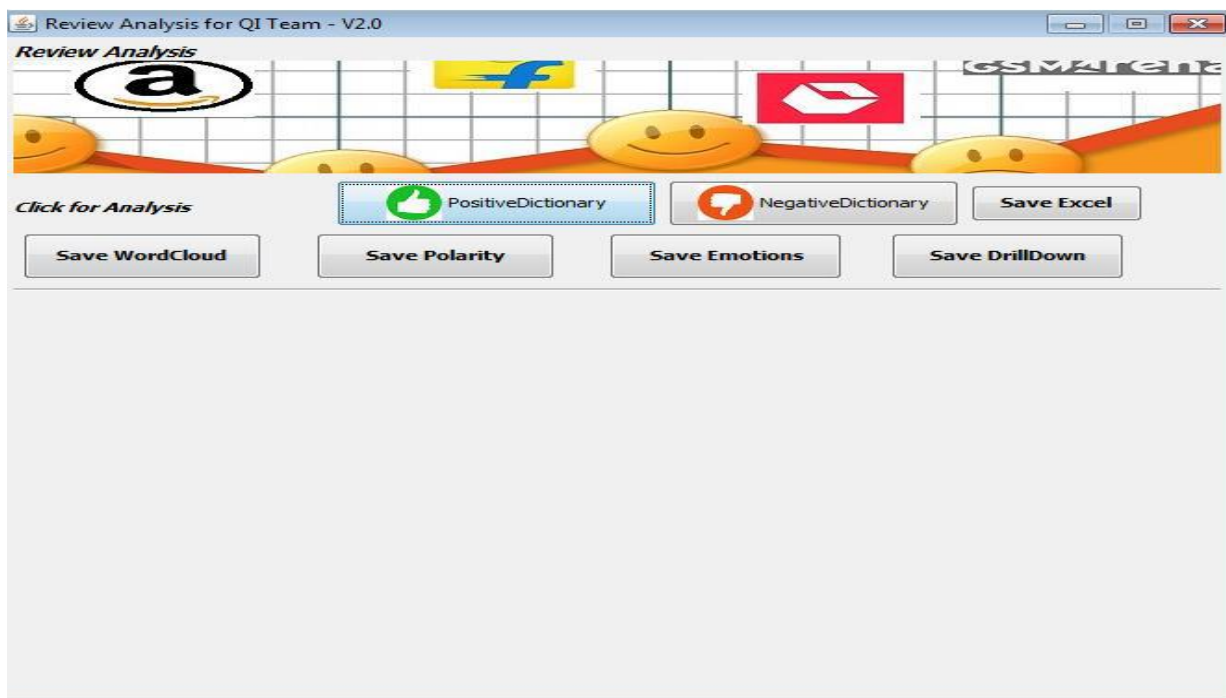


Fig. 14. *Data analysis*

- a. Click on Save Word Cloud Button for capturing word cloud of reviews.
- b. Click on Save Polarity button for Saving Polarity (Positive, Negative, Neutral)
- c. Click on Save Emotions button to show Emotions of Reviews (Joy, anger, sad, fear)
- d. Click on Save Drilldown to save feature wise Negative reviews.
- e. Click on Save Excel to save the output in Excel format.

Once Excel is saved it can be viewed by clicking on OUTPUT button of first window









	Samsung GALAXY J7 Prime_17_4_2017_12...	17-04-2017 12:08	Microsoft Excel 97...	276 KB
	Oppo F1s_13_4_2017_16_46_06.xls	13-04-2017 17:11	Microsoft Excel 97...	655 KB
	Samsung GALAXY A7_13_4_2017_15_24_2...	13-04-2017 15:26	Microsoft Excel 97...	126 KB
	Samsung GALAXY A7 (2017)_13_4_2017_1...	13-04-2017 12:20	Microsoft Excel 97...	49 KB
	Samsung GALAXY C9 Pro_13_4_2017_11_...	13-04-2017 11:24	Microsoft Excel 97...	384 KB
	Samsung GALAXY C9 Pro_13_4_2017_11_...	13-04-2017 11:15	Microsoft Excel 97...	332 KB
	Samsung GALAXY J2 Ace_13_4_2017_10_5...	13-04-2017 11:00	Microsoft Excel 97...	80 KB
	Samsung GALAXY A7_13_4_2017_09_57_1...	13-04-2017 09:58	Microsoft Excel 97...	161 KB

Fig. 15. Saved output excel

Troubleshooting Steps

While launching SentimentMiner.exe, if it does not executes check the log folder in C:\Program Files\Sentiment Miner\logs\logc.txt.

If Error like "Registry key Error: Java version has value '1.8', but '1.7' is required" comes, kindly remove all java.exe, javaw.exe and javaws.exe from C:\Windows\System32, C:\Windows\SysWOW64.

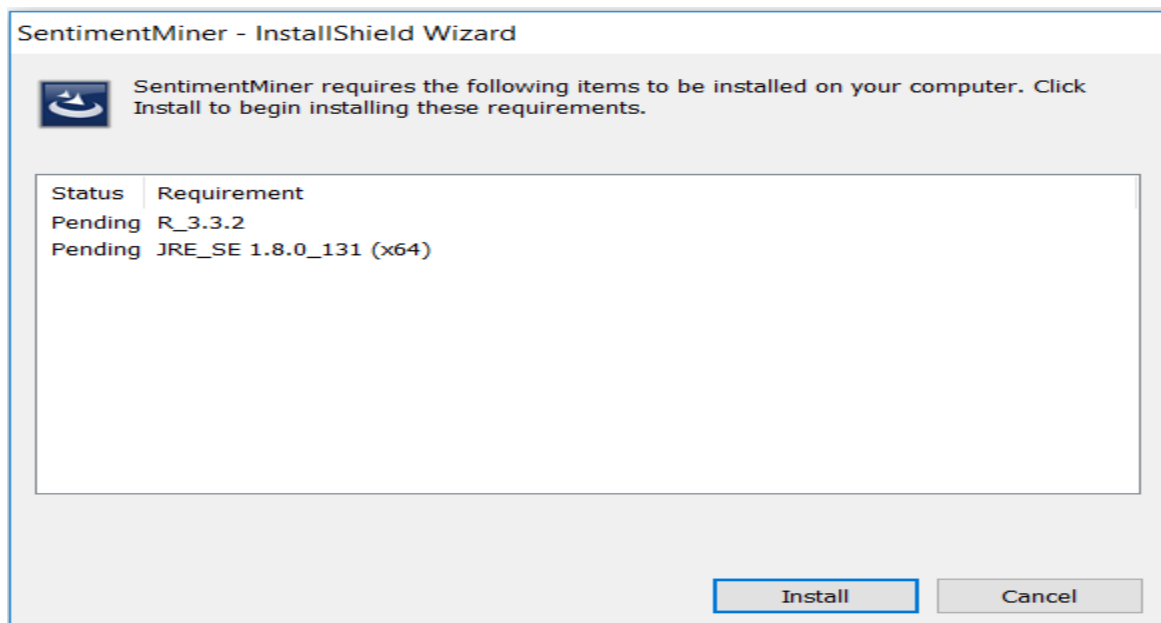


Fig. 16. Installation

- R Library
- Python
- Operating System Windows 7, Windows Server 2008

4.2 HARDWARE REQUIREMENTS

The minimum hardware platform is

Processor	Intel core i5, G570, 4 GHz CPU
RAM Memory	5GB
Hard Disk	2GB

CHAPTER 5

PROJECT DESIGN AND IMPLEMENTATION

Solution: Algorithm and event flow

Fig.1 shows a flowchart that can be used to describe major events and key method to take into consideration for scoring system to calculate polarity [18].

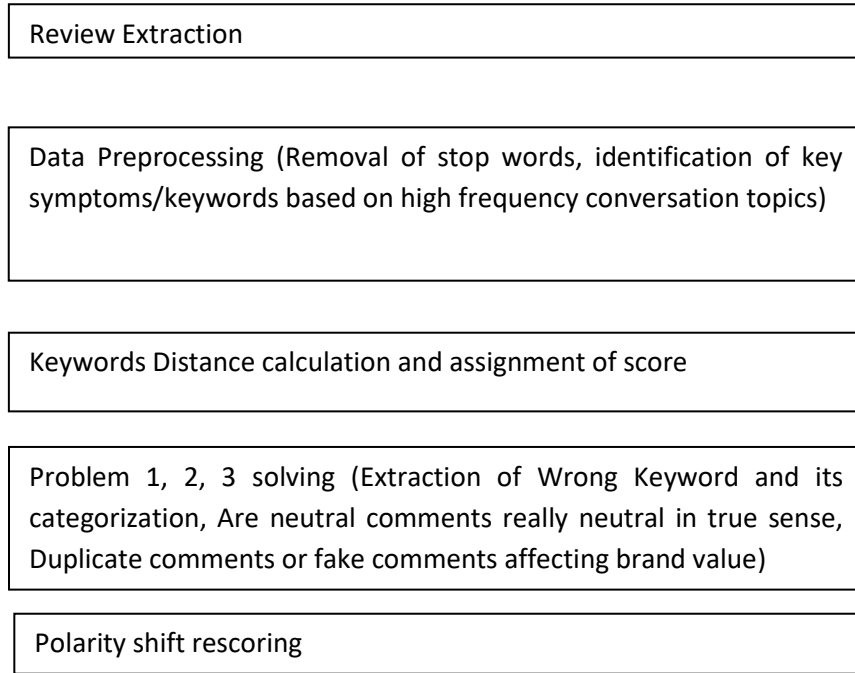


Fig. 17. Flow chart for scoring

TABLE I. ALGORITHM

Algorithm	Online review Polarity score calculator
Input	<i>Online reviews</i>
Output	<i>Refined Polarity score</i>
Step 1	Identify frequently used keywords. Extract keyword K_i
Step 2	Distance calculator () { Calculate distance of adverb/adjectives from K_i (keyword)

Algorithm	Online review Polarity score calculator
Input	<i>Online reviews</i>
Output	<i>Refined Polarity score</i>
	Assign S_i (Score) to nearest Distance D_i Assign Emotional intensity E_i
Step 3	Calculate Sum S_i to define Polarity P For $S_i < 0$: Assign Negative Polarity count as N For $S_i > 0$, Assign Positive Polarity count as P For $S_i = 0$, Assign Neutral Polarity count as N_u Call function Neutral Drill Down ()
Step 4	Neutral Drill Down () { Calculate frequency of Negative words Update N_i Calculate frequency of Positive words Update P_i Link Emphasis words E_i to N_i & P_i Recalculate and update score S' If $N_i > P_i$ && $S' < 0$, Assign New Polarity $P' = N$ else if $N_i < P_i$ && $S' > 0$, Assign New $P' = P$ else Assign $P' = N_u$ } Print New Polarity P'

Code snippet w.r.t ALGORITHMS

Drill Down

Below mentioned code is used for drilling down of symptoms. The drilldown code counts the number of occurrences and updates frequency of occurrence of symptoms. This is used to plot graph as described in subsequent snippet.

```
setwd("C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber")

# required packages
#####install.packages("http://107.108.168.111/ftms/QI_VOC_GRABBER/sentiment.tar.gz"
, repos = NULL, type="source")
library(tm)
library(sentiment)
library(plyr)
library(ggplot2)
library(wordcloud)
library(RColorBrewer)
library(foreign)
library(stringi)
library(Cairo)

path <- paste (getwd(), "/backtrackDrilldown.txt", sep="")
conn <- file(path,open="r")
lines <- readLines(conn)

for (i in 1:length(lines)){
  hotCount = as.numeric(lines[1])
  restartCount = as.numeric(lines[2])
  soundCount = as.numeric(lines[3])
  displayCount = as.numeric(lines[4])
  slugCount = as.numeric(lines[5])
  sensorCount = as.numeric(lines[6])
  batteryCount = as.numeric(lines[7])
  cameraCount = as.numeric(lines[8])
  networkCount = as.numeric(lines[9])
}

#plot graph
df <-
data.frame(Parameters=c("Heat", "Restart", "Sound", "Display", "Slug", "Sensor", "Battery", "Cam
era", "Network"),
           len=c(hotCount, restartCount,
soundCount, displayCount, slugCount, sensorCount, batteryCount, cameraCount, networkCount))
```

```

#####make DrillDown Graph#####
#drillDownImage = paste ("/images/drillDown_", currentTime, sep="")
#drillDownImage = paste (drillDownImage, ".png", sep="")

drillDownImage = paste ("/images/drillDown", ".png", sep="")

#png(filename=drillDownImage, width=500, height=500)
Cairo(file=drillDownImage, type="png",units="in", width=6.2, height=4, pointsize=8,
dpi=92)
ggplot(data=df, aes(x=reorder(Parameters,-len), y=len, fill=Parameters)) +
  geom_bar(stat="identity") +
  #scale_fill_brewer(palette="Dark2") +
  labs(x="Parameters", y="No of emotions") +
  ggtitle("Parameters wise Negative reviews") +
  geom_text(aes(label=len), vjust=-0.3, size=3.5)+
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=12,face="bold"))
dev.off()

```

SAMSUNG GALAXY J2 Ace - ONLINE REVIEWS SUMMARY												
S. No	Symptoms	Flipkart	Amazon	Snapdeal	GsmArena	Total		S. No	Symptoms	Rating (0,1,2,3)	Rating (4,5)	Total
1	Sluggish	2	15	2	0	19		1	Sluggish	4	15	19
2	Sound	1	3	1	0	5		2	Sound	1	4	5
3	Network	0	13	1	0	14		3	Network	3	11	14
4	Battery	3	34	6	0	43		4	Battery	6	37	43
5	Camera	3	36	9	0	48		5	Camera	6	42	48
6	Heat	1	17	3	0	21		6	Heat	3	18	21
7	Restart	0	1	0	0	1		7	Restart	0	1	1

Fig. 18. Symptom count

Emotion graph

Emotion graph is plotted based on word categorization as happy, sad, joy, anger etc.. based upon signs like exclamation marks, feelings, emotions. These are subsequently plotted in a graph.

```
setwd("C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber")

# required packages
library(tm)
library(sentiment)
library(plyr)
library(ggplot2)
library(wordcloud)
library(RColorBrewer)
library(png)
library(Cairo)

# Read the text file
filePath <- paste (getwd(), "/vocText.txt", sep="")
some_txt <- readLines(filePath)
#remove extra spaces from text file
some_txt <- some_txt[sapply(some_txt, nchar) > 0]

emotionfileConn<-file(paste(getwd(),"emotionsBacktrack.txt", sep=""))

# define "tolower error handling" function
try.error = function(x)
{
  # create missing value
  y = NA
  # tryCatch error
  try_error = tryCatch(tolower(x), error=function(e) e)
  # if not an error
  if (!inherits(try_error, "error"))
    y = tolower(x)
  # result
  return(y)
}
# lower case using try.error with sapply
some_txt = sapply(some_txt, try.error)

# remove NAs in some_txt
some_txt = some_txt[!is.na(some_txt)]
names(some_txt) = NULL
```

```
##### For Emotion Graph #####
# classify emotion
class_emo = classify_emotion(some_txt, algorithm="bayes", prior=1.0)
# get emotion best fit
emotion = class_emo[,7]
# substitute NA's by "Neutral"
emotion[is.na(emotion)] = "Neutral"

#data frame with results
emo_df = data.frame(text=some_txt, emotion=emotion, stringsAsFactors=FALSE)

##### writing reviews and emotions in text file
write.table(emo_df,emotionfileConn,sep="\t",row.names=FALSE)

#sort data frame
emo_df = within(emo_df, emotion <- factor(emotion, levels=names(sort(table(emotion),
decreasing=TRUE))))

currentTime <- format(Sys.time(), "%Y_%m_%d_%H_%M_%S")

#####plot distribution of emotions#####
#emotionImage = paste ("./images/emotions_", currentTime, sep="")
#emotionImage = paste (emotionImage, ".png", sep="")

emotionImage = paste ("./images/emotions", ".png", sep="")

#png(filename=emotionImage, width=500, height=500)
Cairo(file=emotionImage, type="png",units="in", width=6, height=4, pointsize=12, dpi=96)
ggplot(emo_df, aes(x=emotion)) +
  geom_bar(aes(y=..count.., fill=emotion)) +
  scale_fill_brewer(palette="Dark2") +
  labs(x="Emotions", y="No of emotions") +
  geom_text(stat="count", aes(y = ..count..+1.7, label=..count..))+
  ggtitle("Emotion wise reviews")
dev.off()
#####
```

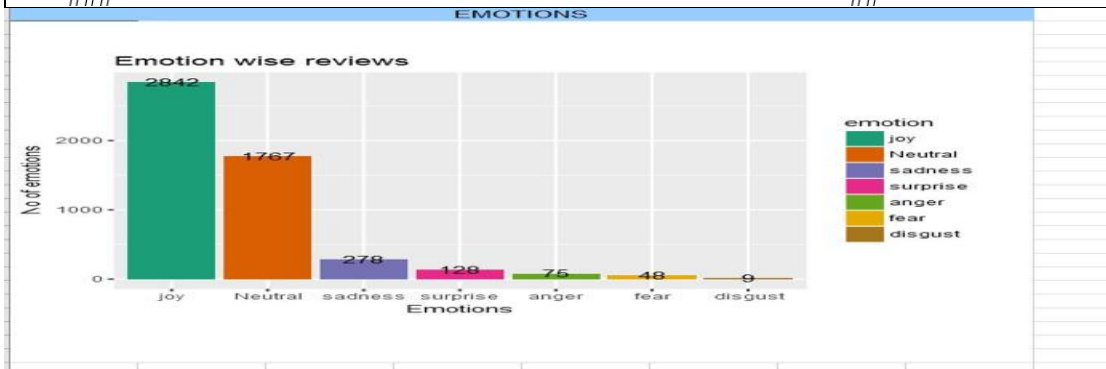


Fig. 19. A30 Emotional polarity

Polarity

As explained in algorithm, data dictionary used for calculating frequency of positive, negative & neutral reviews is used to identify sentiment polarity and rescoring is done for neutral reviews based upon identifying detailed inclination of reviews.

```
setwd("C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber")

library(ggplot2)
library(syuzhet)
library(Cairo)

path2 <- paste (getwd(), "/polarityCount.txt", sep="")
conn2 <- file(path2,open="r")
lines2 <- readLines(conn2)

for (i in 1:length(lines2)){
  positiveCount = lines2[1]
  negativeCount = lines2[2]
  neutralCount = lines2[3]
}
####Convert string into numeric count####
positiveCount <- as.numeric(positiveCount)
negativeCount <- as.numeric(negativeCount)
neutralCount <- as.numeric(neutralCount)

df <- data.frame(Emotion=c("Positive", "Neutral", "Negative"),
                 len=c(positiveCount,neutralCount, negativeCount))

polarityImage = paste (".\\images\\polarity", ".png", sep="")

Cairo(file=polarityImage, type="png",units="in", width=6, height=4, pointsize=12, dpi=96)
ggplot(data=df, aes(x=reorder(Emotion,-len), y=len, fill=Emotion)) +
  geom_bar(stat="identity") +
  scale_fill_manual(values=c("red1", "green4", "skyblue4")) +
  labs(x="Polarity", y="No of emotions") +
  ggtitle("Polarity wise Reviews") +
  geom_text(aes(label=len), vjust=-0.3, size=3.5)+
  theme_minimal()
dev.off()
```

Word cloud

In the below code, data stemming is done, text is converted to lower case & punctuations are removed. Based upon the frequency of occurrence, font size and placement near to center is assigned to high frequency words. Lower frequency words are assigned smaller fonts and moved away from center. Positive and negative words are colored accordingly.

```
##set working Directory
setwd("C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber")

library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
print("comes here")
# Read the text file
filePath <- paste (getwd(), "/vocText.txt", sep="")
text <- readLines(filePath)
# Load the data as a corpus
docs <- Corpus(VectorSource(text))

toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
docs <- tm_map(docs, toSpace, "@")
docs <- tm_map(docs, toSpace, "\\")

# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords,
c("click", "expand", "iphone", "just", "will", "smartphone", "mi", "note", "redmi", "vivo", "oppo", "samsung",
"snapdeal", "flipkart", "plus", "on8", "amazon", "mobile", "phone", "prime", "grand", "galaxy", "ace", "pro"))
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
# Text stemming
docs <- tm_map(docs, stemDocument)

dtm <- TermDocumentMatrix(docs)
```


Polarity calculator

```
posWords = []
negWords = []
multWord = []
feedbacklines = []
feature = {}
f = open('C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber\\vocText.txt')
for line in f:
    feedbacklines.append(line.strip().replace("?", " ").replace("/", " ").replace("\\", " ").replace("=", " ")
    ).replace("#", " ").replace("-", " ").replace("_", " ").replace(".", " ").replace(" ", " ").lower())
f.close()

f = open('C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber\\positiveText.txt')
for line in f:
    if len(line.strip().split(" ")) > 1:
        multWord.append(line.strip())
        posWords.append(line.strip().replace(" ", ""))
f.close()

f = open('C:\\Program Files\\Sentiment Miner\\RDocs\\EmotionsGrabber\\negativeText_1.txt')
for line in f:
    parts = line.strip().split("##")
    feature[parts[0].strip().replace(" ", "")] = parts[1]

    if len(parts[0].strip().split(" ")) > 1:
        multWord.append(parts[0].strip())
        negWords.append(parts[0].strip().replace(" ", ""))
f.close()
posWords.sort()
negWords.sort()

feedbacklen = len(feedbacklines)-1
for tempmultWord in multWord:
feedbacklen = len(feedbacklines)-1
for tempmultWord in multWord:
    for i in range(0,feedbacklen):
        if(feedbacklines[i].find(tempmultWord) != -1):
            feedbacklines[i] =
feedbacklines[i].replace(tempmultWord,tempmultWord.replace(" ", ""))
            #print tempmultWord
            #print feedbacklines[i]+"\\n"
def binarySearch(alist, item):
    first = 0
    last = len(alist)-1
    found = False

    while first<=last and not found:
```

```

midpoint = (first + last)//2
    if alist[midpoint] == item:
        found = True
    else:
        if item < alist[midpoint]:
            last = midpoint-1
        else:
            first = midpoint+1
return found

file = open("C:\\Program Files\\SentimentMiner\\RDocs\\EmotionsGrabber\\backtrack.txt","w")
positiveCount = 0
negativeCount = 0
neutralCount = 0
import re
for tempFeedback in feedbacklines:
    d = {}
    for c in re.findall(r"[\w']+)", tempFeedback):
        print(c)
        try:
            d[c.strip()] += 1
        except:
            d[c.strip()] = 1
score = 0;
polarity = ""
for k in d.keys():
    t = binarySearch(posWords,k)
    if t:
        score += (d[k]*1)
    t = binarySearch(negWords,k)
    if t:
        print(k.strip().replace(" ",""))
        score += (d[k]*int(feature.get(k.strip().replace(" ",""))))
if score > 0:
    polarity = "Positive"
    positiveCount += 1
else:
    if score < 0:
        polarity = "Negative"
        negativeCount += 1
    else:
        polarity = "Neutral"
        neutralCount += 1

file = open("C:\\Program Files\\SentimentMiner\\RDocs\\EmotionsGrabber\\backtrack.txt","a+")

```

```

file.write(tempFeedback+"\t"+str(score)+"\t"+polarity+"\n")

polarityFile = open("C:\\Program
Files\\SentimentMiner\\RDocs\\EmotionsGrabber\\polarityCount.txt", "w")
polarityFile = open("C:\\Program
Files\\SentimentMiner\\RDocs\\EmotionsGrabber\\polarityCount.txt", "a+")
polarityFile.write(str(positiveCount)+"\n"+str(negativeCount)+"\n"+str(neutralCount-1))

```

POLARITY-WISE REVIEWS			
S.NO	Symptoms	Reviews	Polarity
1		fastest delivery in the country best device in the world right now and amazing support i wish there was something new i could say, but i could just sum up saying that you guys are the best all the best! :)	Positive
2	Camera	best smartphone ever never thought this phone would be so more fantastic to hold apart from pictures flipkart delivered it 3 days before so happy to buy from flipkart its next generation phone kind of feel that it is from mars jk	Positive
3	Battery	no words to describe the beauty of this phone best design ever created battery life is best worth every penny	Positive
4	Battery,Camera,Heat,Display,Sluggish	i have always been a fan of samsung phones don't want to drag too much about the delivery after a day of use, here are few things about the phone the phone is jus a beast and runs without any lag the fone got heated a little bit heating is negligible as most of the fones these days get heated as u use them i don't use	Negative
5	Display,Sluggish	everything from display to performance is just awesome!! this 5* rating won't do any justice!	Positive
6		the product has some problems in software & i also did not get my free wireless charger, now i need to replace this product and get another replaced product alongwith the free charging instrument	Neutral
7	Network,Battery,Camera,Display,Sensor,Sluggish	first impressions after using it for 2 daysawesome camerasuper displaybattery life is really goodnavigation keys still getting used to as coming from sb edgeoverall performance of the device is really smooth and enjoyablebeen using iris instead of finger print and it works really well and is fastfew things that bother me or	Neutral

Fig. 21. Polarity Wise reviews

CHAPTER 6

EXPERIMENTAL RESULTS

Amazon e-commerce platform was used to analyze a product performance and customer reviews. 5K comments were extracted for a popular product and comparative study was done by previous approach [5][6] with respect to proposed approach and a score was assigned to each comment to calculate overall emotion of the comments to be negative, neutral or positive,. It was still observed that neutral comments have underlying information that was not yet disclosed so the algorithm was modified in second to refine these type of comments.

Stop words, grammatical sequences such as “a, an, the” are segregated. This method of text refinement and dictionary based processing of identified words is done. Data dictionary of negative and positive words was created using approach used in [3]. After assigning positive score and negative score to respective word category, total cumulative score is calculated that helps to establish overall polarity of any statement. Emphasis words identification was also used such as the one that really slow on long term use ”, then multiple emphasis on “really” indicates high emotional intensity E_i and hence more word “slow” in this case is assigned a little higher negative score.

If we look at fig.22 we will observe that neutral review count is still too high. Hence, we modified proposed algorithm further to enhance the polarity results as shown in Fig. 23[18].

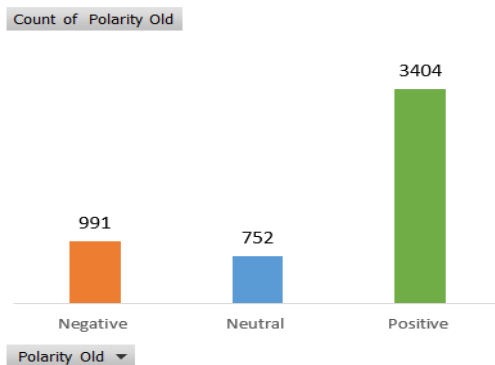


Fig. 22. Neutral comments high count from conventional approach

Table. 1 give an algorithm to further refine neutral comments by drilling them in 2nd iteration. The reason for doing so is that neutral comments are not necessarily neutral in nature but can tilt towards positive or negative. In both the scenarios, this information can be extracted.

Whenever a review polarity comes out to be neutral, a new implemented drill down function is called using if-else condition. This drill down module checks the occurrences of emphasis words along with positive words & negative words to calculate scores and frequency again for those particular statements. This enables correct categorization of earlier neutral comments to later positive or negative category.

The experimental result on Amazon comments (approximately more than 5K) is shown in fig 6. On comparing with new output represented in Fig. 22, it is clearly visible that neutral reviews occurrences have dropped to 294. This enables us to identify that those comments which were initially neutral were actually wrongly placed and the output now has improved by 70% (comparative study of Fig.22 with respect to Fig.23) [18].

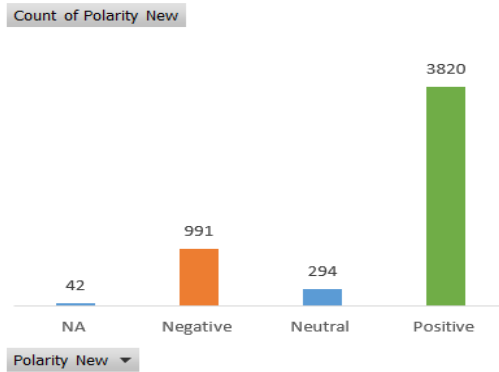
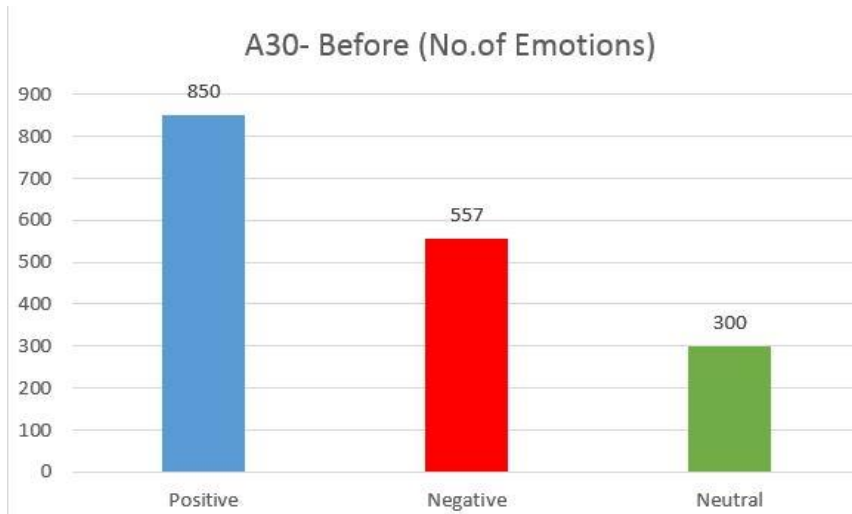


Fig. 23. Reduced neutral review count

6.1 Model analysis

Study 1: Model analysis for few more data sets is studied via tool. Amazon reviews for A30 are extracted. Approximately 1707 reviews are identified by conventional approach into respective polarity



On applying proposed algorithm, it is observed that neutral review count significantly reduces from 300 as before to 215 neutral reviews as in below figure. This depicts that almost 75 reviews are further categorized into positive and negative and moved from neutral polarity.

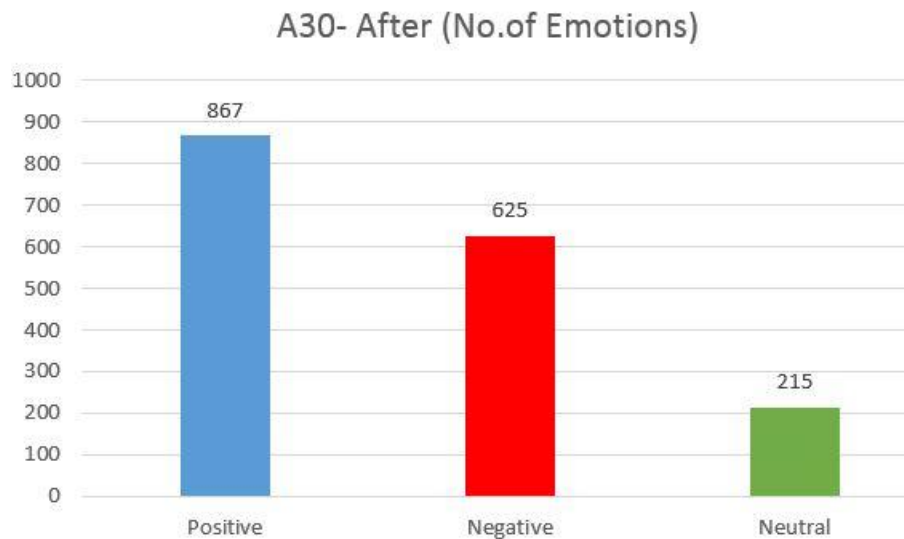
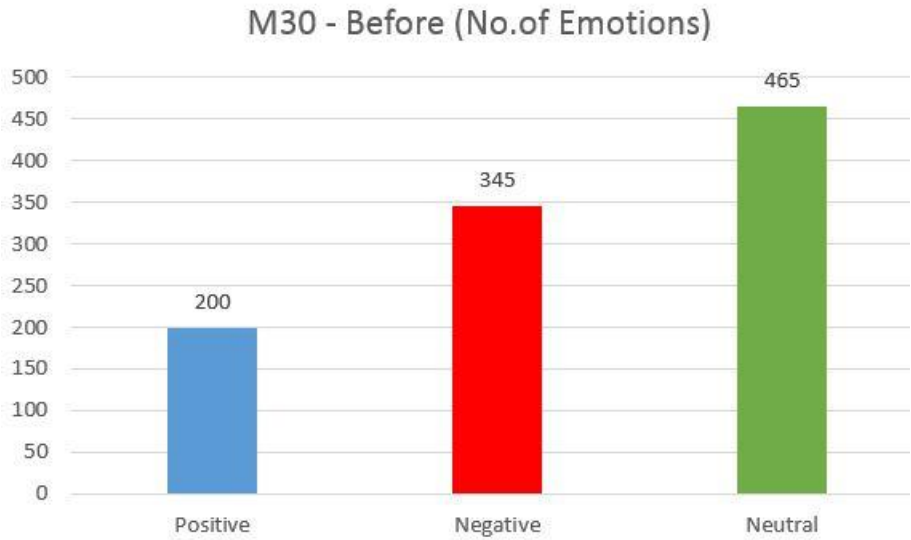


Fig. 24. A30 Before vs After polarity

Study 2: Flipkart reviews for M30 are extracted. Approximately 1010 reviews are identified by conventional approach into respective polarity



On applying proposed algorithm, it is observed that neutral review count significantly reduces from 465 as before to 203 neutral reviews as in below figure. This depicts that almost 262 reviews are further categorized into positive and negative and moved from neutral polarity.

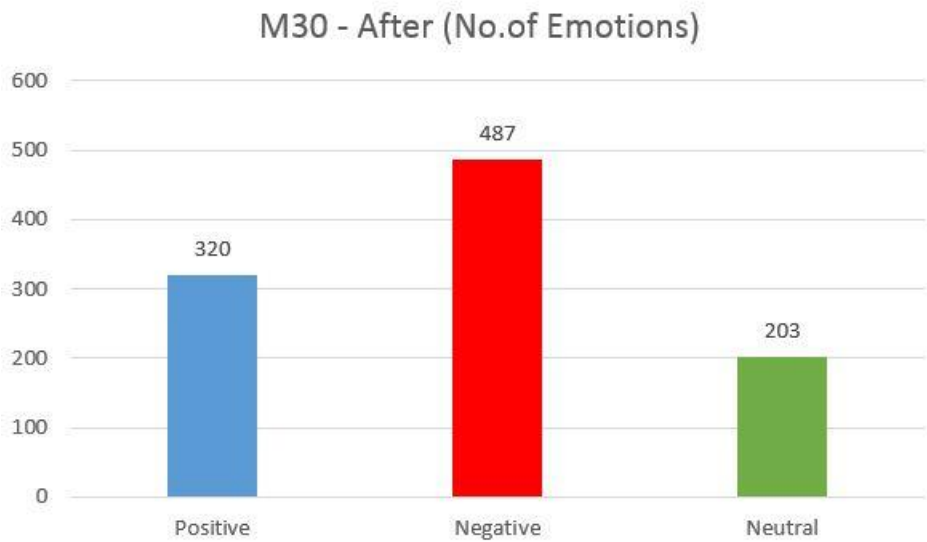


Fig. 25. M30 Before vs After polarity

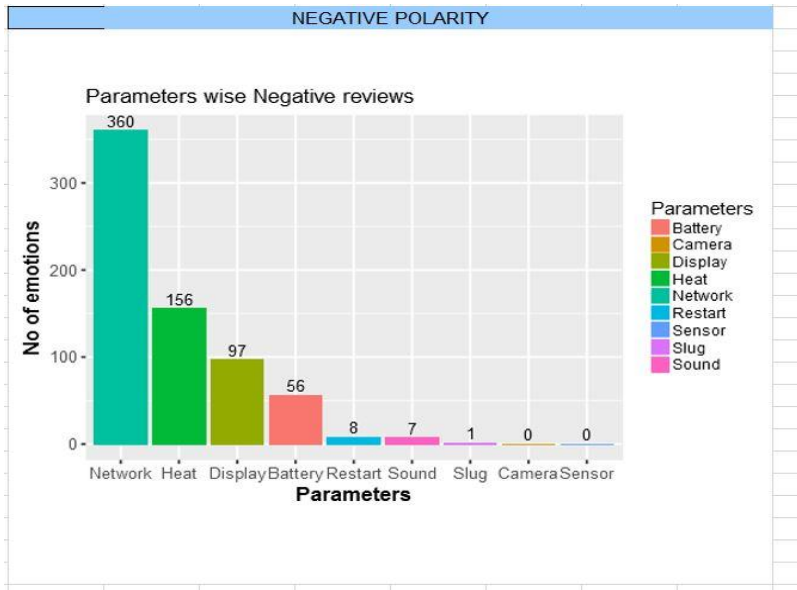


Fig. 26. M30 Emotional Polarity

VISUAL OUTPUT



Fig. 27. Visual Output

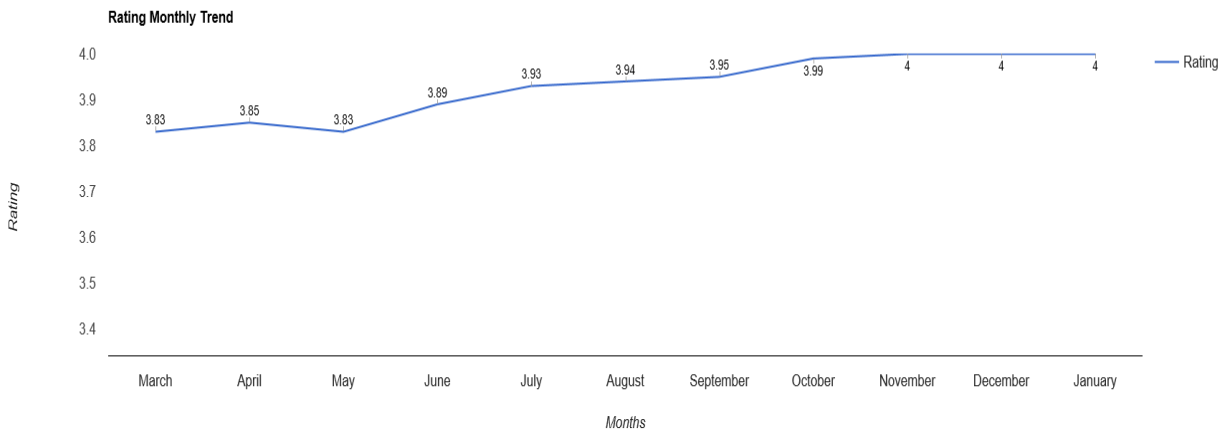
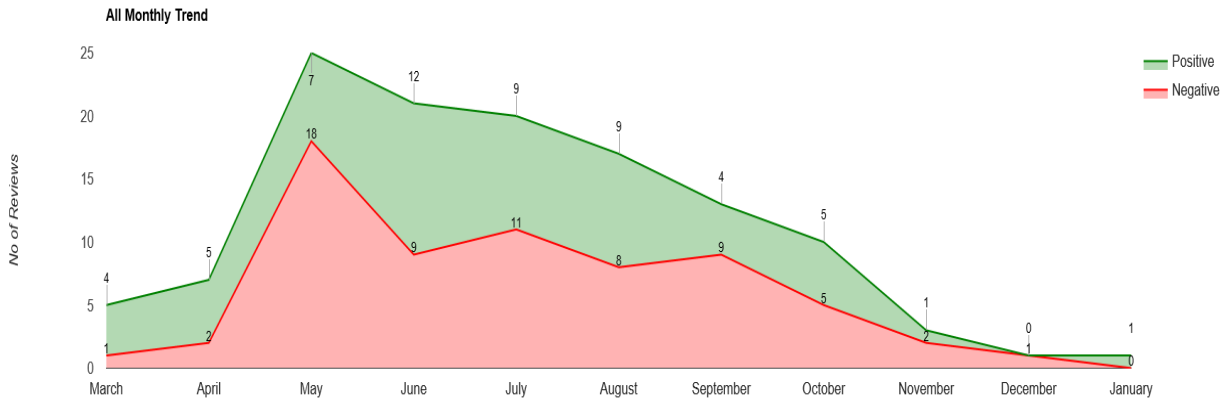
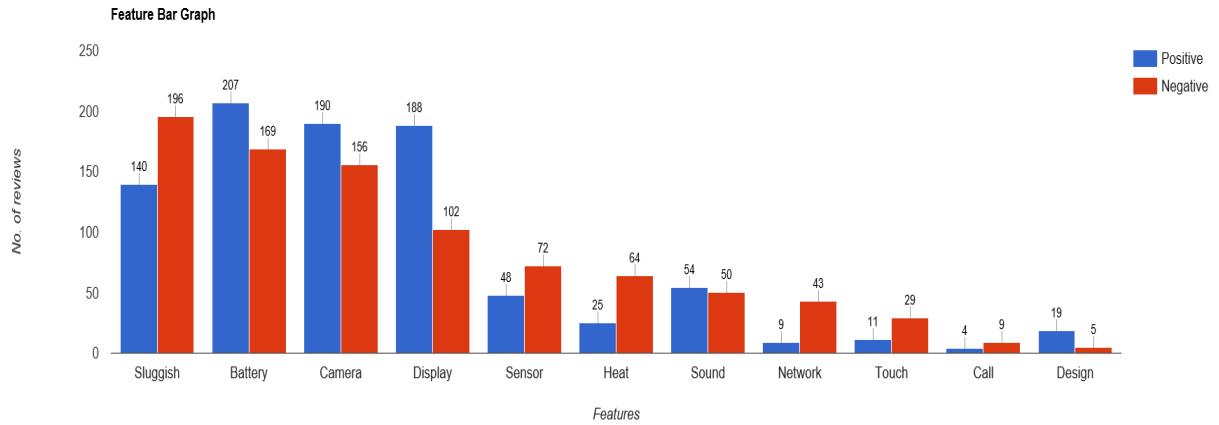


Fig. 28. Trend

CHAPTER 7

CONCLUSION

This report helps in placement of earlier calculated neutral reviews into positive or negative category or discarding meaningless comments. A 2nd iteration drill down approach is used. The outline is that based on emotional intensity, a neutral comment can have underlying information which gets ignored due to conventional mathematical score-based calculation scheme. A clearer picture of a product performance is observed by re-categorizing those into positive or negative polarity [18].

Various social networking sites and e commerce websites were used as data source for extracting reviews. Reviews that were identified as neutral by conventional approach were rescored by proposed algorithm to shift their polarity to positive or negative.

CHAPTER 8

FUTURE PROSPECTS

A vast scope of research exist in this area of opinion mining. An area that was identified and can be taken up in future is for those fake comments that posted by competitors to destroy brand value. These fake comments can have negative impact for future consumers and market [18]. If a fake comment identifier can be created that can help influence consumers to take unbiased decision.

Another area is to identify source of reviewers that can help eliminate duplicate comments that tweaks the real results.

Further Slang or non-dictionary word can be identified and categorized which do not have real meaning in any dictionaries but are used by many people across globe [18]. There are many words that are used by people but do not hold real meaning in any language. However, these words can also be categorized into positive and negative. This will help to attain more accurate results.

CHAPTER 9

REFERENCES

- [1] Wararat Songpan, "The analysis and prediction of Customer review rating using opinion mining", Khon Kaen University Thailand, IEEE SERA, June 2017.
- [2] Pragya Tripathi, Santosh Kr Vishwakarna, and Ajay Lala, Sentiment Analysis of English Tweets Using Rapid Miner, International conference on computational intelligence and communication networks, 2015.
- [3] <https://www.r-bloggers.com/twitter-sentiment-analysis-with-r/>
- [4] K. Shaalan and S. Atia, Increasing the accuracy of opinion mining in Arabic, Proceeding of the 1st International conference on Arabic computing linguistics, 2015, pp.106-113.
- [5] Srinidhi Bhat, Saksham Garg and Poornalatha G, "Assigning sentiment score for Twitter tweets" 2018, IEEE.
- [6] Yi Ding, Bing Li, Yuqi Zhao & Can Cheng, Scoring Tourist Attractions Based on Sentiment Lexicon, 2017, IEEE
- [7] T. Chumwatana, Using sentiment analysis technique for analyzing Thai customer satisfaction from social media, 5th International conference on computing and Informatics, 2015, pp.659-664.
- [8] S.Ahmed and A.Danti, A novel approach for sentiment analysis and opinion mining based on sentiwordnet using web data, International conference on trends in Automation, Communications and computing technology, 2015, pp. 1-5
- [9] R.K. Bakshi, R. Kaur, N. Kaur and G.Kaur, Opinion mining and sentiment analysis, 3rd International conference on computing for sustainable Global development, 2016
- [10] P. Ghaffari, J.G. Breslin, P.Barnaghim, Opinion mining and sentiment polarity on twitter and correlation between events and sentiment, 2nd international conference on Big Data Computing service and application, 2016, pp.52-57.
- [11] L. Lin, J. Li, R. Zhang, W. Yu and C. Sun, Opinion mining and sentiment analysis in Social Networks: A retweeting structure-aware approach. Proceeding of the 7th International conference on utility and cloud computing, 2014, pp.890-895.

- [12] Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model by Yakup Yildirim, Adnan Yazici., IEEE, January 2013.
- [13] Scaling Big Data Mining Infrastructure: The Twitter Experience by Jimmy Lin and Dmitriy Ryaboy (Twitter,Inc.).
- [14] Mining Large Streams of User Data for Personalized Recommendations by Xavier Amatriain (Netix).
- [15] A. Vo, Q. Nguyen, C. Ock, Opinion Aspect Relations in Cognizing Customer Feelings via Reviews, 2018
- [16] S.Bhat, S.Garg, ASSIGNING SENTIMENT SCORE FOR Twitter tweets, 2018
- [17] Y.Ding, B.Li, Y.Zhao, C.Cheng, Scoring Tourist Attractions Based on Sentiment Lexicon, 2017

PUBLICATION FROM THESIS

- [18] M.Jain, R.Jindal, Accurate Information Extraction from Customer comments posted Online, (International Journal of Recent Technology & Engineering, 2019)
IJRTE : <https://www.ijrte.org/wp-content/uploads/papers/v8i4/D7731118419.pdf>