SENTIMENT ANALYSIS OF SOCIAL MEDIA REPORTING ON DEMONETIZATION

A project report submitted in partial fulfillment of the requirements for B.Tech. Project

B.Tech.

by

Ambuj Mishra (2014IPG-012)
Sheshan Sheniwal (2014IPG-080)
Sunil Kumar (2014IPG-117)

विश्वजीवनामृतं ज्ञानम्

ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT
GWALIOR-474 010

2017

CANDIDATES DECLARATION

We therefore ensure that the work, which is being exhibited in the report, entitled **Sentiment Analysis of Social Media Reporting on Demonetization**, in incomplete satisfaction of the necessity for the honor of the Degree of **Bachelor of Technology** and submitted to the organization is our very own real record work completed amid the period *May 2017* to *September 2017* under the supervision of **Dr. Pradip Swarnakar**. We additionally refered to the reference about the text(s)/figure(s)/table(s) from where they were taken.

Date: Signatures of the Candidates

This is to ensure that the above articulation made by the candidates is right to the best of my insight.

Date: Signatures of the Research Supervisor

ABSTRACT

Demonetization was taken into action to tackle the problem of black money in India. Since then there is a lot of debate going on among many people over social media whether demonetization was good or bad. Any tweet about any event on twitter is reported by the person in his/her way of showing the event i.e., the matter in the tweet is biased towards the thinking of the users.This paper presents sentiment of society on a particular topic “demonetization”. We have collected data of demonetization of different months and weeks using twitter APIs in CSV format. Then we extracted useful information from CSV documents, like tweets, created date,created time and screen name etc using different Python libraries. Our purpose is to conduct sentiment analysis of people’s perception of demonetization by parsing the tweets with respect to time and comparing their sentiments with different algorithms. We also tried to conclude how sentiments of our society changed with respect to time.

Keywords: Sentiment analysis, demonetization, naive bayes, support vector machine, tweets.

ACKNOWLEDGEMENTS

It gives us immense pleasure in bringing out this synopsis of the report entitled "**Sentiment Analysis of Social Media Reporting on Demonetization**". We express our heartful indebtedness and owe a deep sense of gratitude to **Dr. Pradip Swarnakar** who encouraged us to work on this topic and readily agreed to supervise our work. He inspite of various engagements, showed his keen interest at every stage in completing the present study and devoted much time in conducting this study. We have been greatly benefited by his inspiring influence and able guidance. He provided us with unfailing support, guidance, simulating discussions and continuous encouragement throughout the research. This accomplishment would not have been possible without him.

Finally, we are thankful to our Institute and colleagues whose consistent encouragement served to rejuvenate our spirit, refocus our attention and enthusiasm and assisted us in carrying out this work.

(Ambuj Mishra) (Sheshan Sheniwal) (Sunil Kumar)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

RBI	Reserve Bank of India
FGOV	Foreign governmental organisation
GOV	Indian governmental organisation
IGOV	Independent governmental organisation
INGO	Independent non-governmental organisation
IRES	International research organisation
RES	Research organisation
SVM	Support Vector Machine
TH	The Hindu
TOI	Times of India

CHAPTER 1

INTRODUCTION AND LITERATURE SURVEY

In this section we briefly describe the background and related work, based on which sentiment analysis of twitter data has been performed.

1.1 INTRODUCTION

Social media is changing the way we consume news as a result of exponential increase in users day by day. Here we described our project in which sentiment analysis is performed on social media reporting on demonetization.

1.1.1 News Reporting on Social Media Platform

Dominant sources for news in our society have traditionally been broadcast and print media, especially because of their deemed reliability. With the increasing fame of internet which is surpassing that of television, a new element comes into picture (Sagan and Leighton, 2010). India has worlds highest number of users on Facebook with over 195 million users amongst which 76% are men and 24% are women. Aaj Tak is most famous Indian Brand on Facebook with above 14.4 million fans. Twitter on the other hand consists 17% of Indian social networking crowd. Prime Minister Narendra Modi is currently the most famous Indian on both the platforms having more than 33 million followers on twitter. News consumption way has been changed a lot due to social media as a result of exponential increase in users day by day. According to a study by RISJ, we got to know that more than half of the data use online platforms as a source of news across 26 countries (Fletcher et al., 2015).

Reporting means collecting the information considering the current events and background material and presenting it in such a way which makes it easier to analyze the data. Any news agency gathers news through interviews, investigation and observation by the reporter. Reporters are directed by the editors to cover a particular event, which is known as assignment. This is known as news framing (Scheufele, 1999). Digital journalism is basically broadcasting the information over any internet medium for the public bodies to read. Social networking sites are a medium to pass the news from digital news ventures to end users.

So the system is consist of mainly 3 parts:

(a) Source news venture

(b) Intermediate platform

(c) End users

Reporters are the eyes and ears of the very front in any news agency and work as a mediator between any occurring event and readers.

1.1.2 Matter in the Reporting

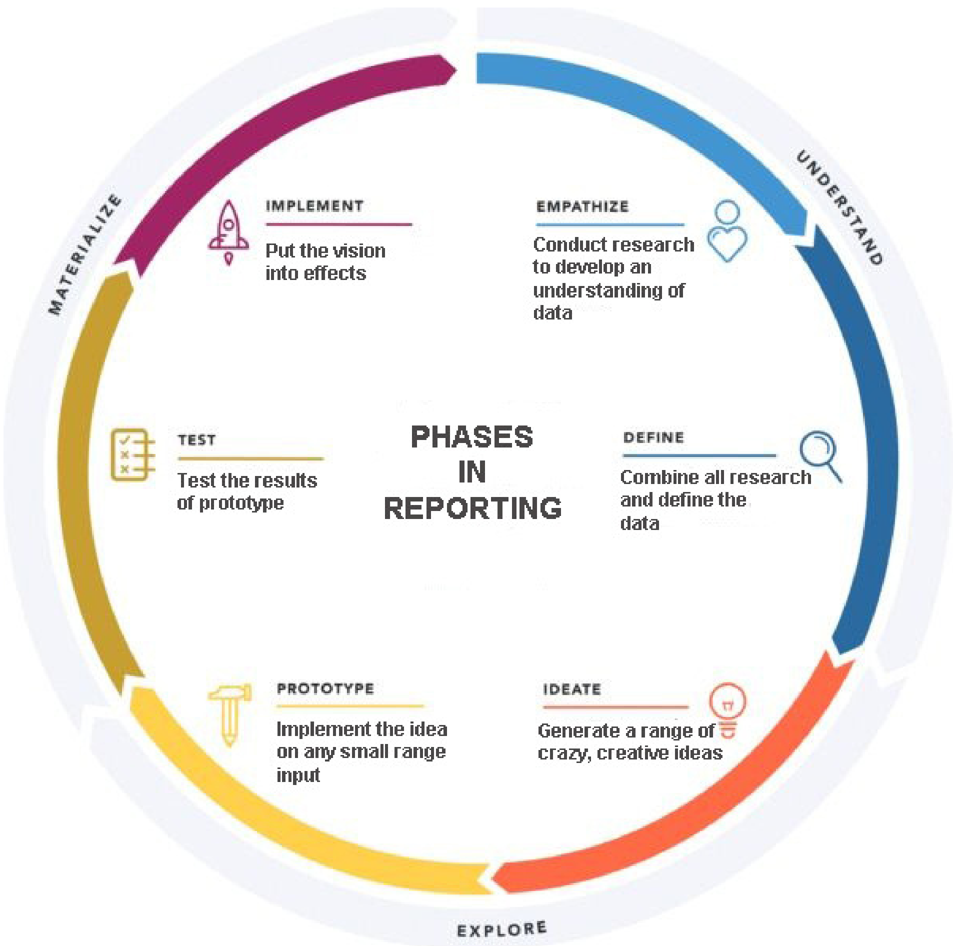

Figure 1.1: Different phases in reporting process.

Digital news reporting consists wide variety of genres such as: political reporting (which includes material on various political speeches, current political events and campaign), environment reporting (which includes the weather forecast information), religious and cultural reporting (news on different religious and cultural organizational setup), sports reporting (critical evaluation of different sports), education and research reporting (academic activity and developments) etc. But some contents cannot be altered or modified by reporters/authors during reporting such as direct speech by an individual person. We shall extract these matters (direct speech) regarding "demonetization" and shall perform sentiment analysis on them in this project and compare the conclusion using different algorithm on online media platform (such as 'twitter' in our case).

1.1.3 Why demonetization?

There is a lot of debate going on among many people over social media whether demonetization was good or bad. But looking at the data and facts, we can conclude on one common fact that GDP of the country has definitely decreased. There were many reasons to initialize demonetization in India, tackling black money is one of them. With the increasing rate of black money in India, corruption was also increasing. Black money worth more than half a million crore was supposed to scoop out (Kaur and Sharma, 2017).

Finance and Opportunity in India
20th Lalit Doshi Memorial Lecture

Response from Dr. Raghuram Rajan: I am not quite sure if what you meant is demonetise the old notes and introduce new notes instead. In the past demonetisation has been thought off as a way of getting black money out of circulation. Because people then have to come and say "how do I have this ten crores in cash sitting in my safe" and they have to explain where they got the money from. It is often cited as a solution. Unfortunately, my sense is the clever find ways around it. They find ways to divide up their hoard in to many smaller pieces. You do find that people who haven't thought of a way to convert black to white, throw it into the *Hundi* in some temples. I think there are ways around demonetization. It is not that easy to flush out the black money. Of course, a fair amount may be in the form of gold, therefore even harder to catch. I would focus more on the incentives to generate and retain black money. A lot of the incentives are on taxes. My sense is the current tax rate in this country is for the most part reasonable. We have a reasonable tax regime, for example, the maximum tax rate on high-incomes is 33%, in the US it is already 39% *plus* State taxes, etc., it takes it to near 50. We are actually lower than many industrial countries. Given that, there is no reason why everybody who should pay taxes is not paying taxes. I would focus more on tracking data and better tax administration to get at where money is not being declared. I think it is very hard in this modern economy to hide your money that easily.

Figure 1.2: Dr. Raghuram Rajans View on Demonetization.

So, for lowering the cash circulation and eliminating fake currency and dodgy funds, demonetization became necessary because many terrorist groups Parasite on fake currencies.

1.2 LITERATURE REVIEW

We are trading mainly with two literature, one of them is about the implementation of different sentiment analysis algorithms and the other one handles the sentiment analysis on twitter data.
Sentiment analysis or opinion mining is a very vast area of research which focuses at users perception by analyzing texts having users opinions. Mahmood in 2013 investigated the tweets available on twitter to predict the winner of 2013 Pakistan elec-

tions(Mahmood et al., 2013). Rapid miner was used as data mining tool by him. Author used twitter APIs to collect tweets and Naive Bayes and Support Vector Machine (SVM) algorithms for performance analysis. H. Saif, Y. He and H. Alani (Knowledge Media Institute, The Open University, United Kingdom) in a paper of November 2012 "Semantic Sentiment Analysis of Twitter" have discussed a modern approach of adding semantics as new elements into the training set for sentiment analysis(Saif et al., 2012). From tweets, they added the semantic concepts (e.g. "Apple product") as new elements corresponding to the each retrieved entity (e.g. iPhone) and evaluate the relationship of those concepts with negative/positive sentiment. Alec Go, Richa Bhayani, Lei Huang (all from Stanford University) has tried different machine learning algorithms (Naive Bayes, Maximum Entropy and SVM) for differentiating the twitter messages sentiments using distant supervision in their paper "Twitter Sentiment Classification using Distant Supervision" (Go et al., 2009). Training data was consist of twitter messages with emoticons. Authors tried to show that machine learning algorithms (Naive Bayes, Maximum Entropy and SVM) have more than 80% accuracy when data is trained with emoticons data. Best accuracy achieved is 82.2% for unigrams.

Sentiment analysis is one of most popular field these days. Everyday people are developing different algorithms and techniques to get better results. Pang and Lee (Pang et al., 2008), Liu (Liu and Zhang, 2012) , Kumar and Vadlamani Ravi (Ravi and Ravi, 2015) and Harshali P. Patil (Patil and Atique, 2015) have taken detailed surveys on opinion mining and sentiment analysis. The focus of Pang and Lee on what challenges Sentiment Analysis is facing and what could be the applications of them. Sentiment analysis can be done on anything. Movie reviews or opinion reviews is one of the major topic for sentiment analysis. ZHU Nanli and others (Nanli et al., 2012) have performed a detailed literature survey on sentiment analysis. (Zhou et al., 2012), (Hu and Li, 2011) and (He and Zhou, 2011) has done sentiment analysis on movie reviews and customer reviews. Like this (Cao et al., 2011) has done sentiment analysis on feedback of software programs. Sentiment analysis on social media data is similarly popular. Brooks (Brooks et al., 2014), Z. Jianqiang, G. Xiaolin (Jianqiang and Xiaolin, 2017), A. Rezgui,D. Fahey,I. Smith (Rezgui et al., 2016) have also worked to perform sentiment analysis on twitter data. Xiangfeng Dai have used social media data for public health surveillance (Dai et al., 2017). So we thought this can also be used on our twitter data sentiment analysis project.

We have tried to further explore the results of the study in which evaluation of sentiments on demonetization is performed by Prabhsimran Singh, Ravinder Singh Sawhney, Karanjeet Singh Kahlon in 2017 (Singh et al., 2017) and examine the results after a week by week study of the sentiments.

Table 1.1: Literature review

Sr. No.	Authors	Algorithm	Polarity	Data	Task
1	Sadeghian and Sharafat(N/A)	Random forest	Pos/Neg	Movie Reviews	Sentiment Classification
2	Pang and Lee (2008)	SVM	Pos/Neg	Movie Reviews	Sentiment Analysis
3	Cao et al. (2011)	Semantic , LSA- based	G	Software programs, users feedback	Sentiment Classification
4	Hu and Li (2011)	Graph-Based approach	Pos/Neg	Movie, Product Reviews	Sentiment Analysis
5	He and Zhou (2011)	Weakly and semi supervised classification	Pos/Neg	Movie Reviews, Multi-domain sentiment data set	Sentiment Classification
6	Tan and Wu (2011)	Random walk algorithm	G	Electronics, Stock, Hotel Reviews	Sentiment Analysis
7	Wu and Tan (2011)	Ranking algorithm	G	Book, Hotel, Notebook Reviews	Sentiment Classification
8	Xu et al. (2012)	Markov Blanket, SVM, NB, ME	Pos/Neg	Movie Reviews, News Articles	Sentiment Classification
9	Liu (2012)	Graph-Based approach, NB, SVM	Pos/Neg	Camera Reviews	Sentiment Classification
10	Hye-Jin Min, Jong C. Park(2012)	NLP	G	Blogs	Building Resource
11	Michael Hagenau, Michael liebmann, Dirk Neumann(2013)	Chi-square, BNS, SVM	G	Stock Market	Sentiment Analysis

12	Fermin L. Cruz, Jose A. Troyano, Fernando Enriquez, F. Javier Ortega, Carlos G. Vallejo(2013)	Taxonomy-based, Corpus-based	Pos/Neg	Headphones, Car, Hotel reviews	Sentiment Classification
13	Liang-chih Yu, Jheng-Long Wu, Pei-Chann Chang, Hsuan-Shou Chu(2013)	PMI-Based	G	Stock News	Sentiment Classification
14	Brooks (2014)	A n Dictionary	Pos/Neg	Twitter	Sentiment Analysis
15	Desheng Dash Wu, Lijuan Zheng, and David L. Olson (2014)	SVM	Pos/Neg	Stock forum data	Sentiment Analysis
16	Lorenzo Gatti, Marco Guerini, Marco Turchi (2015)	SWN,SVM	Pos/Neg	Movie Review	Sentiment Analysis
17	Zhao Jianqiang,Gui Xiaolin (2016)	Random Forest,Naive Bayes	Pos/Neg	Twitter Data	Sentiment Analysis
18	Kesong Liu,Jianwu Yang,Dan Zhang (2016)	SVM,Naive Bayes	Pos/Neg	Micro Blog	Sentiment Analysis

19	A. Rezgui, Daniel Fahey,Ian Smith (2016)	Naive Bayes	Pos/Neg	Twitter Data	Sentiment Analysis
20	Bradley Meyer, Marwan Bikdash, Xiangfeng Dai (2017)	SVM	Pos/Neg	Public Health Surveillance	Sentiment Analysis

1.3 MOTIVATION

1.3.1 Media Biasness on the Matter

Online media has got tremendous power to set up cultural guidelines and also to shape political discourse. It is necessary that digital media, along with other institutions, should be challenged, to be unbiased and accurate. The first step to unbiased news coverage is documenting bias. The biasness is seen in social media as low class people have not much access to it.

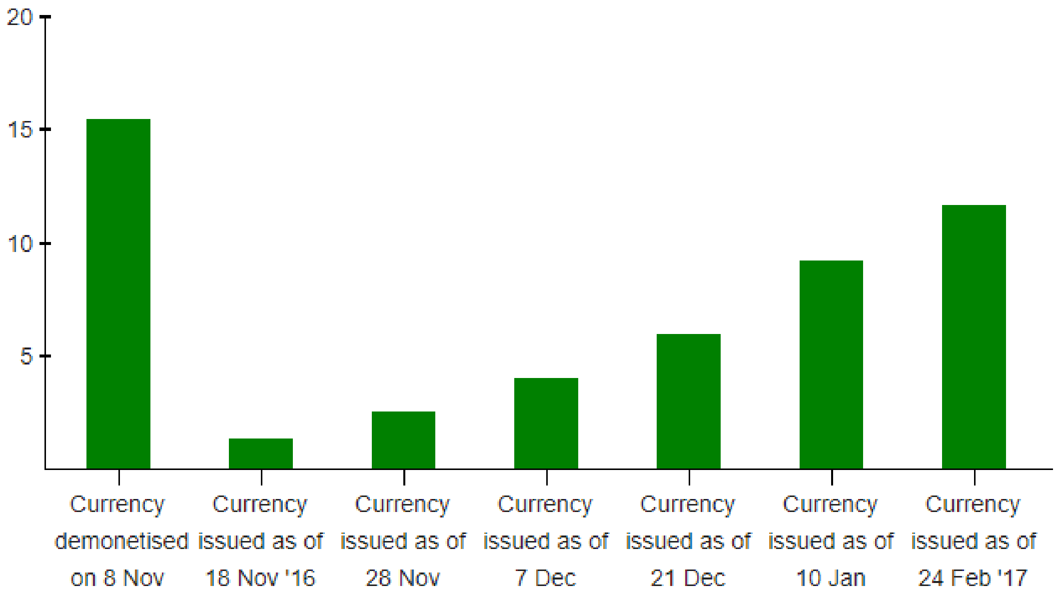

Figure 1.3: How much money is demonetized (in trillion rs.)?

Practical limitations restricting social media neutrality include the inability of reporting the news, to report all available stories and facts related to the event, and the necessity of linking selected facts into a coherent narrative and the access of resources. These days, faith in news reporting on online platforms has been degrading day by day. 77% of those surveyed by the Pew Research Center said that the online media tend to favour one side. This has surely increased if we compare it with the fact that 53% said so in 1985. But does the media really acts more biased? Or, then again is it only an instance of observation besting reality? Indeed, there is little to demonstrate that in the course of recent decades online media made news revealing more ideal to the other side. It doesn't mean however that scientists have not discovered predisposition in reporting. They have, yet it is discovered that no one side is reliably supported and neither one of the favoritism has been developing like a malevolent weed. The social media seems to be more biased as it is accessible to resourceful people.

1.3.2 Demonetization being a controversial issue in media and society

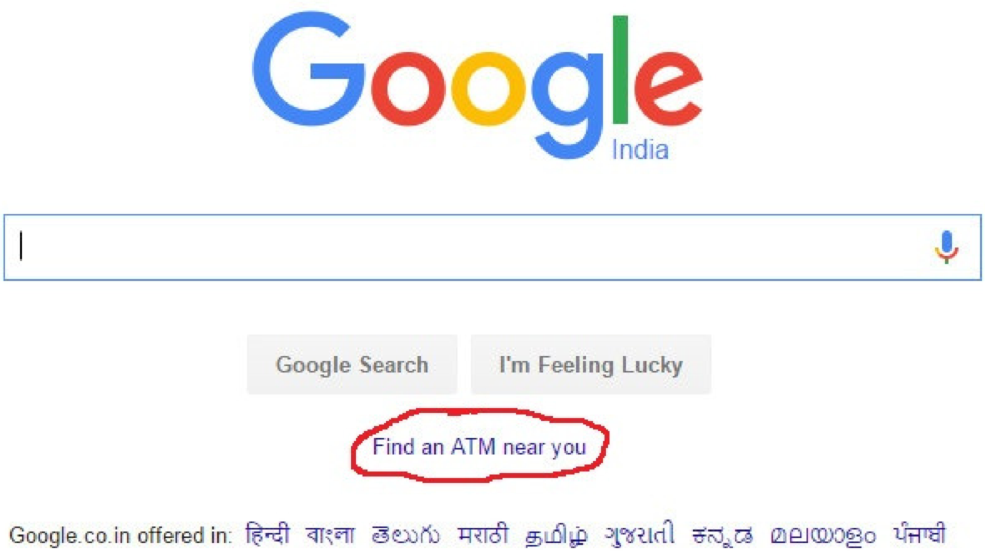

Figure 1.4: Google India homepage during demonetization

A lot of problems were faced by people during note-exchange which created a chaos around the whole country. The government claimed to curtail the shadow economy and counterattack the terrorism, human-trafficking, etc. Many parties like Communist Party of India (Marxist) (CPM) claimed that the BJP leaders of Bengal had pre-knowledge of demonetization.
Many people had been counted to die during demonetization while standing in queue to exchange their notes.
There were so many debates to justify the move made by the BJP government. Some

people have to say that this action has been taken for the benefit of Indian people. Many data shows some decrement in human-trafficking but not much as thought of. Others have justification of failure of demonetization as about 90% of old notes have been exchanged and there is nothing much for the country on the misery of common men.

1.4 GAP ANALYSIS

In this venture, a few troubles were confronted amid sparing such expansive information of tweets. Amassing and searching for topics has turned out to be imperative in industry, culture, sciences and humanities, which is not a simple errand as information is extremely gigantic and complex. So we use python libraries to deal with this type of problem.
Thousands of people tweeted on demonetization. It is necessary to analyze sentiment of tweets to check biasness.
If some event occurs at some place, then the tweets we get related to it could be biased. Whatever twitteraties tells influenced by their own point of view because tweets are influenced by personal views. If one person is following some ideology then each tweet by that person will be influenced by his/her ideology. If any negative thing happens in any place then the people of that locality tweets negatively about the action. So we analyze sentiment of these articles to gain more accuracy.

1.5 AIM

To conduct sentiment analysis of people perception of demonetization by parsing through the tweets and comparing their sentiments with different algorithms over time.

1.6 OBJECTIVES AND RESEARCH QUESTIONS

1.6.1 Objectives

- To examine people's perception of Demonetization by parsing the tweets with respect to time(8 weeks)
- To compare public sentiments through outcomes of naive bayes and support vector machine algorithms

1.6.2 Research Questions

- Which part of our community adores demonetization?
- What is the use of different algorithms?
- How print media and social media reflect our society differently?

CHAPTER 2

METHODOLOGY AND IMPLEMENTATION

Online tweets has been extracted with the help of python libraries like tweepy , twitter API,etc. Tweepy enable us to scrap online tweets on the topic "Demonetization".

2.1 DATASET

2.1.1 Twitter data

Twitter is one of the most frequently used social site. We extract useful twitter data and convert the data in .csv format. There are many parameters like id , date , sender , no. of retweets , etc are present in our dataset. The tweeter data contains the tweets during demonetization,and after that. The dataset is stored according to the date of corresponding tweets.

	A	B	C	D	E	F	G	H	I	J	K	L	
1		X	text	favoritec	favoriteC	replyToSl	created	truncatec	replyToSl	id	replyToU	statusSource	screenN
2	1	1	RT @rssurjewala: Critical question: Was PayTM informed about #Demonetization edict by	FALSE	0	NA	11/23/2016 18:40	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	HASHTA
3	2	2	RT @Hemant_80: Did you vote on #Demonetization on Modi survey app?	FALSE	0	NA	11/23/2016 18:40	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	PRAMOI
4	3	3	RT @roshankar: Former FinSec, RBI Dy Governor, CBDT Chair + Harvard Professor	FALSE	0	NA	11/23/2016 18:40	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	rahulja:
5	4	4	RT @ANI_news: Gurugram (Haryana): Post office employees provide cash exchange to pat	FALSE	0	NA	11/23/2016 18:39	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	deeptiy
6	5	5	RT @satishacharya: Reddy Wedding! @mail_today cartoon #demonetization #ReddyWed	FALSE	0	NA	11/23/2016 18:39	FALSE	NA	8.01E+17	NA	<a href="http://cpimharyana.c	CPIMBa
7	6	6	@DerekScissors1: India's #demonetization: #Blackmoney a symptom, not the disease htt	FALSE	0	DerekSci	11/23/2016 18:39	FALSE	NA	8.01E+17	2.6E+09	<a href="http://twitter.com" re	ambaza
8	7	7	RT @gauravcsawant: Rs 40 lakh looted from a bank in Kishtwar in J&K. Third such inc	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	bhodia:
9	8	8	RT @Joydeep_911: Calling all Nationalists to join...	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	KARUNA
10	9	9	RT @sumitbhati2002: Many opposition leaders are with @narendramodi on the	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	sumitbl
11	10	10	National reform now destroyed even the essence of sagan. Such instances urge giving #c	FALSE	0	NA	11/23/2016 18:38	TRUE	NA	8.01E+17	NA	<a href="https://mobile.twitte	HelpInc
12	11	11	Many opposition leaders are with @narendramodi on the #Demonetization	FALSE	1	NA	11/23/2016 18:37	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	sumitbl
13	12	12	RT @Joydas: Question in Narendra Modi App where PM is taking feedback if people supp	FALSE	0	NA	11/23/2016 18:37	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	Monish
14	13	13	@Jaggesh2 Bharat band on 28??<ed><U+00A0><U+00BD><ed><U+00B8><U+0082>Those wh	FALSE	0	Jaggesh2	11/23/2016 18:37	FALSE	8.01E+17	8.01E+17	1.2E+09	<a href="http://twitter.com/dc	yuvaraj_
15	14	14	RT @Atheist_Krishna: The effect of #Demonetization !!	FALSE	0	NA	11/23/2016 18:36	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	PMKejri
16	15	15	RT @sona2905: When I explained #Demonetization to myself and tried to put it down in	FALSE	0	NA	11/23/2016 18:36	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	hkgupta
17	16	16	RT @Dipankar_cpiml: The Modi app on #DeMonetization proves once again that the govt	FALSE	0	NA	11/23/2016 18:35	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	aazaad
18	17	17	RT @roshankar: Former FinSec, RBI Dy Governor, CBDT Chair + Harvard Professor	FALSE	0	NA	11/23/2016 18:35	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	darkdes
19	18	18	RT @Atheist_Krishna: BEFORE and AFTER Gandhi ji heard they are standing there	FALSE	0	NA	11/23/2016 18:34	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	snoovei
20	19	19	RT @pGurus1: #Demonetization The co-operative banking sector in Kerala is as good as	FALSE	0	NA	11/23/2016 18:34	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	Vishwa
21	20	20	RT @roshankar: Former FinSec, RBI Dy Governor, CBDT Chair + Harvard Professor	FALSE	0	NA	11/23/2016 18:34	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com" re	Politica
22	21	21	RT @Hemant_80: Did you vote on #Demonetization on Modi survey app?	FALSE	0	NA	11/23/2016 18:34	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	MdShua
23	22	22	RT @roshankar: Former FinSec, RBI Dy Governor, CBDT Chair + Harvard Professor	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	BharatP
24	23	23	RT @Atheist_Krishna: BEFORE and AFTER Gandhi ji heard they are standing there	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	ihavnth
25	24	24	RT @MahikaInfra: @narendramodi	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	GajjarBI
26	25	25	RT @Hemant_80: Did you vote on #Demonetization on Modi survey app?	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com" re	Rss_cha
27	26	26	RT @roshankar: Former FinSec, RBI Dy Governor, CBDT Chair + Harvard Professor	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	mukesh
28	27	27	RT @kapil_kausik: #Doltiwal I mean #JaiChandKejriwal is "hurt" by #Demonetization as t	FALSE	0	NA	11/23/2016 18:32	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	mrx565
29	28	28	RT @roshankar: Former FinSec, RBI Dy Governor, CBDT Chair + Harvard Professor	FALSE	0	NA	11/23/2016 18:32	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com" re	iMirzaG
30	29	29	RT @kapil_kausik: #Doltiwal I mean #JaiChandKejriwal is "hurt" by #Demonetization as t	FALSE	0	NA	11/23/2016 18:32	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dc	PontiFu
31	30	30	RT @AAPVind: #Demonetization Is Disaster! @naam_pk	FALSE	0	NA	11/23/2016 18:31	FALSE	NA	8.01E+17	NA	<a href="http://www.twitter.cc	sehgal_
32	31	31	RT @Hemant_80: Did you vote on #Demonetization on Modi survey app?	FALSE	0	NA	11/23/2016 18:31	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com" re	BoscoU

Figure 2.1: Dataset.

2.1.2 Training dataset

Kaggle is a platform for predictive modeling and analytical competition on which firm post their informations . It is an open source website . We have used training data from kaggle for the sentiment analysis on the topic "Demonetization".

2.2 WORK MODEL

After looking at recent literature , we have proposed a methodology to fulfill our objectives . The work model and the flow of work can be described by the flow chart present in figure.

2.2.1 Steps

First of all , we scraped data from twitter using Twitter API , Python libraries like tweepy.
Matter in the scraped tweets being unorganized and mixed were tough to directly apply sentiment analysis on them. Also we would not have got more accurate results .so,we extract useful data from scraped tweet like date , id , screenname , no. of retweets etc .
All the data were organised in .csv format . so, data can be made readable to algorithm.
Sentiment analysis from different methods (1. Naive Bayes , 2. SVM) were performed on the extracted tweets .
Graphs were plotted for variation of sentiments with time .
Graphs were analyzed to compare results of their sentiment values.

2.3 FLOW CHART

Below shown is the flowchart for the sentiment analysis on any twitter data -

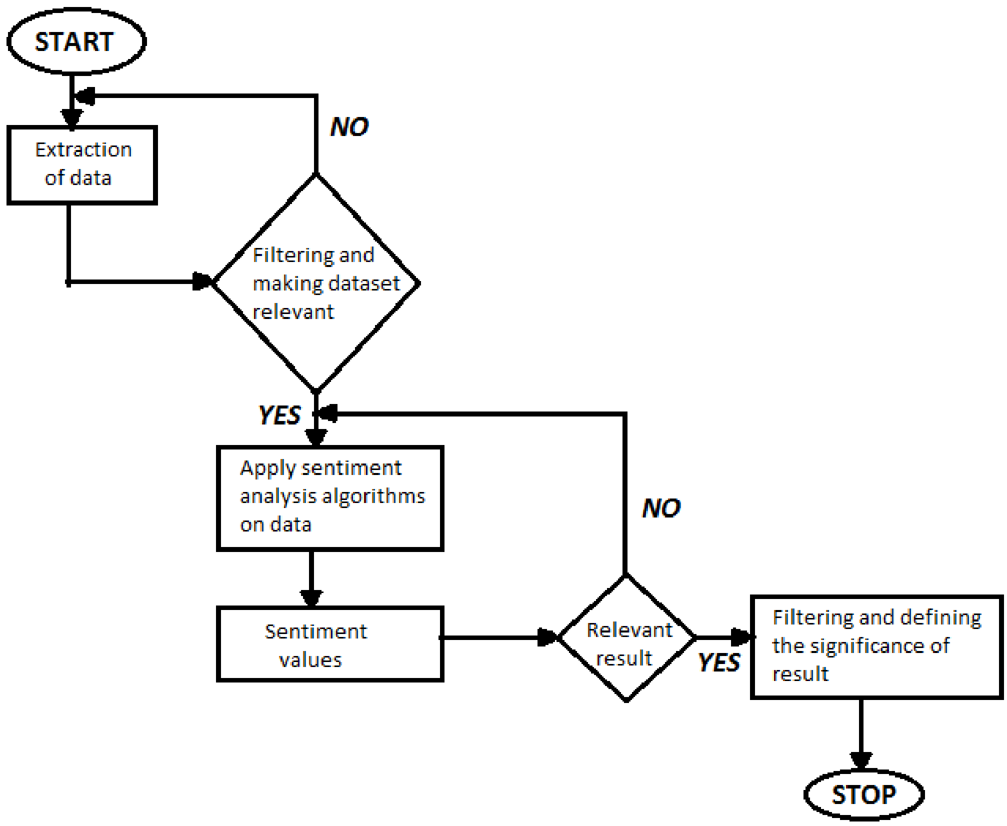

Figure 2.2: Flowchart for sentiment analysis on twitter data.

2.4 METHODOLOGY AND IMPLEMENTATION

We have data , which is unorganized. So we need to extract relevant data from this unorganized data for further analysis. We need to extract id, date, text.So we used pyth and nltk libraries for the extraction of relevant data. We have used two different algorithms for finding the sentiment values.
1.Naive Bayes
2.Support Vector Machine(SVM)

2.4.1 Naive Bayes

Naive Bayes is a classifier based on Bayes' theorem. Classification technique in Naive Bayes assumes that the two features with in a class are independent(unrelated) to each other.For example, a ball may be considered to be a cricket ball if it is round in shape, red in colour,5 inches in diameter and hard in nature.The contribution of each of these

features shape, colour, size and hardness to the probability of a ball being cricket ball is considered independent by a naive Bayes classifier, regardless of any possibility of dependency between these features shape, size and colour.
For different models of probability, people uses NBC(Naive Bayes Classifier), as it is efficient and easy to train Naive Bayes Classifier in supervised learning.
NBC uses the following model of probability for each of k possible classes

p(Ck/x1,x2,....,xn)

for bigger estimation of n above formula doesn't function admirably So,Bayes theorem deteriorates the conditional likelihood as

p(Ck|x) =p(Ck) p(x|Ck)/p(x))

where Ck is class for each of k possible outcomes and x are the instances to be classified of a given problem represented as a vector.
X = (x1,x2,x3,.....,xn) representing n independent variables or features.
P(c|x) denotes the posterior probability of class.
P(c) denotes the prior probability of class.
P(x|c) denotes the likelihood which is the probability of predictor given class.
P(x) denotes the prior probability of predictor.
NBC mostly used for classification of texts(because of better outcome in multiple class issues and independent rule) have more favourable rate when contrasted with different methods.Thus, it is generally utilized as a part of Spam sifting (recognize spam email) and Sentiment Analysis (in online networking examination, to distinguish positive and negative assumptions).

2.4.2 Support Vector Machine(SVM)

Support Vector Machine is a non-probabilistic binary linear classifier which has the ability to linearly separate the classes by a large margin. Add to it the Kernel, and SVM becomes one of the most powerful classifier capable of handling infinite dimensional feature vectors.
Support Vector Machine is an administered(supervised) ML algorithm which can be utilized for both classification or regression challenges. In any case, it is for the most part utilized as a part of classification problems.
In this algorithm, each data point is plotted as a point in n-dimensional space (where number of features is represented by n) with the estimation of each element being the estimation of a specific facilitate or particular coordinates. At that point, we perform classification by finding the hyper-plane that separate the two classes exceptionally well.
Hyper-plane is selected with following consideration: That plane is selected which separates the two classes better. That plane is selected which is having maximum distance

from nearest data point.And this maximum distance is called margin.
That plane is selected which is having less classification error.

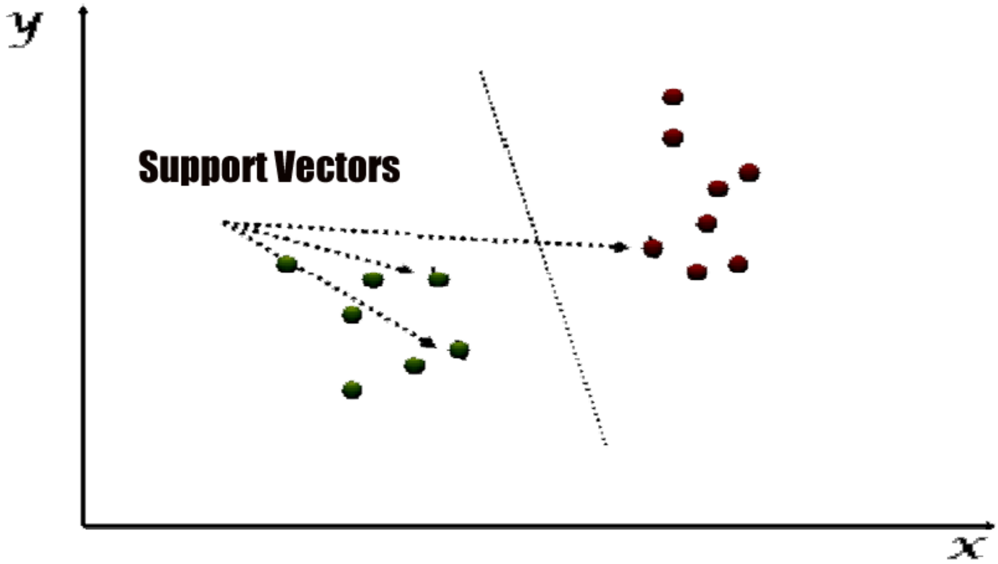

Figure 2.3: SVM

In the problem of high dimensional spaces,Support Vector Machine is extremely powerful ,i.e., in situations where number of samples is less than number of dimensions.Support Vector Machine utilizes a subset of training points in the decisive function(called support vectors) ,so it is likewise memory proficient.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Collective data with sentiment values from different methods

After performing sentiment analysis of all tweets, from the two mentioned methods, we organized the collective data (Serial Number, Date, Time and their sentiment scores) with their corresponding sentiment values with different methods in the data-base (excel sheet). The first methods (SVM) for sentiment analysis gave results in sentiments values of 1, -1 and 0 as Positive, Negative and Neutral respectively.
We, from the second method (Sentiment analysis using Naive Bayes) obtained the fractional sentiment value of the tweets. The values from this method were further used in analyzing and observation through graphs of 'sentiments values vs time' for weeks.

3.2 Variation of sentiment values of tweets over time

We have created tables showing variation in sentiments of tweets with time in accordance of every week with respect to different methods.

Table 3.1: Weekly distribution of dates

Week	From	To
W 1	2016/11/09	2016/11/14
W 2	2016/11/15	2016/11/20
W 3	2016/11/21	2016/11/27
W 4	2016/11/28	2016/12/04
W 5	2016/12/05	2016/12/10
W 6	2016/12/11	2016/12/17
W 7	2016/12/18	2016/12/24
W 8	2016/12/25	2016/12/31

Table 3.2: Weekly distribution of different sentiment values over different algorithms

Week No.	Positive Naive Bayes (in %)	Positive SVM (in %)	Negative Naive Bayes (in %)	Negative SVM (in %)	Neutral Naive Bayes (in %)	Neutral SVM (in %)
W 1	32.90	27.00	15.00	15.00	52.10	58.00
W 2	28.50	23.00	9.80	9.80	61.70	67.20
W 3	33.40	35.60	13.40	12.30	53.20	52.10
W 4	43.70	46.20	16.80	13.30	39.50	40.50
W 5	34.10	77.10	16.30	16.90	49.60	06.00
W 6	46.00	44.20	14.80	13.30	39.20	42.50
W 7	61.20	63.10	11.10	08.50	27.70	28.40
W 8	61.30	88.40	06.17	09.40	32.53	02.20

3.2.1 Variation of sentiment values per week

For comparing the sentiment values from both the methods (Naive Bayes and SVM), we have drawn different graphs for different weeks.

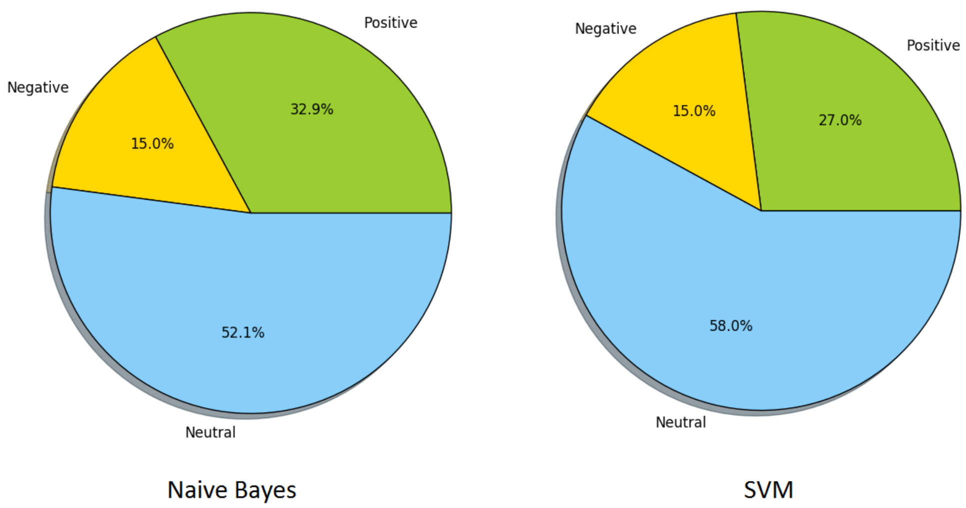

Figure 3.1: 1st week of demonetization

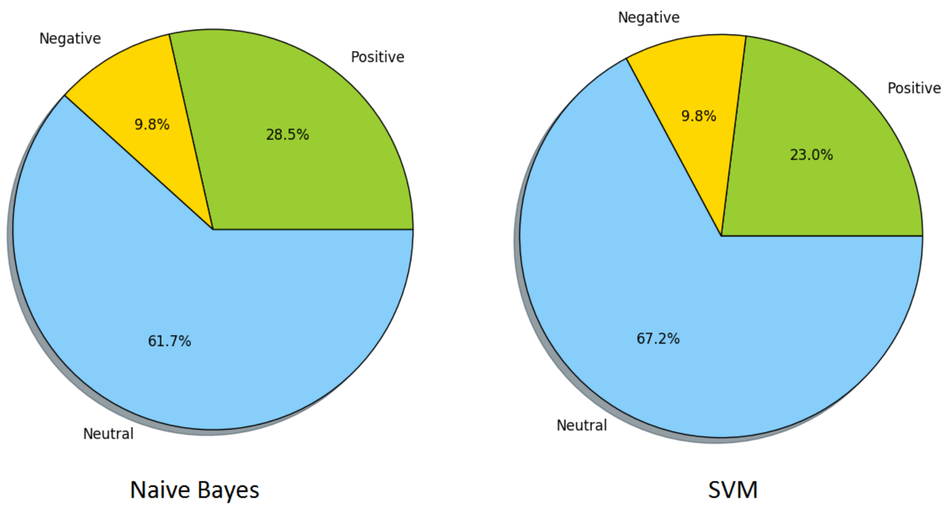

Figure 3.2: 2nd week of demonetization

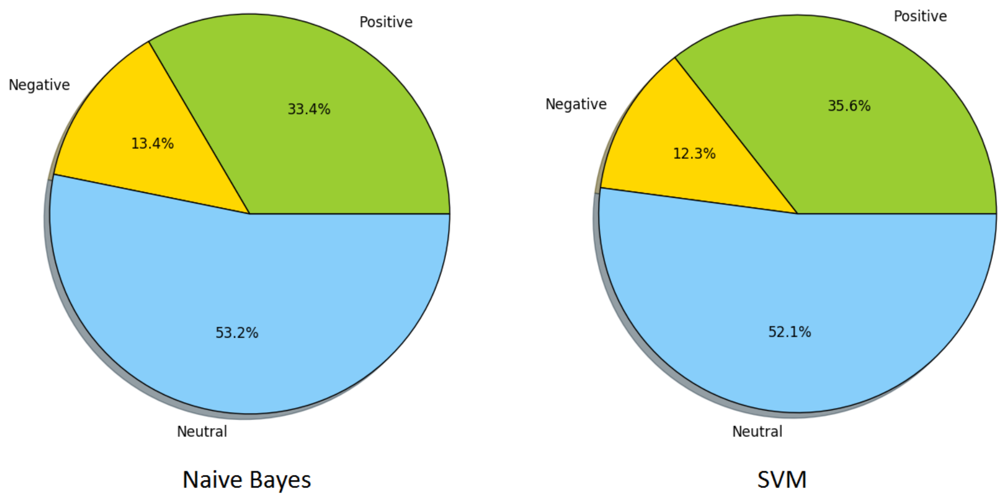

Figure 3.3: 3rd week of demonetization

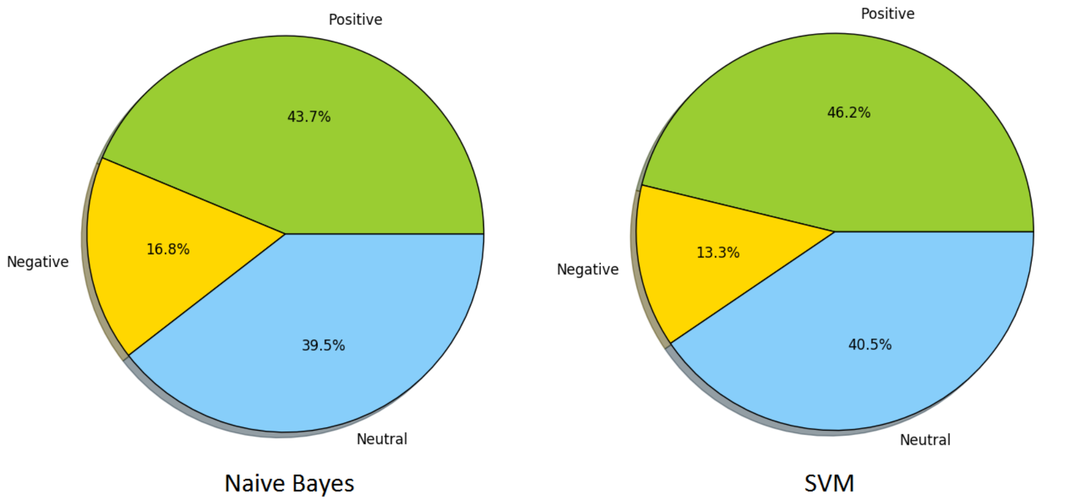

Figure 3.4: 4th week of demonetization

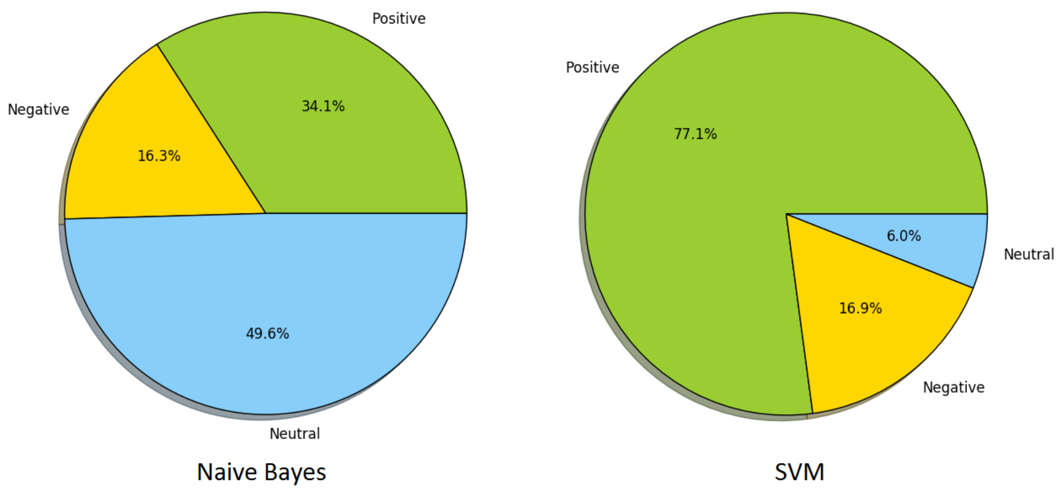

Figure 3.5: 5th week of demonetization

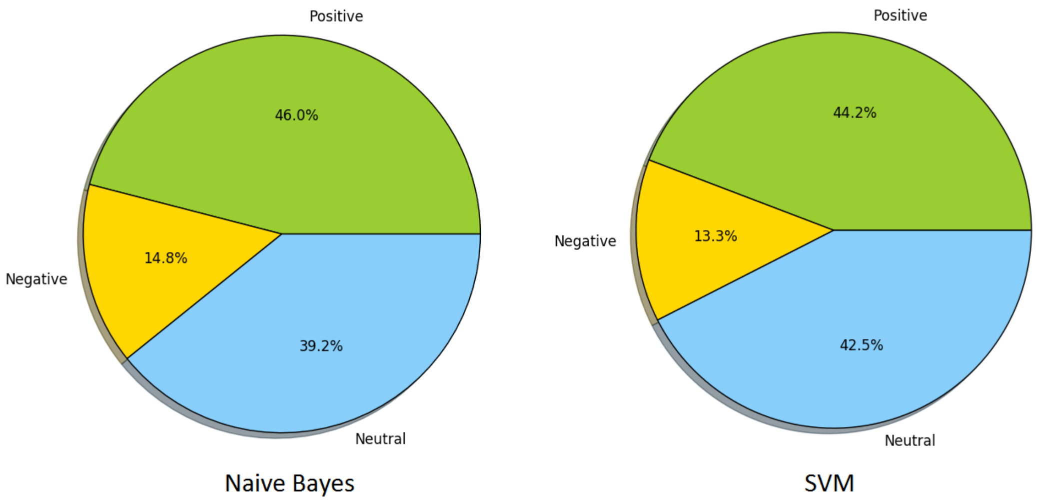

Figure 3.6: 6th week of demonetization

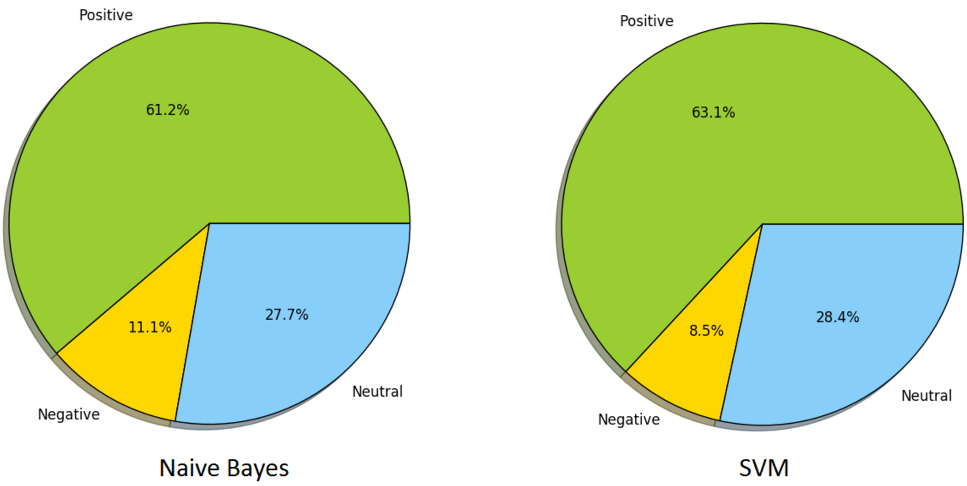

Figure 3.7: 7th week of demonetization

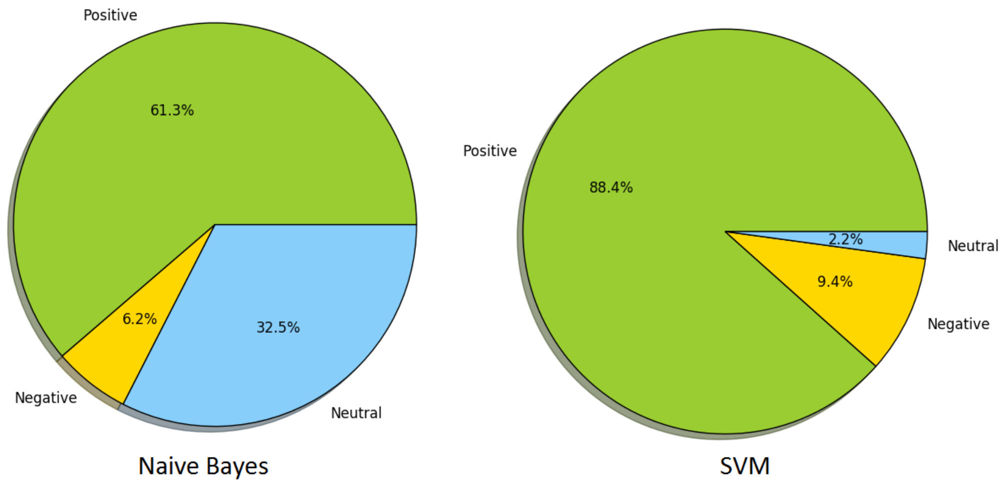

Figure 3.8: 8th week of demonetization

3.2.2 Overall variation of sentiments

We have created graphs for showing sentiments of people throughout the period of demonetization with two different algorithms.

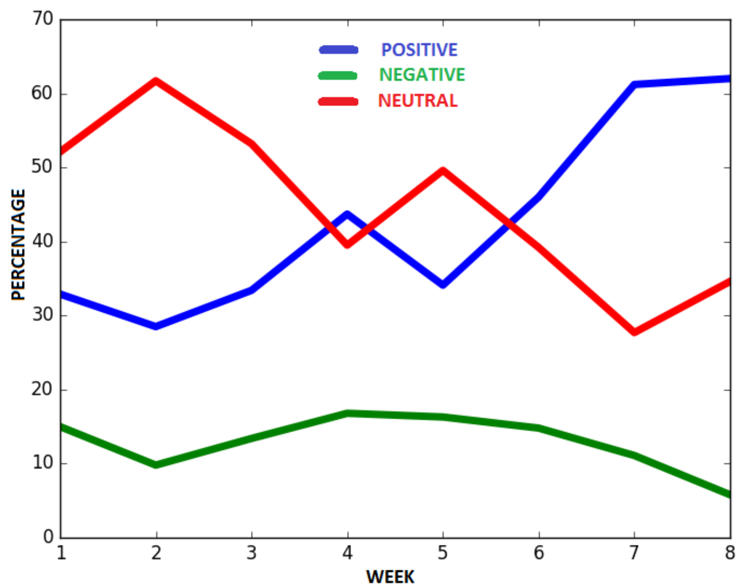

Figure 3.9: Variation of all sentiments through naive bayes over time

Through graphs we can see how the negative sentiments have fallen down and positive sentiments have risen.

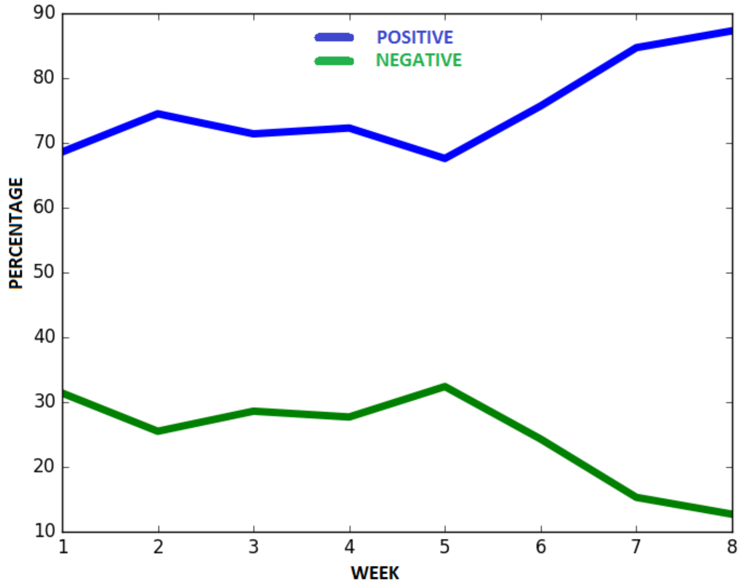

Figure 3.10: Comparison of positive and negative sentiments through naive bayes over time

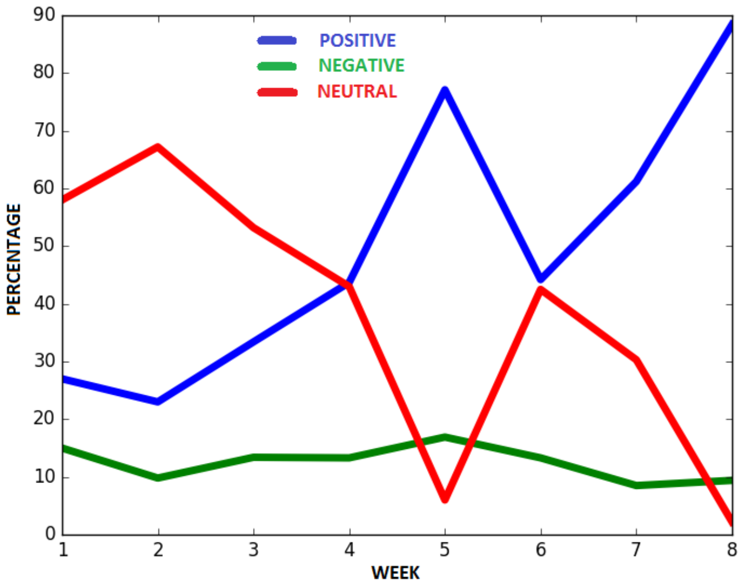

Figure 3.11: Variation of all sentiments through SVM over time

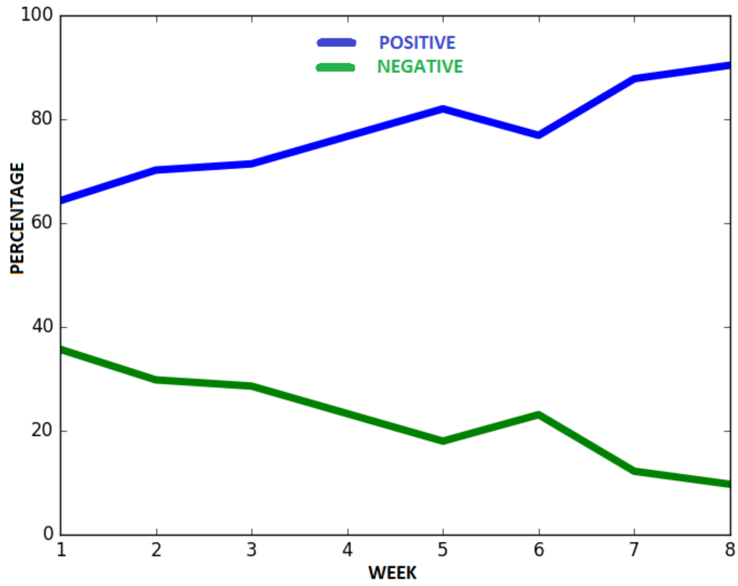

Figure 3.12: Comparison of positive and negative sentiments through SVM over time

3.3 Result

Table 3.3: Overall distribution of different sentiment values over different algorithms

Sentiment	Naive Bayes (in %)	SVM (in %)
Positive	39.60	33.90
Negative	14.80	10.30
Neutral	45.6	55.80

Graph shows overall sentiments values of people during demonetization with respect to two different algorithms.

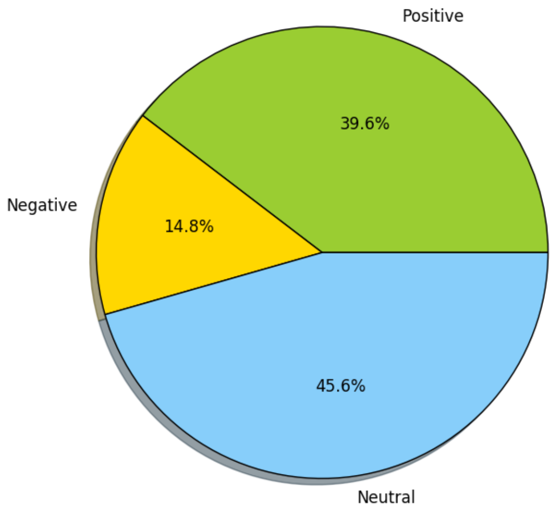

Figure 3.13: Overall sentiments through Naive Bayes

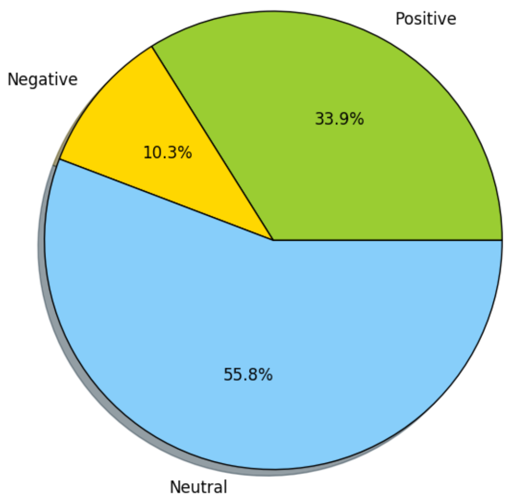

Figure 3.14: Overall sentiments through SVM

We can see overall positive attitude of public for demonetization over time.

CHAPTER 4

CONCLUSION

The biasness in the information makes the analysis of the situation more complex. In such situations, analysis of data and facts proves to be very crucial to analyze the actual perception of society about the topic of interest. It can further open new ways for better solutions and optimal policies.
In our study, we applied different machine learning algorithms to differentiate the sentiments of our society on demonetization with respect to time. The obtained results are compared with each other. The study shows that with every passing week, negativity in the sentiments started to decrease and the number of positive responses increased.
Specially, the conclusive idea about the thesis is as follows:

(a) Twitter includes a great proportion of our population but not everyone. Only upper and middle class of our society has access to social media. So, their opinion mattered the most.

(b) Money exchange was the main issue with demonetization which was mostly faced by lower and working class people due to lack of "digitalization" among them. But our analysis suggests that majority of users believe that demonetization is beneficial for our country. Hence, it concludes that demonetization was not fair for everyone but appreciated by upper class people.

(c) Online social media platforms have widely distributed proportion of our society. Our thesis shows that most of the users reacted positively towards demonetization process as well which makes it a success.

The results of our thesis are based on short term opinions but we cannot predict how much beneficial it will be in future.

REFERENCES

[1] Brooks, M., Robinson, J. J., Torkildson, M. K., Aragon, C. R. et al.: 2014, Collaborative visual analysis of sentiment in twitter events, *International Conference on Cooperative Design, Visualization and Engineering*, Springer, pp. 1–8.

[2] Cao, Q., Duan, W. and Gan, Q.: 2011, Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach, *Decision Support Systems* **50**(2), 511–521.

[3] Dai, X., Bikdash, M. and Meyer, B.: 2017, From social media to public health surveillance: Word embedding based clustering method for twitter classification, *SoutheastCon, 2017*, IEEE, pp. 1–7.

[4] Fletcher, R., Radcliffe, D., Levy, D. A., Nielsen, R. K. and Newman, N.: 2015, Reuters institute digital news report 2015: Supplementary report.

[5] Go, A., Bhayani, R. and Huang, L.: 2009, Twitter sentiment classification using distant supervision, *CS224N Project Report, Stanford* **1**(2009), 12.

[6] He, Y. and Zhou, D.: 2011, Self-training from labeled features for sentiment analysis, *Information Processing & Management* **47**(4), 606–616.

[7] Hu, Y. and Li, X.: 2011, Context-dependent product evaluations: an empirical analysis of internet book reviews, *Journal of Interactive Marketing* **25**(3), 123–133.

[8] Jianqiang, Z. and Xiaolin, G.: 2017, Comparison research on text pre-processing methods on twitter sentiment analysis, *IEEE Access* **5**, 2870–2879.

[9] Kaur, K. and Sharma, N.: 2017, Demonetization in india: Looking through the mirror of twitter sentiments.

[10] Liu, B. and Zhang, L.: 2012, A survey of opinion mining and sentiment analysis, *Mining text data*, Springer, pp. 415–463.

[11] Mahmood, T., Iqbal, T., Amin, F., Lohanna, W. and Mustafa, A.: 2013, Mining twitter big data to predict 2013 pakistan election winner, *Multi Topic Conference (INMIC), 2013 16th International*, IEEE, pp. 49–54.

[12] Nanli, Z., Ping, Z., Weiguo, L. and Meng, C.: 2012, Sentiment analysis: A literature review, *Management of Technology (ISMOT), 2012 International Symposium on*, IEEE, pp. 572–576.

[13] Pang, B., Lee, L. et al.: 2008, Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval* **2**(1–2), 1–135.

[14] Patil, H. P. and Atique, M.: 2015, Sentiment analysis for social media: a survey, *Information Science and Security (ICISS), 2015 2nd International Conference on*, IEEE, pp. 1–4.

[15] Ravi, K. and Ravi, V.: 2015, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, *Knowledge-Based Systems* **89**, 14–46.

[16] Rezgui, A., Fahey, D. and Smith, I.: 2016, Affinityfinder: A system for deriving hidden affinity relationships on twitter utilizing sentiment analysis, *Future Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference on*, IEEE, pp. 212–215.

[17] Sagan, P. and Leighton, T.: 2010, The internet & the future of news, *Daedalus* **139**(2), 119–125.

[18] Saif, H., He, Y. and Alani, H.: 2012, Semantic sentiment analysis of twitter, *The Semantic Web–ISWC 2012* pp. 508–524.

[19] Scheufele, D. A.: 1999, Framing as a theory of media effects, *Journal of communication* **49**(1), 103–122.

[20] Singh, P., Sawhney, R. S. and Kahlon, K. S.: 2017, Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by indian government, *ICT Express* .

[21] Zhou, X., Xu, Y., Li, Y., Josang, A. and Cox, C.: 2012, The state-of-the-art in personalized recommender systems for social networking, *Artificial Intelligence Review* **37**(2), 119–132.