

# BTP 2017

*by* Ambuj Mishra

---

**Submission date:** 18-Sep-2017 04:07PM (UTC+0530)

**Submission ID:** 848545205

**File name:** Final\_Report.pdf (1.37M)

**Word count:** 5602

**Character count:** 29615

# **SENTIMENT ANALYSIS OF SOCIAL MEDIA REPORTING ON DEMONETIZATION**

1

*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)**



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

**2017**

## CANDIDATES DECLARATION

We hereby certify that the work, which is being presented in the report, entitled **Sentiment Analysis of Social Media Reporting on Demonetization**, in partial fulfillment of the requirement for the award of the Degree of **Bachelor of Technology** and submitted to the institution is an authentic record of our own work carried out during the period **May 2017 to September 2017** under the supervision of **Dr. Pradip Swarnakar**. We also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

Date:

Signatures of the Candidates

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date:

Signatures of the Research Supervisors

## 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 Rtf format. Then we extracted useful information from Rtf documents, like tweets, created date, created time and screen name etc using different Python libraries and Natural language toolkit. Our purpose is to conduct sentiment analysis of people perception of demonetization by parsing through the tweets with respect to time and comparing their sentiments with different algorithms. We also try to conclude how sentiments of our society changed with respect to time.

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

## ACKNOWLEDGEMENTS

We are highly indebted to **Dr. Pradip Swarnakar**, and are obliged for giving us the autonomy of functioning and experimenting with ideas. We would like to take this opportunity to express our profound gratitude to them not only for their academic guidance but also for their personal interest in our project and constant support coupled with confidence boosting and motivating sessions which proved very fruitful and were instrumental in infusing self-assurance and trust within us. The nurturing and blossoming of the present work is mainly due to their valuable guidance, suggestions, astute judgment, constructive criticism and an eye for perfection. Our mentor always answered myriad of our doubts with smiling graciousness and prodigious patience, never letting us feel that we are novices by always lending an ear to our views, appreciating and improving them and by giving us a free hand in our project. It's only because of their overwhelming interest and helpful attitude, the present work has attained the stage it has.

Finally, we are grateful to our Institution and colleagues whose constant encouragement served to renew our spirit, refocus our attention and energy and helped us in carrying out this work.

(Ambuj Mishra)

(Sheshan Sheniwal)

(Sunil Kumar)

# TABLE OF CONTENTS

<b>ABSTRACT</b>	<b>ii</b>
<b>LIST OF TABLES</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>vi</b>
<b>1 INTRODUCTION AND LITERATURE SURVEY</b>	<b>ix</b>
1.1 INTRODUCTION . . . . .	ix
1.1.1 News Reporting on Social Media Platform . . . . .	ix
1.1.2 Matter in the Reporting . . . . .	x
1.1.3 Why direct speech? (Brunner,2013) . . . . .	xi
1.1.4 Why demonetization? . . . . .	xi
1.2 LITERATURE REVIEW . . . . .	xii
1.3 MOTIVATION . . . . .	xvi
1.3.1 Media Biasness on the Matter . . . . .	xvi
1.3.2 Demonetization issue . . . . .	xvi
1.4 GAP ANALYSIS . . . . .	xvii
1.5 AIM . . . . .	xviii
1.6 OBJECTIVES and RESEARCH QUESTIONS . . . . .	xviii
1.6.1 Objectives . . . . .	xviii
1.6.2 Research Questions . . . . .	xviii
<b>2 METHODOLOGY AND IMPLEMENTATION</b>	<b>xix</b>
2.1 DATA SET . . . . .	xix
2.1.1 Twitter data . . . . .	xix
2.1.2 Training data set . . . . .	xx
2.2 WORK MODEL . . . . .	xx
2.2.1 Steps . . . . .	xx
2.3 FLOW CHART . . . . .	xxi
2.4 METHODOLOGY AND IMPLEMENTATION . . . . .	xxi
2.4.1 Naive Bayes . . . . .	xxi

**TABLE OF CONTENTS**

v

2.4.2 SVM . . . . .	xxii
<b>3 RESULTS AND DISCUSSION</b>	<b>xxiv</b>
3.1 Collective data with sentiment values from different methods . . . . .	xxiv
3.2 Variation of sentiment values of tweets over time . . . . .	xxiv
3.2.1 Variation of sentiment values per week . . . . .	xxv
3.2.2 Overall variation of sentiments . . . . .	xxviii
3.3 Result . . . . .	xxx
<b>4 CONCLUSION</b>	<b>xxxii</b>
<b>REFERENCES</b>	<b>xxxii</b>

## **LIST OF TABLES**

1.1	Literature review . . . . .	xiv
3.1	Weekly distribution of different sentiment values over different algorithms	xxv
3.2	Overall distribution of different sentiment values over different algorithms	xxx

# LIST OF FIGURES

1.1	Different phases in reporting process. . . . .	x
1.2	Dr. Raghuram Rajan's View on Demonetization. . . . .	xii
1.3	How much money is demonetized (in trillion rs.)? . . . . .	xvii
2.1	Dataset. . . . .	xix
2.2	Flowchart for sentiment analysis on twitter data. . . . .	xxi
2.3	SVM . . . . .	xxiii
3.1	Week 1 of demonetization . . . . .	xxv
3.2	Week 2 of demonetization . . . . .	xxvi
3.3	Week 3 of demonetization . . . . .	xxvi
3.4	Week 4 of demonetization . . . . .	xxvi
3.5	Week 5 of demonetization . . . . .	xxvii
3.6	Week 6 of demonetization . . . . .	xxvii
3.7	Week 7 of demonetization . . . . .	xxvii
3.8	Week 8 of demonetization . . . . .	xxviii
3.9	Variation of all sentiments through naive bayes over time . . . . .	xxviii
3.10	Comparison of positive and negative sentiments through naive bayes over time . . . . .	xxix
3.11	Variation of all sentiments through SVM over time . . . . .	xxix
3.12	Comparison of positive and negative sentiments through SVM over time . . . . .	xxx
3.13	Overall sentiments through Naive Bayes . . . . .	xxx
3.14	Overall sentiments through SVM . . . . .	xxx

## **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**

This chapter includes the details of social media reporting, our objectives, platform used to implement the project and literatures review related to work done in this field.

### **1.1 INTRODUCTION**

In this section we briefly describe 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 Leighto, 2010:119). 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. Social media is changing the way we consume news as a result of exponential increase in users day by day. In a study by the Reuters Institute for the Study of Journalism (RISJ), we got to know that 51% of the sample use social media as a news source across 26 countries (Newman, Fletcher Levy Nielsen, 2016:8).

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

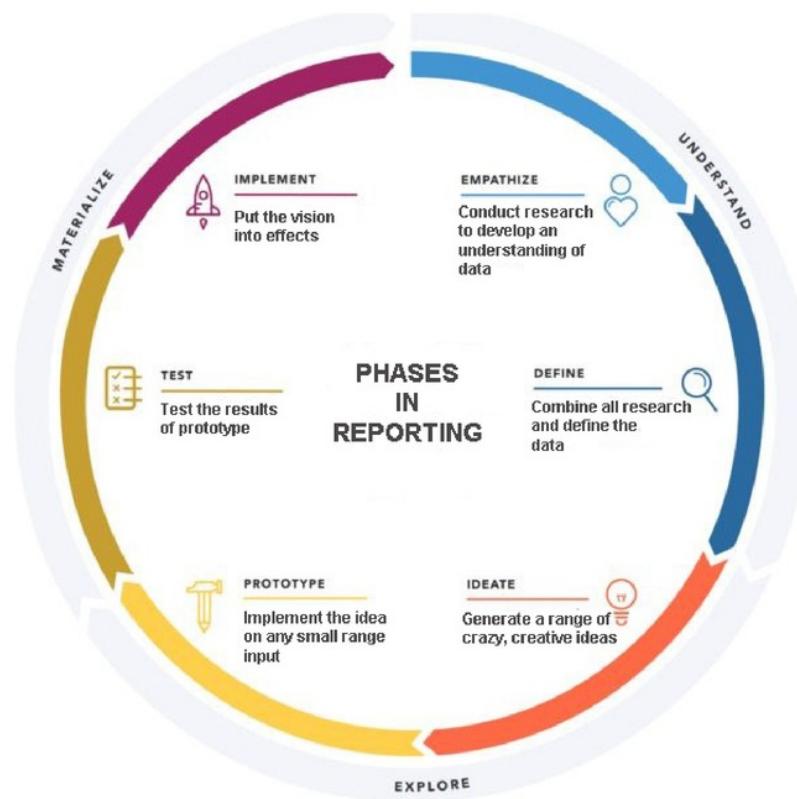


Figure 1.1: Different phases in reporting process.

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 direct speech? (Brunner,2013)**

Any information about any news event is reported by a personal body (reporter/user) in their own way of reporting i.e. thought process of any reporter/editors affects the article very much. Reporters are the eyes and ears of the very front in any news agency. They gather information with the help of their previous knowledge, public interaction, police records and other private and government informants. These types of data loses its accuracy and legitimacy because of the biased thinking of person. In ideal conditions, reporters should present a balanced story. Online news coverage system allows only one side of a debate to reach people. This is why unbiasness in reporting is really crucial.

### **1.1.4 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. It was supposed to <sup>8</sup>scop out more than 5 lakh crore Rs black money from the economy. Cash circulation is directly related to corruption in the country (according to PM Modi).

***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 Rajan's View on Demonetization.

So, for lowering the cash circulation and eliminating fake<sup>8</sup> currency and dodgy funds, demonetization became necessary because fake currencies and dodgy funds have been used by terrorist groups to fund terrorism in India.

## 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 <sup>27</sup> 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) and Liu (Liu and Zhang, 2012) have taken two detailed surveys. 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. (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 .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 “Sentiment analysis of demonetization of 500 1000 rupee banknotes by Indian government” 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 <sup>15</sup> e (2004)	SVM	Pos/Neg	Movie Reviews	Sentiment Analysis
3	Cao et al. (2011) <sup>15</sup>	Semantic , LSA- based	G	Software programs, users feedback	Sentiment Classification
4	Hu and Li (2011) <sup>9</sup>	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) <sup>9</sup>	Ranking algorithm	G	Book, Hotel, Notebook Reviews	Sentiment Classification
8	Xu et al. (2012) <sup>9</sup> IE	Markov Blanket, SVM, NB,	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 [17]. Mallejo(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 Jian-qiang, Gui Xiaolin (2016)	Random For-est,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 Surveil- lance	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.

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, just about every new poll of public sentiment shows us that confidence in the news reported by the online media is low. 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 is it just a case of perception trumping reality? In fact, there is little to show that over the past few decades online media made news reporting more favorable to one side. It does not mean though that researchers have not found bias in reporting. They have, but it is found that no one side is consistently favored and neither favoritism has been growing like a pernicious weed. The social media seems to be more biased as it is accessible to resourceful people.

### 1.3.2 Demonetization issue

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

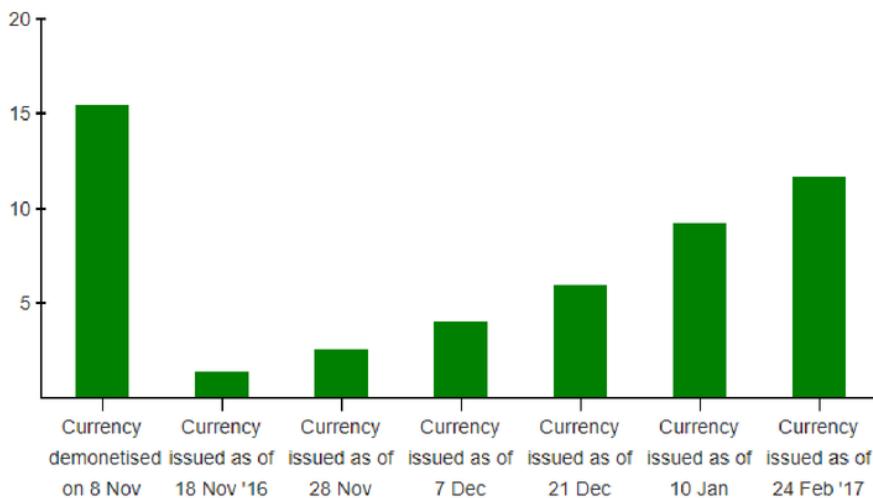


Figure 1.3: How much money is demonetized (in trillion rs.)?

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 project, <sup>1</sup>some difficulties were faced during saving such large data of tweets. Assembling and looking for themes has become important in industry, culture, sciences and humanities, which is not an easy task as data is very huge and complex. So we use python libraries pyth, nltk 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 twitterati tells influenced by his 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.

## 1.6 OBJECTIVES and RESEARCH QUESTIONS

### 1.6.1 Objectives

- To examine people perception of demonetization by parsing through the tweets with time.
- To establish the relationship between outcomes of different <sup>23</sup>algorithms (Naive Bayes, Maximum Entropy, Support Vector Machine).
- To compare the public opinion as per social media and print media.

### 1.6.2 Research Questions

- Which part of our community adores demonetization?
- What is the use of different algorithms?
- How to ensure perfect outcomes with maximum efficacy and efficiency?
- 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 DATA SET

#### 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	Tweet		retweeted	favorited	replied	toStatusId	retweeted	repliedToId	statusSource	Screen	
2	1	RT @Rakeshwarlal: Critical question: Was PM informed about #Demonetization on edict or	FALSE	0	NA	11/23/2016 18:40	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/M5hTfa	
3	2	RT @Hemant: 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/dt/PRAMoT	
4	3	RT @JoshiKaran: 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/dt/rahuvalai"	
5	4	RT @ANI_news: Gangaram (Haryana): Post office employees provide cash exchange to pa	FALSE	0	NA	11/23/2016 18:39	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/deeptiy	
6	5	RT @ANI_news: Veedu (Gujarat): Gram panchayat can't demonetize because #demonetized	FALSE	0	NA	11/23/2016 18:39	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/krishna1995	
7	6	@gauravkawant: India's demonetization - #blackmoney a symptom, not the disease itself	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	0.2E+09	<a href="http://twitter.com/dt/kmdebashis"	
8	7	RT @gauravkawant: Rs 40 lakh looted from a bank in Kathua in Jammu&Kashmir. Third such inc	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/kashidac"	
9	8	RT @Joydeep_912: Calling all nationalists to join	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/KARKUUP	
10	9	RT @sumithrasat2002: Many opposition leaders are with #demonetizemad on the	FALSE	0	NA	11/23/2016 18:38	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/purnibh"	
11	10	Nationalist reform now demands that the govt. ban the use of cagay. Such instances urge giving #t	FALSE	0	NA	11/23/2016 18:38	TRUE	NA	8.01E+17	NA	<a href="https://twitter.com/dt/ajitjigingle"	
12	11	RT @ANI_news: Many leaders are with #demonetization on the #Demonetization	FALSE	1	NA	11/23/2016 18:38	TRUE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/ANI_mnch"	
13	12	RT @Sneha_Gupta: In Narendra Modi App where we're raising feedbacks if people says	FALSE	0	NA	11/23/2016 18:37	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/Sneha_Gupta"	
14	13	@Jagesh21 want bind on 28? <a href="https://t.co/0ACdn1u920D">https://t.co/0ACdn1u920D</a> <url>#0089-110002> those w/	FALSE	0	Jagesh21	11/23/2016 18:37	FALSE	8.01E+17	0.2E+09	<a href="http://twitter.com/dt/yuvarti"		
15	14	RT @Adheesh_Krishna: The effect of #Demonetization ??	FALSE	0	NA	11/23/2016 18:36	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/PMKejari"	
16	15	RT @sonaksh205: 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/dt/nkgupta"	
17	16	RT @Deepak_Agami: The Modi govt. says one #Demonetization problem is that the govt.	FALSE	0	NA	12/21/2016 18:35	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/akshayd	
18	17	RT @ANI_news: The Modi govt. says one #Demonetization problem is that the govt.	FALSE	0	NA	12/21/2016 18:35	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/ANI_mnch"	
19	18	RT @Adheesh_Krishna: BEFORE and AFTER Gandhi ji said they are standing there	FALSE	0	NA	11/23/2016 18:34	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/snoozee"	
20	19	RT @pGurust: #Demonetization on The co-operative banking sector in Kerala is all good as i	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/Vishwas"	
21	20	RT @JoshiKaran: 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/dt/Politic9"	
22	21	RT @Hemant: 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/dt/akshayd"	
23	22	RT @Deepak_Agami: 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/dt/akshayd"	
24	23	RT @Adheesh_Krishna: BEFORE and AFTER Gandhi ji said they are standing there	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/Invarsh"	
25	24	RT @MuktinathP: @maretamodi...	FALSE	0	NA	11/23/2016 18:33	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/Gajapati"	
26	25	RT @Hemant: 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/dt/Rosa_Che"	
27	26	RT @reshankar: 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/dt/mukundp"	
28	27	RT @ANI_news: #Demonetization is a disaster. It will bring India to its knees. "Hurt" by #Demonetization s.t	FALSE	0	NA	11/23/2016 18:32	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/ANImnch"	
29	28	RT @ANI_news: #Demonetization is a disaster. It will bring India to its knees. "Hurt" by #Demonetization s.t	FALSE	0	NA	11/23/2016 18:32	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/Mirza"	
30	29	RT @Apil_Lukic: #Demotilak means #blackandwhite is "Hurt" by #Demonetization s.t	FALSE	0	NA	11/23/2016 18:32	FALSE	NA	8.01E+17	NA	<a href="http://twitter.com/dt/Pontif"	
31	30	RT @AAVInd: #Demonetization is Disaster! #nraem_pk	FALSE	0	NA	11/23/2016 18:31	FALSE	NA	8.01E+17	NA	<a href="http://www.twitter.com/dt/sehgal_j"	
32	31	RT @Hemant: 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/dt/BescoU"	

Figure 2.1: Dataset.

### 2.1.2 Training data set

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.

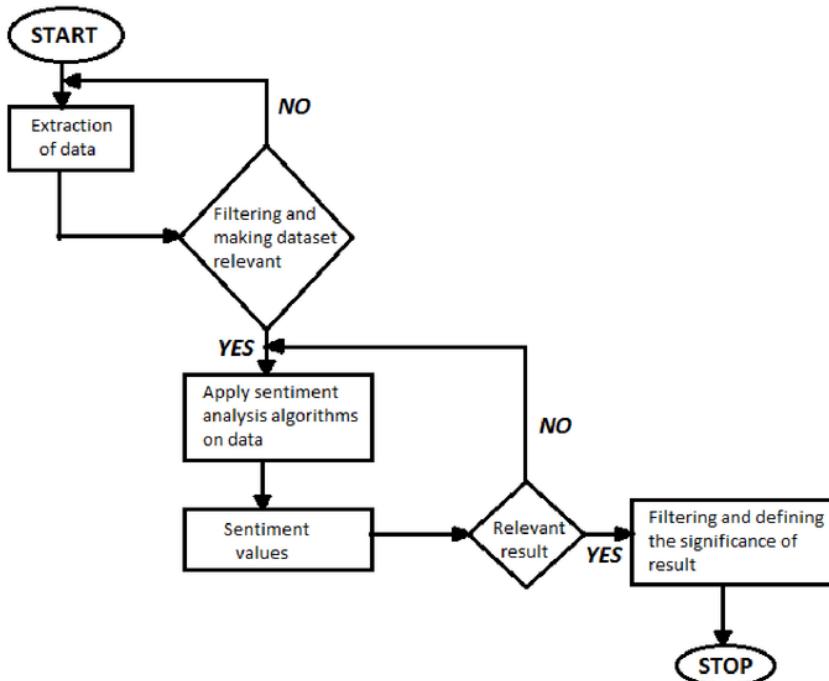
Matter in the scraped tweets being unorganized and mixed were tough to directly apply sentiment analysis on them. Also we would not have got more accurate results .so,we extract useful data from scraped tweet like date , id , screenname , no. of retweets etc . All the data were organised in .csv format . 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 -



16 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.SVM

13

### 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 various types of probability models, researcher uses naive bayes classifiers, because in a supervised learning <sup>13</sup> naive bayes classifiers can be trained <sup>10</sup> very easily and <sup>10</sup> efficiently. Naive bayes works using probabilistic model given below for each of k possible outcomes or classes

$$p(C_k|x) = p(C_k) p(x|C_k)/p(x)$$

But for larger value of n above formula doesn't work well So, Bayes theorem decomposes the conditional probability as <sup>12</sup>

$$p(C_k|x) = p(C_k) p(x|C_k)/p(x)$$

where  $C_k$  is class for each of k possible outcomes and  $x$  are the instances to be classified of a given problem represented as a vector.

<sup>5</sup>  $X = (x_1, x_2, x_3, \dots, x_n)$  representing n independent variables or features.

$P(c|x)$  is the posterior probability of class.

$P(c)$  is the prior probability of class.

$P(x|c)$  is the likelihood which is the probability of predictor given class.

$P_2(x)$  is the prior probability of predictor.

Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative sentiments).

## 2.4.2 <sup>19</sup> 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.

SVM is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems.

In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very 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.

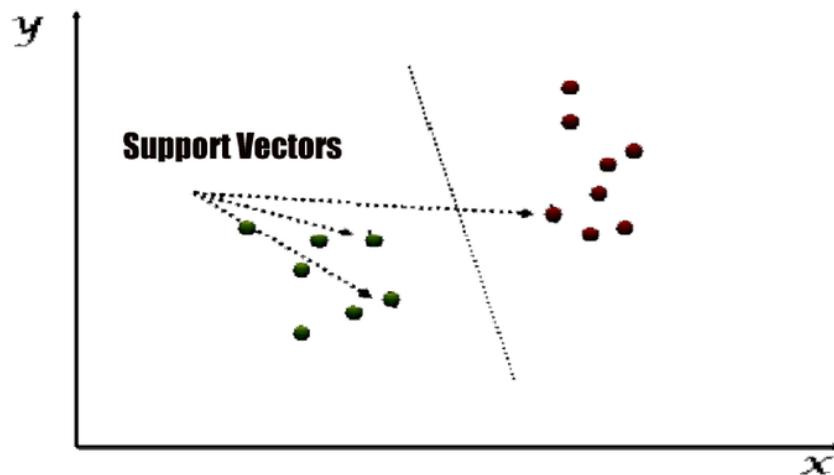


Figure 2.3: SVM

7 SVM is very effective in case of high dimensional spaces,i.e., in cases where number of dimensions is greater than number of samples. SVM uses a subset of training points in the decision function(called support vectors) ,so it is also memory efficient.

# **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 a table showing variation in sentiments of tweets with time in accordance of every week with respect to different methods.

Table 3.1: 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 %)
Week 1	32.90	27.00	15.00	15.00	52.10	58.00
Week 2	28.50	23.00	9.80	9.80	61.70	67.20
Week 3	33.40	35.60	13.40	12.30	53.20	52.10
Week 4	43.70	46.20	16.80	13.30	39.50	40.50
Week 5	34.10	77.10	16.30	16.90	49.60	06.00
Week 6	46.00	44.20	14.80	13.30	39.20	42.50
Week 7	61.20	63.10	11.10	08.50	27.70	28.40
Week 8	61.20	88.40	06.17	09.40	32.63	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.

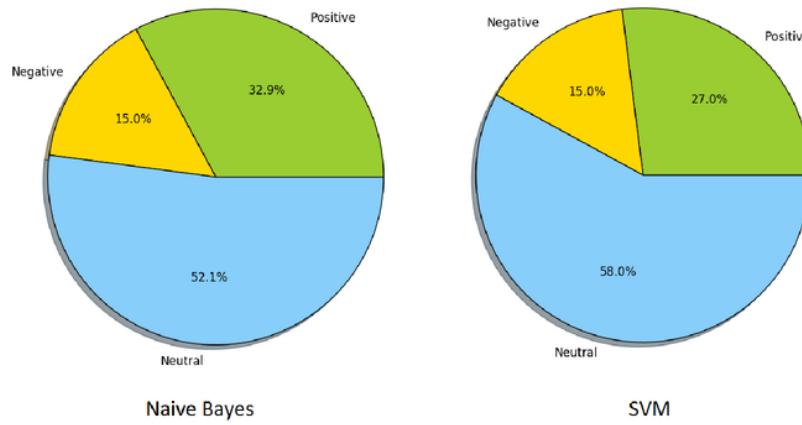


Figure 3.1: Week 1 of demonetization

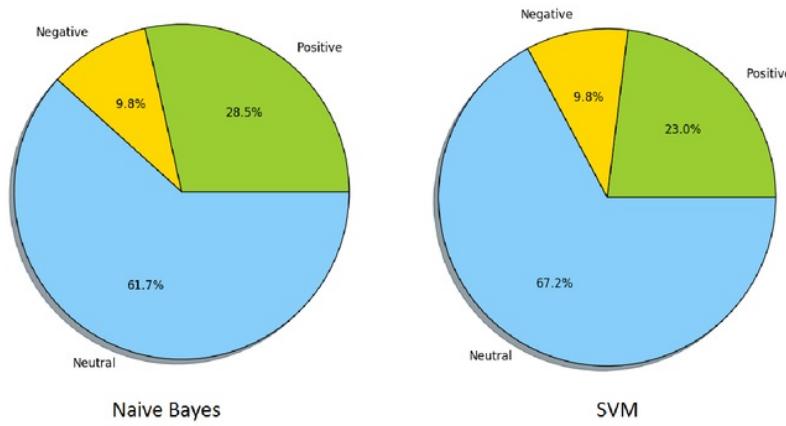


Figure 3.2: Week 2 of demonetization

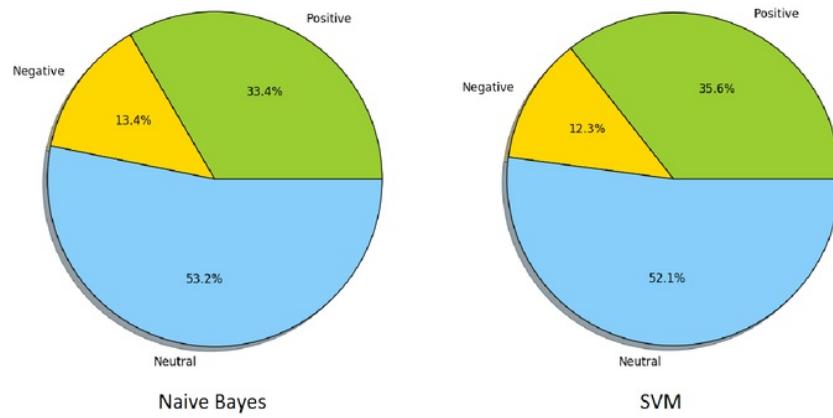


Figure 3.3: Week 3 of demonetization

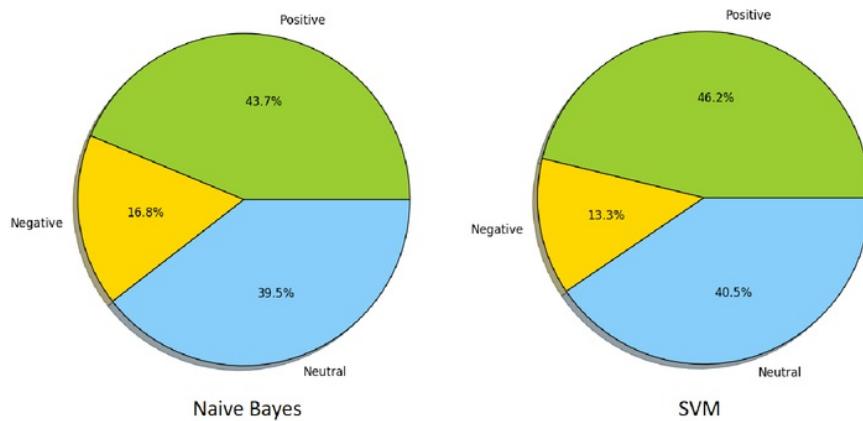


Figure 3.4: Week 4 of demonetization

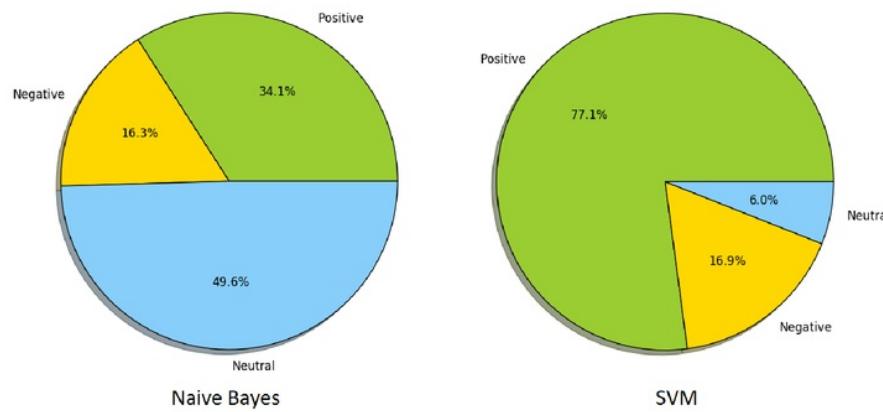


Figure 3.5: Week 5 of demonetization

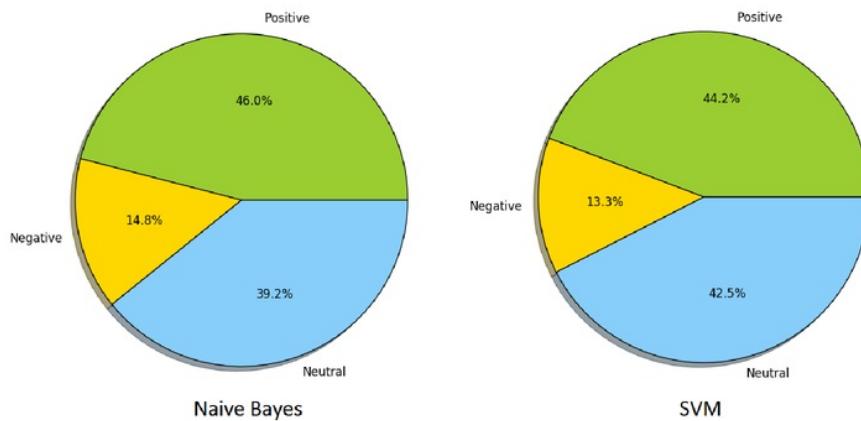


Figure 3.6: Week 6 of demonetization

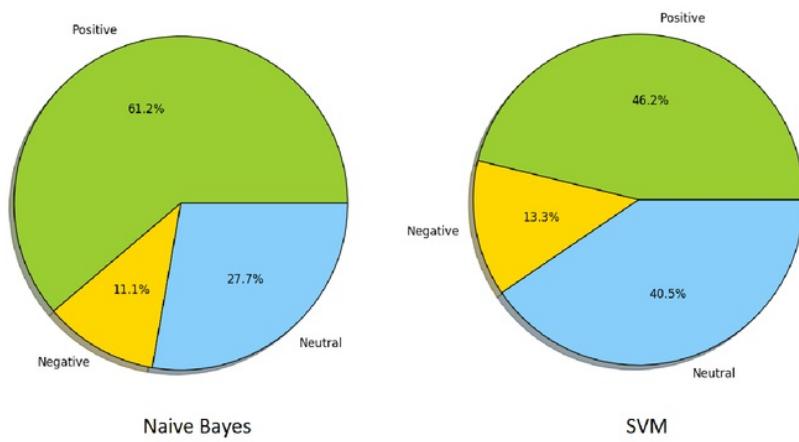


Figure 3.7: Week 7 of demonetization

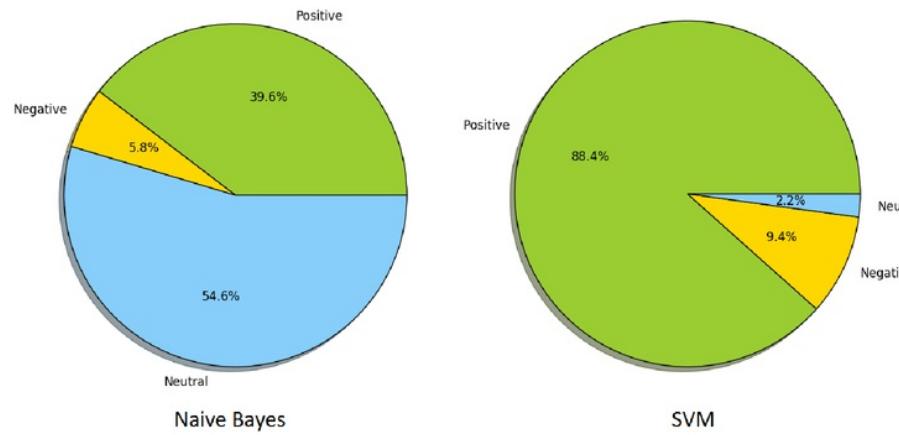


Figure 3.8: Week 8 of demonetization

### 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.

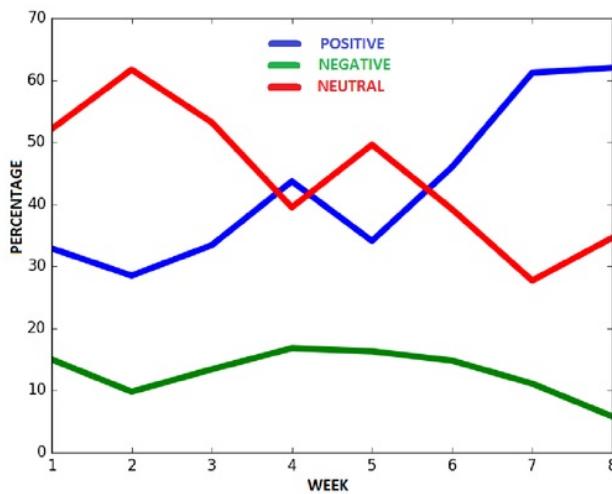


Figure 3.9: Variation of all sentiments through naive bayes over time

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

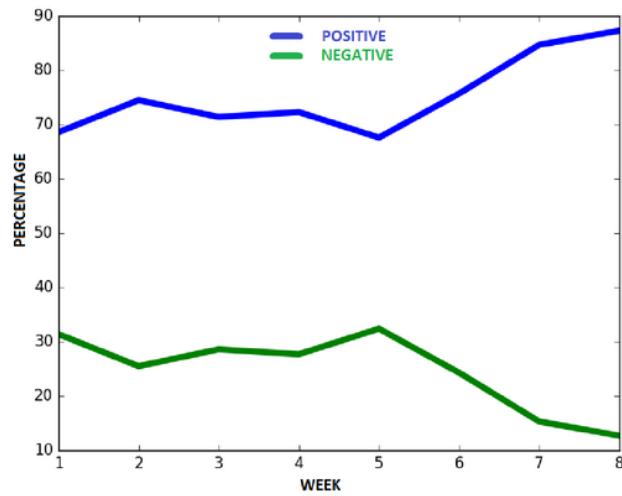


Figure 3.10: Comparison of positive and negative sentiments through naive bayes over time

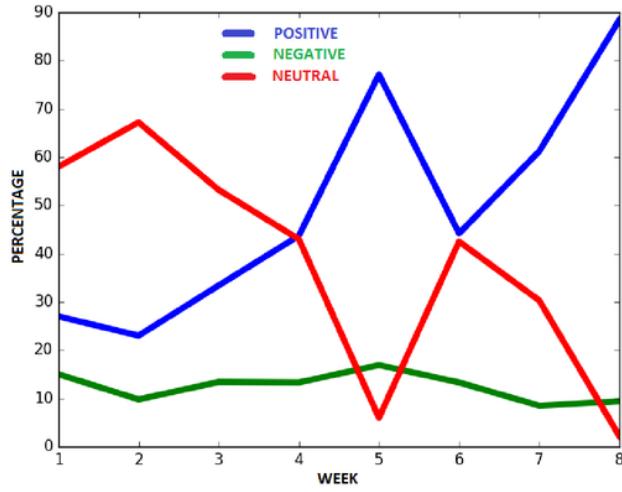


Figure 3.11: Variation of all sentiments through SVM over time

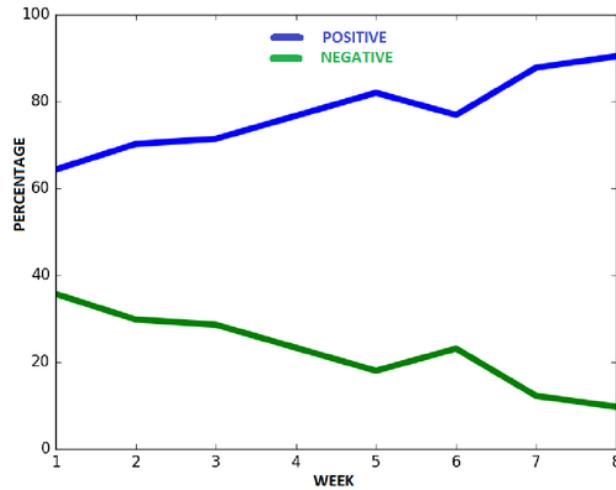


Figure 3.12: Comparison of positive and negative sentiments through SVM over time

### 3.3 Result

Table 3.2: 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.

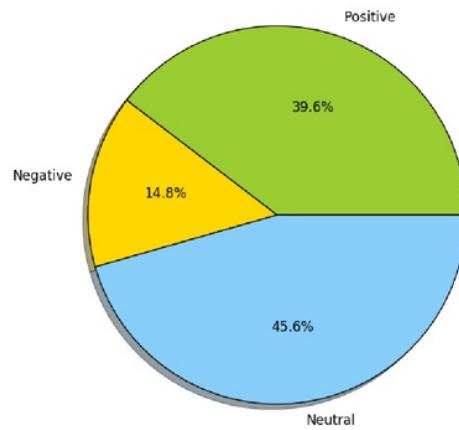


Figure 3.13: Overall sentiments through Naive Bayes

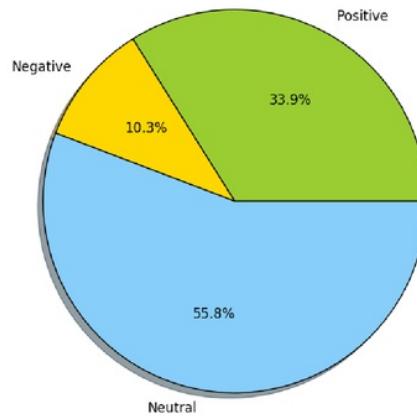


Figure 3.14: Overall sentiments through SVM

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] Go, A., Bhayani, R. and Huang, L.: 2009, Twitter sentiment classification using distant supervision, *CS224N Project Report, Stanford* **1**(2009), 12.
- [4] He, Y. and Zhou, D.: 2011, Self-training from labeled features for sentiment analysis, *Information Processing & Management* **47**(4), 606–616.
- [5] 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.
- [6] Jianqiang, Z. and Xiaolin, G.: 2017, Comparison research on text pre-processing methods on twitter sentiment analysis, *IEEE Access* **5**, 2870–2879.
- [7] Liu, B. and Zhang, L.: 2012, A survey of opinion mining and sentiment analysis, *Mining text data*, Springer, pp. 415–463.
- [8] 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.
- [9] Pang, B., Lee, L. et al.: 2008, Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval* **2**(1–2), 1–135.
- [10] 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.

- [11] Saif, H., He, Y. and Alani, H.: 2012, Semantic sentiment analysis of twitter, *The Semantic Web-ISWC 2012* pp. 508–524.
- [12] Singh, P., Sawhney, R. S. and Kahlon, K. S.: 2017, Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by indian government, *ICT Express* .
- [13] 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.



### PRIMARY SOURCES

- |   |   |    |
|---|---|----|
| 1 | Submitted to ABV-Indian Institute of Information Technology and Management Gwalior      | 8% |
| 2 | www.analyticsvidhya.com<br>Internet Source  | 2% |
| 3 | Submitted to Higher Education Commission Pakistan                                       | 2% |
| 4 | Submitted to Poway Unified School District<br>Student Paper                             | 1% |
| 5 | docplayer.net<br>Internet Source  | 1% |
| 6 | Submitted to University of KwaZulu-Natal<br>Student Paper                               | 1% |
| 7 | Submitted to The Robert Gordon University<br>Student Paper                              | 1% |
| 8 | "What is demonetisation and why was it done. (Policy)", The Economic Times, Nov 10 2016 | 1% |

## Issue

Publication

9	<a href="http://www.verticale.net">www.verticale.net</a> Internet Source	1 %
10	Submitted to CSU, San Jose State University Student Paper	<1 %
11	Submitted to Contra Costa Community College District Student Paper	<1 %
12	Submitted to Savitribai Phule Pune University Student Paper	<1 %
13	<a href="http://www.di.unito.it">www.di.unito.it</a> Internet Source	<1 %
14	<a href="http://iswc2012.semanticweb.org">iswc2012.semanticweb.org</a> Internet Source	<1 %
15	Submitted to Pusan National University Library Student Paper	<1 %
16	<a href="http://kt.ijss.si">kt.ijss.si</a> Internet Source	<1 %
17	<a href="http://dl.acm.org">dl.acm.org</a> Internet Source	<1 %
18	<a href="http://www.aclweb.org">www.aclweb.org</a> Internet Source	<1 %
19	<a href="http://www.evsys.net">www.evsys.net</a>	

- 
- 20 Wu, Desheng Dash, Lijuan Zheng, and David L. Olson. "A Decision Support Approach for Online Stock Forum Sentiment Analysis", IEEE Transactions on Systems Man and Cybernetics Systems, 2014. <1 %  
Publication
- 
- 21 gradworks.umi.com <1 %  
Internet Source
- 
- 22 Liu, Li Luo, Dashi Liu, Ming Zhong, Jun . "A self-adaptive hidden Markov model for emotion classification in Chinese microblogs.(Research Artic", Mathematical Problems in Engineering, Annual 2015 Issue <1 %  
Publication
- 
- 23 aaltodoc.aalto.fi <1 %  
Internet Source
- 
- 24 aclweb.org <1 %  
Internet Source
- 
- 25 Submitted to American University in Cairo <1 %  
Student Paper
- 
- 26 Albert Bifet. "Sentiment Knowledge Discovery in Twitter Streaming Data", Lecture Notes in Computer Science, 2010 <1 %  
Publication

27	<a href="http://www.scribd.com">www.scribd.com</a> Internet Source	<1 %
28	Submitted to University of Applied Sciences Berlin Student Paper	<1 %
29	Studies in Computational Intelligence, 2016. Publication	<1 %
30	Submitted to Bath Spa University College Student Paper	<1 %

---

Exclude quotes

On

Exclude matches

Off

Exclude bibliography

On