

# Implementing sentiment analysis in relevance with Indian Elections

Amandeep Kaur<sup>1</sup>  
Research Scholar, PhD (Comp. Appl.)  
Guru Kashi University, Talwandi Sabo  
Punjab, India.

Dr. Rajinder Singh<sup>2</sup>  
Assistant Professor, UCCA  
Guru Kashi University, Talwandi Sabo  
Punjab, India.

## Abstract

The sentiment analysis has emerged as a prominent topic of research in recent past. Tracking different types of opinions and sentiments has enabled different political leaders and political parties to have insight of the general public view regarding their political performances. The sentiment analysis guides those electorates who are confused and fail to decide who to vote for. The research paper emphasized on collecting tweets and Facebook posts related to Indian elections and performing sentiments analysis on them. The paper illustrates the methodology adopted to perform sentiment analysis and reach a definite conclusion.

**Keywords** – *Elections, negative words, sentiment analysis, positive words, text mining, tweets.*

## I. INTRODUCTION

The process of investigating data to collect valuable information is referred as text mining. Text mining comprises of data mining algorithms, statistical operations, machine learning, and NLP to come out with informative content from unstructured formats. The text mining has found its applications in different sectors. The primary reason behind this the importance gained by data to reach general masses and availability of non-expensive tools. Text analytics has assisted businesses to extract insights about customer's preferences and expectations. The feedback data can be used to identify trends and patterns. Manually analyzing can't capture this level of insight because of huge volume and complexity carried by the text. So, text mining can be defined as a process of analyzing data in order to get hold of key concepts and expose hidden relationships without prior knowledge of specific words or terms used by authors to express those notions. The basic steps involved in text mining are mentioned as under and shown in Fig. 1.

- First in the process is to collect unstructured data from different multiple data sources like web pages, pdf files, blogs, plain text and much more.
- Perform pre-processing of the collected data and clean up the data by removing noise and anomalies from the collected data.
- The extracted information should be converted into structured format.
- Analyze the patterns (visible and hidden) within the data.
- Finally, the valuable information should be properly stored in secure database for assisting decision making process of the organization.

# Text Mining

Text mining involves a series of activities to be performed in order to efficiently mine the information. These activities are:

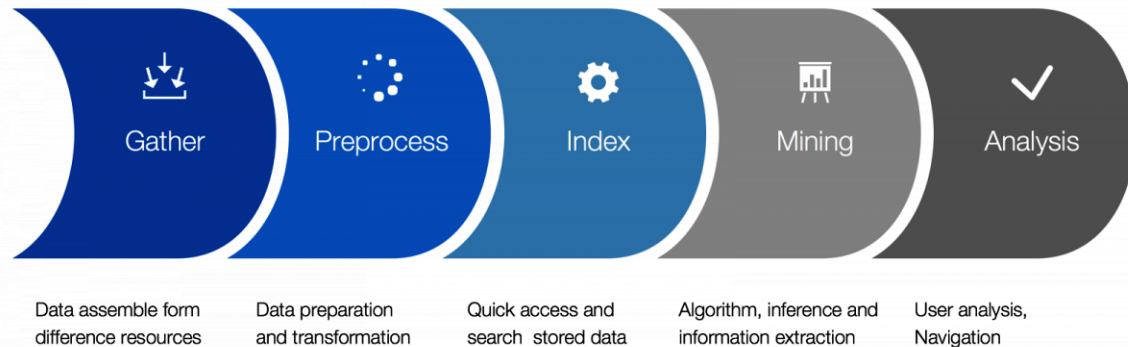


Fig. 1: Figure depicts the sequential steps involved in text mining

Few advantages of text mining are mentioned as under.

Text mining helps to utilize available resources to optimum level and performs better than human brains.

- Text mining assists in tracing opinions overtime.
- It extracts notions from text and presents it more effectively.
- It helps in summarizing large documents.
- The information extracted from text mining can be effectively used for predictive analysis.
- It helps in finding relationships and correlations among different entities using appropriate techniques.
- A particular hypothesis can be tested using text mining to see whether the document confirm or deny the hypothesis under study.
- Text mining assists in smart decision making.
- It helps to discover knowledge in different application areas.
- It provides outputs in different visualizations lie tables, graphs, and charts.

Different techniques involved in text mining are mentioned as under.

- **Extracting information:** This includes extraction of valuable information from huge textual data. It emphasizes on identifying extraction of relationships among participating entities and attributes. The extracted information is stored in database. The effectiveness of the outcomes is periodically checked and evaluated using precision and recall processes.
- **Retrieving information:** This refers to conducting extraction of associated patterns based on phrases or words. It uses different algorithms for tracking and monitoring user behaviors and find out data accordingly. For instance, Google and Yahoo search engines.
- **Categorization:** It refers to assigning normal language texts to predefined set of topics depending upon their contents. Natural Language Processing is another name for categorization. Co-referencing is often used to extract relevant abbreviations and synonyms from textual data.

- **Clustering:** Clustering is supposed to recognize intrinsic structures in textual information and organizing them into relevant subgroups. The major challenge faced by clustering is to come up with formation of meaningful clusters from unlabeled textual data without earlier information on them.
- **Summarization:** Summarization is intended for generation of compressed version of text having valuable information. It involves browsing through multiple sources and constructing information in a concise format keeping meaning of the original document essentially the same. The methods like decision trees, regression models, swarm intelligence, and neural networks can be used for this.

Technique	Characteristics	Tools
Retrieval	Retrievals valuable information from unstructured text	Intelligent Miner, Text Analyst
Extraction	Extract information from structured database	Text Finder, Clear Forest Text
Summarization	Reduce length by keeping its main points and overall meaning as it is	Tropic Tracking Tool, Sentence Ext Tool
Categorization	Document based categorization	Intelligent Miner
Cluster	Cluster collection of documents, Clustering, classification and analysis of text document	Carrot, Rapid Miner

Fig. 2: Figure depicts different techniques involved in text mining

Different methods involved in text mining are mentioned as under.

- **Term Based Method:** Term refers to a word having semantic meaning. The document is analyzed on the basis of term. This method suffers from the problem of synonymy (multiple words having similar meaning) and polysemy (single word having multiple meanings).
- **Phrase Based Method:** This method analyzes the document based on phrases which are less obvious to more meanings and are more discriminative. This method has drawbacks of poor statistical properties to terms, low existence frequency, and adequate noisy phrases.
- **Concept Based Method:** This method analyzes the document on the basis of sentence and document level. This method is good enough to differentiate between important and unimportant words. Firstly, the meaningful part of the sentences is examined. Secondly, a conceptual ontological graph is produced to explain the structures. Lastly, the main components are extracted based on earlier doings.
- **Pattern Taxonomy Method:** This method makes use of patterns in analyzing documents. Data mining techniques like sequential pattern mining, association rule mining, closed pattern mining, and frequent item set mining are used. This method use pattern deploying and pattern evolving and is superior as compared to other methods.

## II. TEXT MINING TOOLS/SOFTWARES

Some of the popular text mining software's/tools are described below.

### RapidMiner

RapidMiner is open source platform for data science comprising text mining, data mining, and predictive analytics. The most appreciated feature of RapidMiner is that there is no need to write the code and don't have to know about working math behind it. RapidMiner offers user friendly environment for sentiment analysis. RapidMiner can read data from Twitter, and also read csv and excel files. It offers powerful text processing features in combination with clustering algorithms and machine learning operators. RapidMiner is proficient in performing word counts and word frequency where word occurs in corpus. It is capable of doing sentiment analysis using machine learning algorithm.

### KNIME Analytics platform

KNIME provides an in-built visual workflow for programming free data mining. It offers as many operators as RapidMiner does. It is capable of performing below mentioned tasks.

- Stemming: It converts variations on key terms into basic forms.
- Stop word filtering: KNIME is capable of removing insignificant terms like “for”, “in”, and “the”.
- Tokenization: Tokenization deals with breaking of text strings into smaller units.
- KNIME is capable of reading directly from Twitter and flat files such as csv.

### Apache mahout

Mahout is a data mining framework which break down a computational task into multiple segments and assigns each segment to a different machine. Mahout runs united with Hadoop infrastructure and manages enormous amount of data. Mahout implements prevalent machine learning algorithms like recommendation, classification, and clustering. Mahout performs good in distributed environment. It makes use of Hadoop library to perform efficiently in the cloud. Mahout provides ready to use user friendly framework for performing mining on voluminous data. It enables assessing of large amount of data in no time with high precision. It comprises of many MapReduce clustering implementations like k-means, Mean-shift, Canopy, and Fuzzy –Means. It also provides support to Naïve Bayes and complementary Naïve Bayes classification implementations.

## III. IMPLEMENTATION AND RESULTS

The steps adopted for performing sentiment analysis on election relevant data are mentioned as under.

1. An unstructured database has been constructed by collecting tweets, messages, and comments from social media in a text file.
2. A dictionary is constructed which comprises of positive and negative words. Each word is assigned a score depending upon the intensity of the word. For instance, abandon -2, abandoned -2, abandons -2, abducted -2, affection 3, affectionate 3, agrees 1 etc.
3. A source code in Java is developed and made to run in Eclipse. The keywords in the constructed database are extracted and on the basis of the occurrence and the score assigned to them in the dictionary, the value of each tweet is calculated.
4. The values of all the tweets are finally calculated and the final score is obtained. If the final score is a positive value, then it can be concluded that the overall impact of the collected tweets ended on a positive note and vice versa.

Some of the tweets having relevance with CongressManifesto (#CongressManifesto) are mentioned below.

- *The best part i found from Congress Manifesto is : No application fee for Govt Jobs.*
- *At the #CongressManifesto release, Manmohan Singh says their manifesto should be subject to debate across the country. We were fortunate to call him our PM once.*
- *South feels hostility from Narendra Modi": Rahul Gandhi's Wayanad retort*
- *Here are some promises which I personally love in the #CongressManifesto Education will get 6% share in Budget (currently it's lowest in decade at 3.5%) Removal of Electoral Bonds (Modi got this disaster in 2017 which legalized corruption in funding of pol. parties).*
- *Sam Pitroda, Indian Overseas Congress Chief, on party's election manifesto: Don't underestimate our ability to deliver. We promised we'll deliver. We've delivered in the past and we'll deliver in future (news agency ANI).*
- *A quick calculation of the various promises in #CongressManifesto suggest they would add up to around Rs 10 lakh crores; budget estimate for 2019-20 is Rs27.8 lakh crore! So where is the addl money coming from? As is often said: 'if you can't keep a promise, don't pledge it!'*
- *#CongressManifesto 1. NEET Exemption 2. NYAY, minimum guarantee scheme 3. Separate budget for Farmer 4. Abolishing Angel Tax 5. 22 lakhs government jobs by March 2020 6. Reforming Goods and Services Tax (GST) 7. National security (Reco. of Lt Gen. (Retd) DS Hooda*
- *Congress Manifesto is for the people , by the people,of the people -it captures the mood of the Nation, reflects the desires & aspirations of young India #CongressManifesto*
- *After the release of the #CongressManifesto, we at @ProfCong are excited to welcome Mr. @PChidambaram\_IN to discuss the economic proposals put forward by @INCIndia. Jobs, farmers, MSMEs, tax reform & more. Join us today at 5.30 PM (sharp) at the Constitution Club in New Delhi.*
- *Dr Manmohan Singh endorses NYAY or minimum income guarantee scheme . He claims UPA got 14 crore out of poverty and says by 2030 everyone should be out of poverty. #CongressManifesto*

The results obtained after running the constructed java source code in Eclipse are mentioned as under.

```

The best part i found from Congress Manifesto is : No application fee for Govt Jobs.=3.0}
Current Counter score=3.0
{At the #CongressManifesto release, Manmohan Singh says their manifesto should be subject to
debate across the country. We were fortunate to call him our PM once.=2.0}
Current Counter score=5.0
{South feels hostility from Narendra Modi": Rahul Gandhi's Wayanad retort=0.0}
Current Counter score=5.0
{Here are some promises which I personally love in the #CongressManifesto Education will get 6%
share in Budget (currently it's lowest in decade at 3.5%) Removal of Electoral Bonds (Modi got
this disaster in 2017 which legalized corruption in funding of pol. parties).=2.0}
Current Counter score=7.0
{Sam Pitroda, Indian Overseas Congress Chief, on party's election manifesto: Don't underestimate
our ability to deliver. We promised we'll deliver. We've delivered in the past and we'll deliver
in future (news agency ANI).=2.0}
Current Counter score=9.0
{A quick calculation of the various promises in #CongressManifesto suggest they would add up to
around Rs 10 lakh crores; budget estimate for 2019-20 is Rs27.8 lakh crore! So where is the addl
money coming from? As is often said: 'if you can't keep a promise, don't pledge it!'=1.0}
Current Counter score=10.0
{#CongressManifesto 1. NEET Exemption 2. NYAY, minimum guarantee scheme 3. Separate budget for
Farmer 4. Abolishing Angel Tax 5. 22 lakhs government jobs by March 2020 6. Reforming Goods and
Services Tax (GST) 7. National security (Reco. of Lt Gen. (Retd) DS Hooda=1.0}
Current Counter score=11.0
{Congress Manifesto is for the people , by the people,of the people -it captures the mood of the
Nation, reflects the desires & aspirations of young India #CongressManifesto=0.0}
Current Counter score=11.0
{After the release of the #CongressManifesto, we at @ProfCong are excited to welcome Mr.
@PChidambaram_IN to discuss the economic proposals put forward by @INCIndia. Jobs, farmers,

```

MSMEs, tax reform & more. Join us today at 5.30 PM (sharp) at the Constitution Club in New Delhi.=5.0}

Current Counter score=16.0

{Dr Manmohan Singh endorses NYAY or minimum income guarantee scheme . He claims UPA got 14 crore out of poverty and says by 2030 everyone should be out of poverty. #CongressManifesto=2.0}

Current Counter score=18.0

{BJP abducted nation. =-2.0}

Current Counter score=16.0

Final Counter score=16.0

Sentiment Analysis ended on a positive note

#### IV. CONCLUSION

The research paper successfully illustrated the process of performing sentiment analysis on unstructured database. The paper illustrated and emphasized on analyzing tweets and facebook posts to have insight of sentiments of particular audience for particular political leader or political party. Sentiment analysis acts as a guidance to those electorates who are still confused to decide that whom they should vote for.

#### REFERENCES

1. Mohey El-Din Mohamed Hussein, D.: A survey on sentiment analysis challenges. J. of King Saud Univ. – Engg. Sci., 1 – 9 (2016) .
2. Jagdev, G. (2018). Sentiment Analysis and its Impact in Modeling Election Scenario. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 5(2), pp.22-27. <http://dx.doi.org/10.20431/2349-4859.0502004>.
3. Bermingham, A., Smeaton, A. F.: Classifying sentiment in microblogs: is brevity an advantage? In Proc. of the 19th ACM int. conf. on Inform. and Know. Manag. 1833 – 1836 (2010).
4. Sharma, Y., Mangat, V., Kaur, M.: A Practical Approach to Sentiment Analysis of Hindi Tweets. In Proc. of the 2015 1st Int. Conf. on Next Gen. Comp. Techno. (NGCT). 677 – 680 (2015).
5. Gagandeep Jagdev et al., “A Comparative study of Conventional Data Mining Algorithms against Map-Reduce Algorithm”, in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN (O) – 2319-8354, ISSN (P) – 2319-8346, Volume – 06, Issue – 05, May 2017.
6. Elghazaly, T., Mahmoud, A., Hefny, H. A.: Political Sentiment Analysis Using Twitter Data. In Proc. of the Int. Conf. on Inter. of things and Cloud Comp., 11 (2016).
7. Sharma, P., Moh, T-S.: Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter. 2016 IEEE Int. Conf. on Big Data (Big Data). 1966 – 1971 (2016).
8. Gagandeep Jagdev et al., “Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data”, in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550, Volume – 05, Issue – 05, May 2017.
9. Kharde, V. A., Sonawane, S.S.: Sentiment Analysis of Twitter Data: A Survey of Techniques. Int. J. of Comp. App. 139(11), 5 – 15 (2016)
10. Medhat, W., Hassan, A. Korashy, H.: Sentiment analysis algorithms and applications: A survey. A. S. Engg. J. 5, 1093–1113 (2014).
11. Amandeep Kaur, & Dr. Gagandeep Jagdev (2017). Exploring Application of Big Data in Elections – From Data to Action, International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 4(4), pp.64-71, DOI: <http://dx.doi.org/10.20431/2349-4859.0404008>.
12. Niyogi, M., Pal, A. K. Discovering conversational topics and emotions associated with Demonetization tweets in India. ArXiv: 1711.04115v1 [cs.CL]. 1 – 6 (2017).
13. Vijayarani, Dr. S., Ilamathi, J., Nithya, M.: Preprocessing Techniques for Text Mining - An Overview. Int. J. of Com. Sci. & Comm. Net. 5(1), 7-16 (2015)
14. Gagandeep Jagdev et al., “Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data”, in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550, Volume – 05, Issue – 05, May 2017.

15. Mohammad, S. M., Kiritchenko, S. Using Hashtags to Capture Fine Emotion Categories from Tweets. Sp. issue on Semantic Analy. in Soc. Med., Comp. Intell. 1 –22 (2013).
16. Gagandeep Jagdev et al., “A Study of Clustering and Classification Techniques involved in Data Mining”, in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550, Volume – 05, Issue – 05, May 2017.
17. Sharma, G.: Extract the context: Sentiment Analysis and Opinion Mining. ParallelDots. (2017).
18. Gagandeep Jagdev et al., “Analyzing and Filtering Big Data concerned with elections via Hadoop Framework” in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN (O): 2319-8354, ISSN (P): 2319-8346, Volume No. 6, Issue No, 4, April 2017.
19. Vu Dung Nguyen, Blesson Vaghese, “Royal Birth of 2013:Analysing and Visualising Public Sentiment in the UK using Twitter,” Research Gate, 2013.
20. Gagandeep Jagdev et al., “Comparing Conventional Data Mining Algorithms with Hadoop based Map-Reduce Algorithm considering elections perspective”, in International Journal of Innovative Research in Science and Engineering (IJIRSE), ISSN: 2454-9665 (O), ISSN: 2455-0663(P), Volume – 3, Issue – 3, March 2017.
21. Sarah E. Shukri, RawanI.Yaghi, “Twitter Sentiment Analysis: Case Study In Automotive Industry,” IEEE-Jordan Conference on Applied Electrical Engineering and Computing Technologies, 2015.
22. Gagandeep Jagdev et al., “Big Data Proposes an Innovative concept for contesting elections in Indian Subcontinent” in International Journal of Scientific and Technical Advancements (IJSTA), ISSN-2454-1532, 2015.
23. Subhabrata Mukherjee, Akshat Malu,“ TwiSent:A MultiStage System For analyzing Sentiment in Twitter,” Proceedings of the 21st ACM International conference on Information and knowledge management, 2012.
24. Mostafa Karamibekr, Ali a.Ghorbani, ”A structure for opinions in Social Domains,” Social Computing (SocialCom), International Conference, 2013.
25. Gagandeep Jagdev et al., “Excavating Big Data associated to Indian Elections Scenario via Apache Hadoop” in International Journal of Advanced Research in Computer Science (IJARCS), ISSN 0976 – 5697.
26. Blessy Selvam, S. Abirami, “A Survey on OM Framework,“ International Journal of Advance Research in Computing & Communication engine- vol 2 Issue 9, Sept 2013 .
27. M.Lovelin Ponn Felciah,R.A nbuselvi, ”A Study on Sentiment Analysis of Social Media Reviews,” IEEE-Second Conference on Innovations in Information Embedded & Communication Systems, 2015.