

Sentimental Analysis on Facebook comments using Data Mining Technique

Rupinder Kaur
Research Scholar,
Punjabi University, Patiala

Dr. Harmandeep Singh
Assistant Professor,
Punjabi University, Patiala

Dr. Gaurav Gupta
Assistant Professor,
Punjabi University, Patiala

Abstract: Data mining is the investigation periods of the “data discovery in documents”. This is a method for deciding plans and extracting the information from huge set of data. It is the procedure of mining knowledge from data. Sentiment analysis refers to the use of natural language processing [4]. Sentiment analysis, also called opinion mining, is the field of study that analyzes people’s sentiments, opinions, and emotions towards entities [7]. These entities might be a thing or a film, surveys of people, products, issues and topics that truly matters. Social sites for example Facebook and Twitter are that, where characters put their status or sentiments. People comment on their facebook account concerning any correct subject of their consideration [3].

Keywords: Data mining, sentimental analysis, facebook comments, classification, slang words

1. INTRODUCTION

1.1 Introduction to data mining

Data mining is the investigation periods of the “data discovery in documents”. This is a method for deciding plans and extracting the information from huge set of data. It is the procedure of mining knowledge from data. Data mining, also called knowledge discovery in databases, is the process of discovering interesting and useful patterns [17]. The aim of data mining process is to gather information from a data set and convert it into understandable form [18].

1.2 Process of data mining or knowledge discovery in database

Knowledge discovery in database is the method of extracting useful information from subordinate databases. Knowledge discovery in database or data mining involves the steps like data processed from data selection, interpretation, cleansing, transforming the raw data into some information, integrating and evaluating the pattern for that information received. Following are the overview of stages in the data identification process [10].

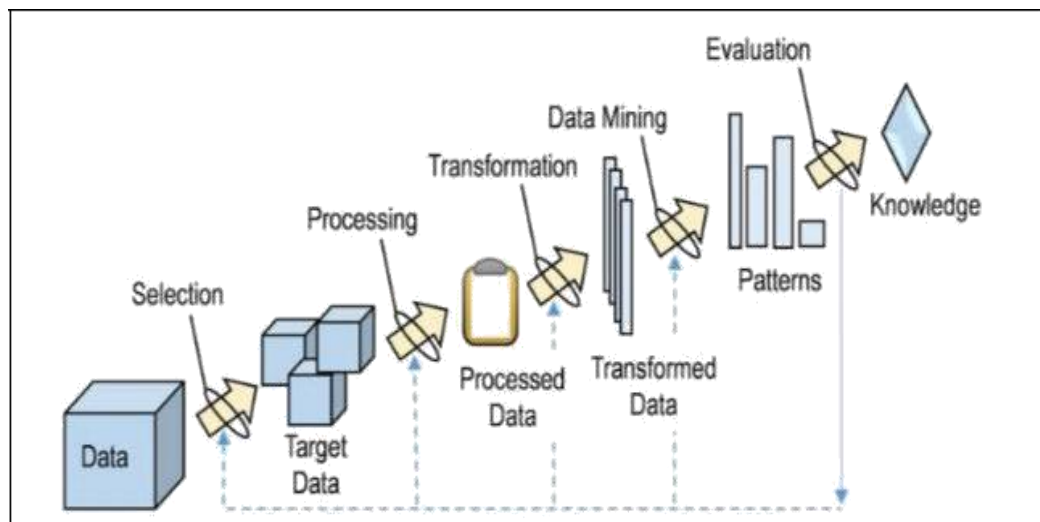


Figure 1: Process of data mining or knowledge discovery process [10]

1.3 Data mining techniques: Following are often used techniques in data mining.

- ❑ **Classification:** It is the most commonly applied data mining technique. But it requires set of pre-defined examples to create a model and applied on large datasets [10].
- ❑ **Support vector machine:** These are supervised learning models with associated learning algorithms that analyses data used for regression analysis and classification [10].
- ❑ **Decision tree:** Decision tree is a graph of decisions and their possible significances, represented in the form of branches and nodes. A decision tree contains a root node, branches, and leaf nodes [10].
- ❑ **Regressions:** A task used to predict the number such as age, height, income distance etc [10].

- **Prediction:** It is a method used in predicting the outcome based on available data by incorporating with unavailable data sets of future [10].

1.4 Data mining hierarchical model

There are various ways to store the huge volumes of information and computational procedures. Models are required to separate the unseen examples and learning. These strategies are utilized to change the information into helpful data, to make advance investigation, distortion discovery, discover the client expectations and so forth.



Figure 2: Data mining hierarchical model [12,18]

The text mining distinguishes the advantageous data in the literary archives or records. Web mining is the technique to gather the useful information from the sites or online audits. Web mining is divided into 3 sub parts. Web usage mining is procedure to discover the use of any sites i.e. how can the clients use some specific site. Web content mining is utilized to discover the valuable data from the real substance or material that is composed on the sites, which can be in any frame like tweets, remarks and audits of various clients. Web structure mining is the strategy to discover the general structure of the online destinations or web journals.

Opinion mining is the further advance in the web content mining [12,18]. Opinion mining is a sub part of web content mining which is also called as sentiment analysis. It is a process of finding users opinion about any topic or event [18]. Web content mining gathers the information from the web destinations and sentiment mining discover the point of view of open towards a particular subject or region.

1.5 Introduction to sentiment analysis

Social media has given web users a venue for expressing and sharing their thoughts on different events [5]. Sentiment analysis is a method of computing and satisfying a view of a person given in a piece of a text, to identify persons thinking about any topic is positive negative or neutral [2]. Sentiment analysis refers to the use of natural language processing [4]. Social sites for example Facebook and Twitter are that, where characters put their status or sentiments. People comment on their facebook account concerning any correct subject of their consideration [3]. Feeling examination is a training to sort the demeanor of the person that might be communicated as tweets. Tweets can be named positive, negative or neutral. For example, the tweet "movie was amazing" is a positive content and the tweet "movie was worst" is a negative content [3, 8].

Facebook become a social site, where lots of people can exchange their judgments & opinion about any current issues. Facebook mainly used to express view on certain topic [2]. Sentiment analysis of Facebook data is providing an effective way to expose user opinion which is necessary for decision making in various fields. Facebook allows the user to post real time short messages called as comments. These comments are restricted to 140 characters in length [2, 14, 16].

1.6 Process of sentiment analysis: Sentiment mining can be examined as a systematization process as shown in figure 3

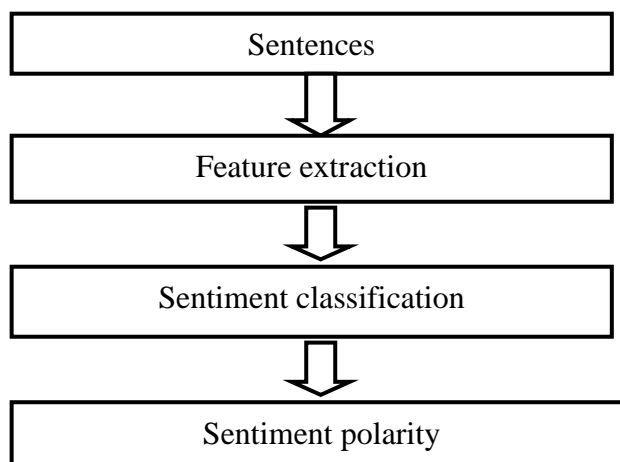


Figure 3: Process of sentiment analysis [7]

1.7 Components of sentiment analysis: The main components of opinion mining or sentiment analysis are as follows:

- **Sentiment holder:** It is the individual that gives a specific opinion on an object. It might any association that is giving data or view point about something [3].
- **Sentiment object:** It is a thing on which an opinion or feeling is expressed by user [3].
- **Sentiment orientation:** It is a view or sentiment of an object done by user. It might be positive, negative or neutral [3]. Figure 4 shows the components of sentiments analysis.
- Components of sentiment analysis e. g. Modi says that the India is a great country

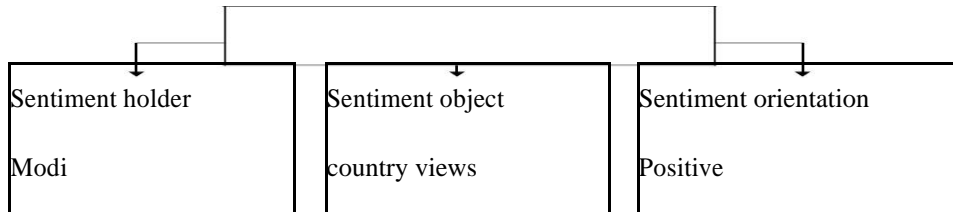


Figure 4: Components of sentiment analysis [3]

1.8 Levels of sentiment analysis: There are three levels of sentiment analysis:

- **Document level:** The whole record or document is considered for slant investigation. The sentiment about the entire record is recognized whether it is positive, negative or neutral [3, 6, 14].
- **Sentence level:** Each sentence is independently regarded as positive, negative or neutral [3, 14].
- **Feature level:** It is called viewpoint level characterization. This level manages specific highlights [3].

1.9 Classification of sentiment analysis: Sentiment analysis basically ordered into 3 categories which are as given below:

- **Positive sentiment:** It is the gathering of good or positive words in the supposition [6, 9, 17].
- **Negative sentiment:** If the negative words are available in the survey then the audit is called negative opinion. These are also called bad words [6, 9, 17].
- **Neutral sentiment:** If the tweet is neither considered as negative nor positive tweet then it is called neutral opinion [6, 9, 17]. Figure 5 shows the positive, negative and neutral sentiments.

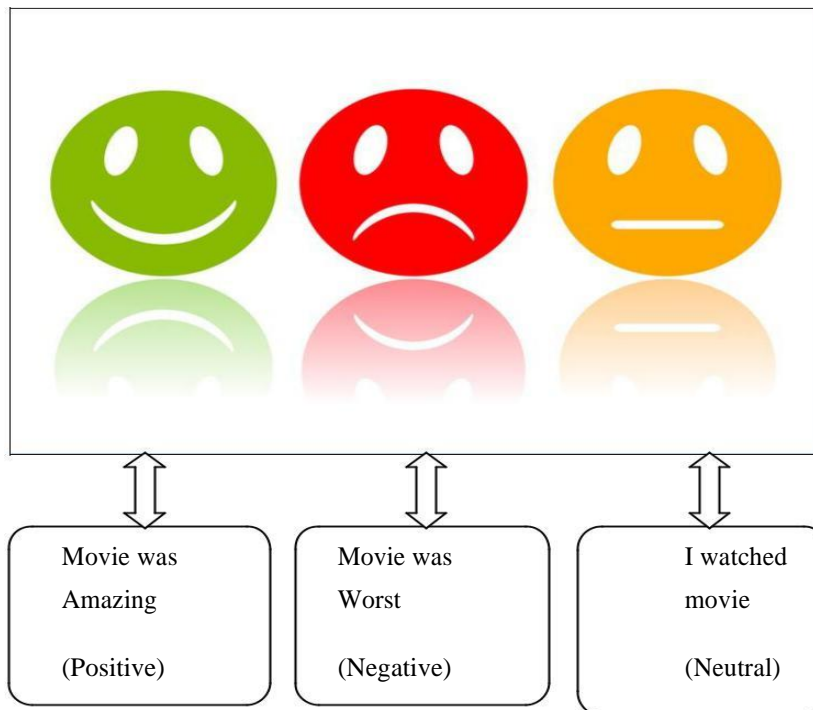


Figure 5: Positive, negative and neutral sentiments [19]

1.10 Facebook: Social media is a platform that allows people to share their thoughts or opinions. Different social media platform have different methods of content sharing, ex. Facebook [13]. Facebook is a most popular social media website [20]. Facebook is framework, online news and person to person communication where clients post and interface with messages, known as "comments." These messages were initially confined to 140 characters [1]. Registered clients can post comments, however the individuals who are unregistered can just read them. Clients get to facebook through its site interface, Short Message Service (SMS) or cell phone application programming ("app"). Facebook [19].

1.11 Comments: Comments are messages. Clients can comment by means of the Facebook site, perfect outer applications for example, for cell phones or by short message service accessible in certain countries. Users may obtain in to other clients comments and postd this is known as "following" and supporters are known as "followers" . Individual comment can be sent by different clients to their own encourage, a procedure known as a sharing.[15]

2. REVIEW OF LITERATURE

2.1 Overview: Data mining techniques offer a standard & great tool set to produce numerous data focused organization systems. This review of literature emphasis on how data mining methods are used for discovers significant arrangement from the database.

2.2 Related work: During the age of time, reading certain of the research papers has been done which is summarized as below:-

Kumar et al. (2019).[1] “**Sentiment Analysis of Electronic Product Tweets Using Big Data Framework**” discussed about the different sale tweets used to examine the sentiments of customers regarding electronic goods. The experimental results of the proposed work will be useful for various business companies to take business verdicts, which will further improve the product sales. In the current scenario, millions of tweets are produced by people every year. But handling these huge unstructured tweets is not possible through the traditional platform.

Gupta et al. (2017).[2] “**Sentiment Analysis of Twitter and Facebook Data Using Map-Reduce**” discussed about Twitter and Facebook’s amusing source of data for opinion mining or sentiment analysis and this vast data can be used to find the sentiments of people on a specified topic or product. In this paper, system is proposed which involves collecting data from social network using the Twitter and Facebook API’s.

Mathapati et al. (2016).[3] “**Sentiment Analysis and Opinion Mining from Social Media : A Review**” discussed about the need for automated analysis techniques to extract sentiments and opinions sent in the user-comments. words provide fine-grained analysis on the customer reviews.This paper focuses on the survey of the existing methods of Sentiment analysis and Opinion mining techniques from social media.

Rastogi et al. (2014).[4] “**A Sentiment Analysis based Approach to Facebook User Recommendation**” discussed about system to offer new friends who have similar interests but having different opinions. The motivation of this work is that users may share similar interests but have dissimilar opinions on them. In this paper, a user recommendation technique based on a novel weighting function is proposed, which consider not only user interests, but also his sentiments.

Gupta et al. (2017).[5] “**Sentiment Analysis of Twitter and Facebook Data Using Map-Reduce**” discussed about Twitter and Facebook’s amusing source of data for opinion mining or sentiment analysis and this vast data can be used to find the sentiments of people on a specified topic or product. In this paper, system is proposed which involves collecting data from social network using the Twitter and Facebook API’s. Then, the challenges of big data are answered using Hadoop through map reduce framework where the complete data is mapped and reduced to smaller sizable data to ease of handling and finally contains analyzing the collected data and represent the results through graphs.

Isah et al. [6] “**Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis**” discussed about user created content from social media platforms that can provide early clues about product allergies, adverse events and product counterfeiting. This paper accounts a work in progress by means of contributions including: the growth of a framework for assembling and analyzing the views and experiences of users of drug and cosmetic goods using machine learning, text mining and sentiment analysis

Gürsoy et al. (2017)[7]. “**Social Media Mining and Sentiment Analysis for Brand Management**” discussed that Corporate companies want to gain from big data studies extra. Although it affects different company dynamics in various areas, especially social media services have become very significant for the marketing and CRM departments of businesses. In this way, communication is always recognized with the customers and the use of Big data in these fields is seen as one of the utmost important steps of the firms in becoming a big trademark.

Patil and Thakare (2017)[8]. “**Analyzing Public Sentiment Variations on Twitter and Facebook**” discussed about interchange of views, ideas, expressions, feelings and opinions on social networking sites like Twitter and Facebook. In this work, analyzation of public sentiment variations in an explicit time period about a explicit target on Twitter and Facebook both is done. This kind of analysis is helpful in various fields for taking proper conclusions and deteriorating public opinion.

Salloum et al. (2017)[9]. “**A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives**” discussed about a mutual practice to not write a sentence with correct grammar and spelling at social networking sites which leads to diverse kinds of uncertainties like lexical, syntactic, and semantic and due to this type of uncertain data, it is inflexible to find out the actual data order. This study aims to describe how studies in social media have used text analytics and text mining methods for the purpose of categorizing the key themes in the information.

3. RESEARCH GAPS AND OBJECTIVES

Today is the universe of innovation. For the most part the work is finished utilizing the web. Web is the new reason for the beginning of learning, shopping and training. With a specific end goal to gather and investigate the information from the online sites a system is utilized which is known as sentiment mining. It is otherwise called notion examination or sentiment analysis. It is utilized to gather the client audits from the place and break down the sentiment of open whether it is positive or negative.

Numerous calculations are accessible to manage slant examination. It should be possible to discover the sentiment of open towards the new cell phones, motion picture evaluations, current issues and some more. Thus it is up and coming field that discovers the individuality of open towards any point. People write their comments frequently & in shortcuts manner, so it is not possible to judging the comments which are positive and which are negative & neutral. To know the views of people in right manner this is the need of today.

3.1 Problem formulation: Sentiment analysis can be seen as a utilization of content order. The primary occupation of content gathering is how to stamp writings with a predefined set of gatherings.

Analyze every last tweet at that point choosing whether it is sure positive or negative is not a simple. A calculation for assumption investigation should be executed to get powerful precision of general feeling. People use very awkward words to express their feelings & most of the people use shortcuts e.g. osm for awesome, lol for laughing out loud & many more, so this is sometime creating difficulty for the person who is not familiar with these words and cannot recognize the sentiments of the person.

3.2 Research objectives

1. To review & explore sentiments of users in comments.
2. To create a comment's database.
3. To preprocess (slang words) & mine the collected data.
4. To build up a proficient calculation for felling investigation.
5. To predict the sentiment of comments.

4. METHODOLOGY

4.1 Existing methods

Sentence level classification is used to analyze the comments. For the purposes of the research, it defines sentiment to be "a personal positive or negative feeling." Data collection, There is no current data indexes of Facebook assumption messages. It gathered its own set of data. The test information was manually. An arrangement of 98 negative comments and 78 positive comments is manually checked. A web interface instrument was worked to help in the manual arrangement undertaking. Following are some existing techniques:

1. **SVM:** Support vector machine learning algorithm used basically for classification and regression problems. It solves an optimization problem of finding the maximum margin hyperplane between the classes. Hyperplane used for classifying the linear and non linear data.
2. **KNN:** K-nearest neighbors is a learning algorithm based on the classes nearest to the point which is to be classified. Based on the values of the K nearest classes a test set is provided the majority voting class.

4.2 Algorithm

1. The tables of database are created; it contains the positive & negative words.
2. The comments will be scored with some numbered values i.e. 1 for positive comment, -1 for negative comments & 0 for neutral comments.
3. Data filtering will be performing to remove the unnecessary data from comments e.g. URLs, usernames, duplicate & repeated character.
4. The slang words (e.g. lol means laughter out loud) will be changed into actual words.
5. The words with negation (never, not, nor etc) will be handle.
6. The single comment will perform the words which will analyze & compare with the database.
7. Sentiments will be shown graphically. Figure 6 shows the functionality of a comment.

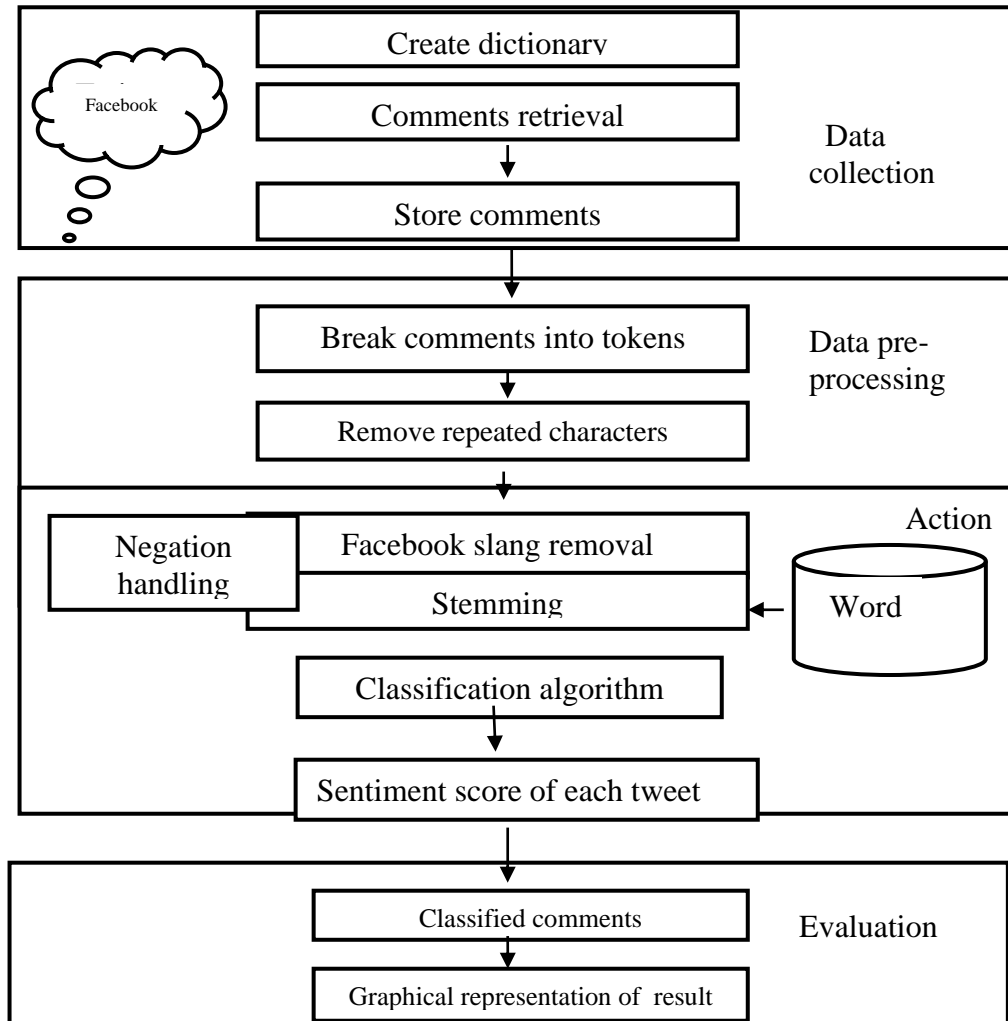


Figure 6: Functionality of a comment

The complete detail of the steps is given in following steps:

4.2.1 Create Dictionary: Make a dictionary of the positive and negative words. Two different tables are created in the sentiment database one for positive words and other for negative words. Firstly made a dictionary of positive and negative words.

Table 1.1: Database table

Table Name	Field Name	Data Type
Negwords	Nwords	Varchar
Poswords	Pwords	Varchar
Tweets Database	Tweet	Varchar
	Sentiment	Int

4.2.2 Comments Collection: The tweets are collected from the twitter. Firstly one have to create a facebook account then login to that account to collect the comments. SQL database is used to store the comments. Wwww.sentiment140.com website is used to collect the tweets. Manually assign the sentiment to each tweet i.e. 0 to neutral tweet, 1 to positive tweet and -1 to negative tweet.

Table 1.2: Demonetization sentiment score database table

Sentiment Source	Tweet	Sentiment Score
Sentiment 140	Scary that we are not yet out of the thoughtless decisions and poor execution #gst #demonetization	-1
Sentiment 140	And someone says #Demonetization wasn't a good move by Modi! I will repeat it was the best step taken by Modi Government so far!	1
Sentiment 140	Demonetization Happened in India.	0

If Comment is positive, then Assign Sentiment Score=1

Comment is Negative, then Assign Sentiment Score=-1

Comment is Neutral, then Assign Sentiment Score=0

4.2.3 Data Pre-Processing: The Preprocessing is done on the retrieved tweets.

4.2.4 Filtering: Filtering helps to create a single data structure that is used by the user for creating single mining method. It helps to use only single or some specific part of document not the whole document. Hence, it reduces the load to carry the whole data. Filters can be used in many ways. Some of them which are used are as follows:

4.2.5 Urls: The comments collected from the comments. contain some links or urls which are not used in estimating the sentiment of the comments. These links does not have any link with actual sentiment. So, these links are replaced by the empty space.

4.2.6 Usernames: Sometimes user in tweets refers to other users so they refer to them by using @ symbol before their name. These names also do not affect the sentiment so replaced by empty space.

4.2.7 Duplicate or Repeated characters: Users sometimes use casual language in tweets. For example, users mostly write 'baaaaaaad' in place of bad word. But actually this the same word bad. Sometimes they write 'happpppppppy' instead of happy. Hence happppppy is replaced by happy. Here, urls and Usernames are replaced by empty space to decrease the complexity and time taken by the algorithm to compare each word with database.

Comments Having	Replaced By
Https://t.co/Htxxx	Empty Space
@ravneet	Empty Space
@gurinder	Empty space
Happpppppy	Happy
Gooooooood	Good

Table 1.3: Data filtering

4.2.8 Facebook slang removal: There is less space offered for writing a comment on Facebook as comment is only of 140 characters. Hence, most of the users prefer to write short form of the actual words. The user created short form is called as slang words. Sometimes public also use some abbreviations. For example, tmrw is used in place of tomorrow, thx in place of thanks. These slang words should be replaced into their original words. For this a different table is created in dictionary that stores the slang words.

Table 1.4: Slang removal

Facebook Slang	Actual Word
Gud	Good
Awsm	Awesome
Fav	Favorite
Thnx	Thanks
Bff	Best friends for ever
Tc	Take care
Sd	Sweet dreams

4.2.9 Stop words removal: Stop words are the words which are mainly used in tweets or comments but this does not add to sentiment. Stop words are articles, prepositions etc. These should be removed from the document and replaced by the empty space.

4.2.10 Negation Handling: There are some words which change the meaning of sentence these words are known as negation words. Words like never, not, does not, no, nor are the negation words. If the tweet is positive these words change the sentiment of tweet to negative. So these are handled with proper method. There are two cases in negation, which are as follows:

4.2.10.1 Negation word used with positive word and it make it negative: In this, if the whole sentiment of sentence is positive, but the positive word preceded by negation then the sentiment of sentence is changed to negative.

"Story of serial is good": This sentence gives the positive sentiment as the positive word good is present here. Now consider the case:

"Story of serial is not good": This sentence has negation word 'not', which changes the sentiment of sentence to negative sentence.

4.2.10.2 Negation word used with negative word and make it positive: In this, if the whole sentiment of sentence is negative, but the negative word preceded by negation then the sentiment of sentence is changed to positive.

"Story of serial is bad": This sentence gives the negative sentiment as the negative word bad is present here. Now consider the case:

"Story of serial is not bad": This sentence has negation word 'not', which changes the sentiment of sentence to positive sentence.

4.2.11 Stemming: It is the process to convert the words into their original form. Sometimes users use the stemmed words for the original words which should be replaced by actual words. For example, hate, hated, hates, hating all belong to the single word hate. It will increase the efficiency of the software.

Table 1.5: Stemming

Original word	Stemmed word
Damaged	Damage
Damages	Damage
Damaging	Damage

4.2.12 Example for Pre-processing of tweets: Following table shows the complete pre-processing of a tweet and its output.

Table 1.6: Example for tweets pre-processing

Actual Tweet	@ravneet And someone says #Demonetization wasn't a good move by Modi! I will repeat it was the best step taken by Modi Government so far! Happppy. Lol!Checkout https://www.seerha.com
Change to Lowercase	@ravneet and someone says #demonetization wasn't a good move by Modi! I will repeat it was the best step taken by modi government so far!Happppy. Lol!Checkout https://www.seerha.com
Remove special characters	@ravneet and someone says demonetization wasn't a good move by Modi! I will repeat it was the best step taken by modi government so farhappppy lol checkout https://www.seerha.com
Remove Usernames	And someone says demonetization wasn't a good move by Modi! I will repeat it was the best step taken by modi government so farhappppy lol checkout https://www.seerha.com
Remove urls	And someone says demonetization wasn't a good move by Modi I will repeat it was the best step taken by modi government so farhappppy lol!Checkout
Remove extra space	And someone says demonetization wasn't a good move by Modi I will repeat it was the best step taken by modi government so farhappppy lol checkout
Remove more than 2 repeated characters	And someone says demonetization was not a good move by modi i will repeat it was the best step taken by modi government so far happylol checkout
Remove slang word	And someone says demonetization wasn't a good move by Modi I will repeat it was the best step taken by modi Government so farhappy laugh out loud checkout
Stop words removal	And says demonetization was not good move by modi will repeat was best step taken by modi government so far happy laugh out loud checkout

4.2.13 Calculating Sentiment Score: Sentiment score is calculated by comparing the words from the tweets with the dictionary words. If the tweet contains more positive words than negative then the tweet is treated as positive.

4.3 Techniques used

1. **Sentiment analysis:** Sentiment analysis can be done through 2 types of procedures as below:

1.1 **Sentiment arrangement utilizing regulated learning:** Supervised learning is actualized by making a classifier. It requires two arrangements of reports for order one is preparing set other is trying set. This strategy is otherwise called machine learning technique.

1.2 **Sentiment arrangement utilizing unsupervised learning:** In the unsupervised order the content is characterized by contrasting it and given words or dictionaries. The feeling an incentive for these words or dictionaries is already characterized. The report is checked and contrasted and positive and negative words.

1.3 **Classification:** Classification assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.

4.4 Parameters

1. Accuracy
2. Time
3. Predictor
4. Automation

5. IMPLEMENTATION

For implementation we have used JAVA language. JAVA is high level object oriented programming language. Netbeans IDE is used as front end. SQL is used as the database to store the comments and the dictionary words. Comments are collected from various fields such as Kesari movie, Bollywood in Politics, Education system in India and Punjab Government.

5.1. Netbeans IDE Interface

Netbeans IDE is a user friendly interface to develop JAVA codes. It provides easy way to create the front end and a proper error handling mechanism. By Netbeans JAVA users got a simple drag and drop system to use any of its tools. To run the project click the green run arrow button on the menu bar.

5.2 Main Window

It consists of 2 buttons and 1 combo box. Combo Box consists of the list of the topics for sentiment analysis. "Check Sentiment" button is used to run the algorithm on the selected dataset. "Clear" button clear all the values of the labels and the variables used in the program. Calculated result field shows the calculated values on the chosen dataset. Accuracy shows the truthfulness of the given algorithm. Actual result field shows the no of actual positive, negative or neutral tweets in the database. Choose the list item to choose the database for which one wants to apply sentiment analysis.

5.4 **Dictionary Creation:** Dictionary for negative and positive words is created separately using two different tables in SQL.

5.4.1 **Positive Words Dictionary:** The list of positive words that are stored in table.

5.4.2 **Negative words Dictionary:** The list of negative words that are stored in table.

5.5 Slang words table: Sometimes people use their own abbreviations to represent any word. These abbreviations are called slang words.

5.6 Stop words table: These are the words that are contained by the tweets but these do not affect the sentiment of the tweets. So these should be removed to save the time of algorithm.

5.7 Comments dataset: To check the accuracy of the algorithm 4 datasets are created collecting the comments. The comments are collected for following topics:

- 1) Kesari movie comments
- 2) Bollywood in Politics comments
- 3) Education System in India comments
- 4) Punjab Government comments

5.8 Summary

In this chapter, screenshot of the thesis implementation are properly explained. Various types of tables that are used in sentiment analysis are also shown. Screenshots of datasets are also explained.

6. RESULT AND DISCUSSIONS

The main motive of the research is to develop this algorithm that easily calculates the sentiment of the tweets collected from the Twitter. Algorithm is applied on the tweets that are collected for a single day. The efficiency of algorithm is measured in terms of accuracy rate which is near about 85 %.

6.1. Results for Kesari movie dataset: Total 24 comments are collected. The Algorithm is applied on them. The software calculated the sentiment with the efficiency of 83.33%. Overall sentiment of comments shows that the opinion of the public towards the Kesari movie is positive. 2 tweets from the total tweets are calculated with wrong sentiment.

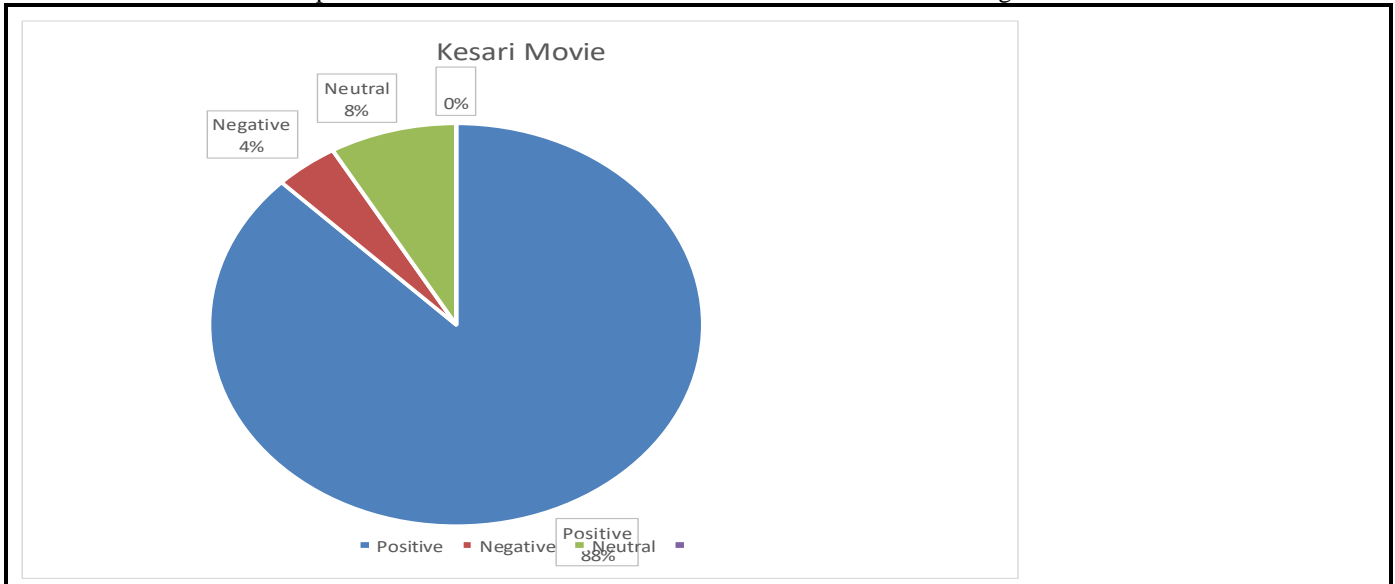


Figure: 7 pie chart for "Kesari movie" comments

6.2. Results for Bollywood in Politics dataset

Total 20 comments are collected. The software calculates the sentiment with efficiency of 60.00%. Figure 7 shows the overall sentiment of the Bollywood in Politics. Result show that public opinion towards this is positive. The system retrieves 5 as positive comments, 4 as negative comments and 11 as neutral comments. Only 11 comments are analyzed wrong. Lesser the amount of wrong comments analyzed more will the accuracy of the system.

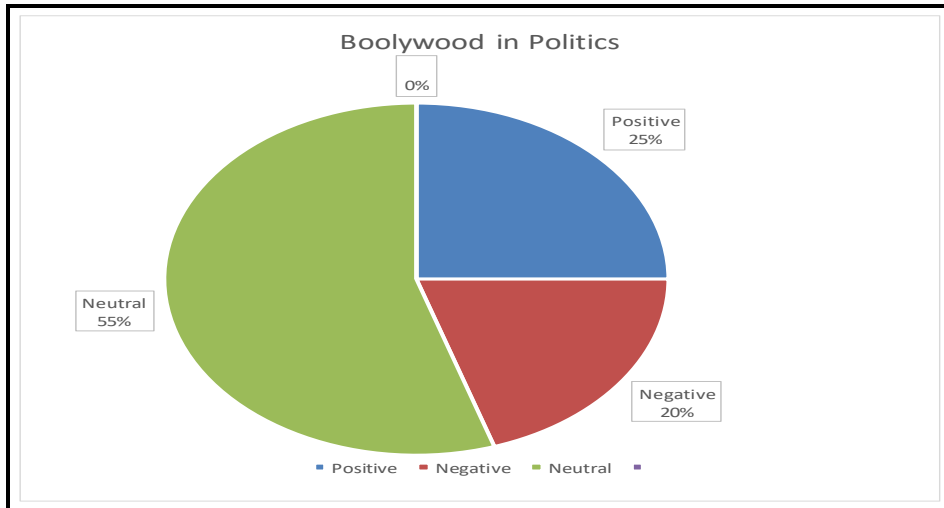


Figure: 8 pie chart for “Bollywood in Politics” comments

6.3 Results for Education System in India Dataset

Total 19 comments are collected and the software calculates the sentiment with efficiency of 63.16%. The results that the sentiment of people towards Education System in India is positive. 6 comments are analyzed with wrong sentiment. After the results retrieved positive tweets are 7, retrieved negative tweets are 6 and retrieved neutral tweets re 6. Lesser the no. of wrong comments analyzed more will be the accuracy of the system. Blue part represents the positive comments, Red part represents the negative comments and green part represents neutral comments.

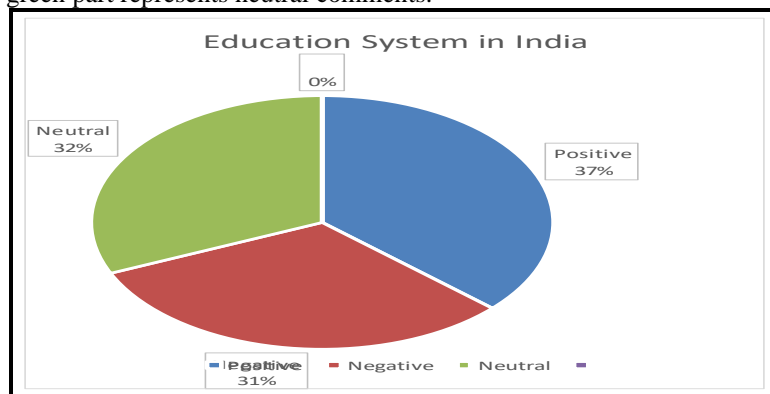


Figure: 9 pie chart for “Education System in India” comments

6.4 Results for Punjab Government comments dataset

Total 15 comments are collected. The Algorithm is applied on them. The software calculated the sentiment with the efficiency of 100%. It is clear from the Figure 9 that overall sentiment of tweets is negative. 2 comments are analyzed with wrong sentiment. Retrieved positive comments are 3, retrieved negative comments are 10 and neutral comments are 2.

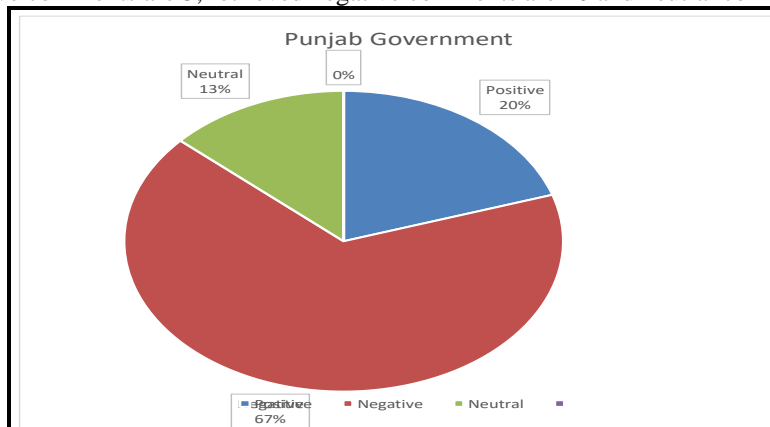


Figure: 10 pie chart for “Punjab Government” comments

6.5 Accuracy comparison of different datasets

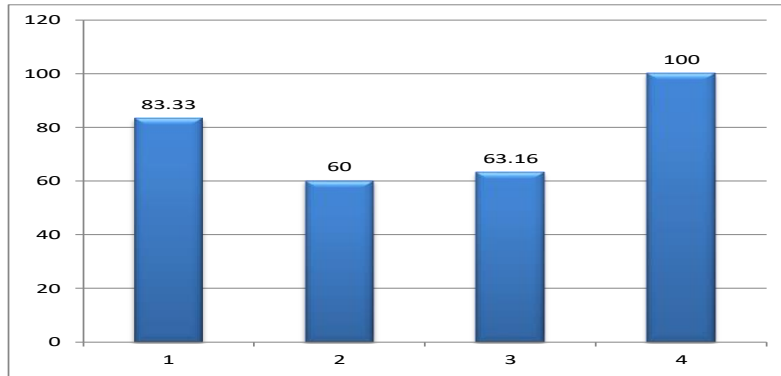


Figure: 11 bar chart showing accuracy of different datasets

6.6 Detail of 6 dataset results

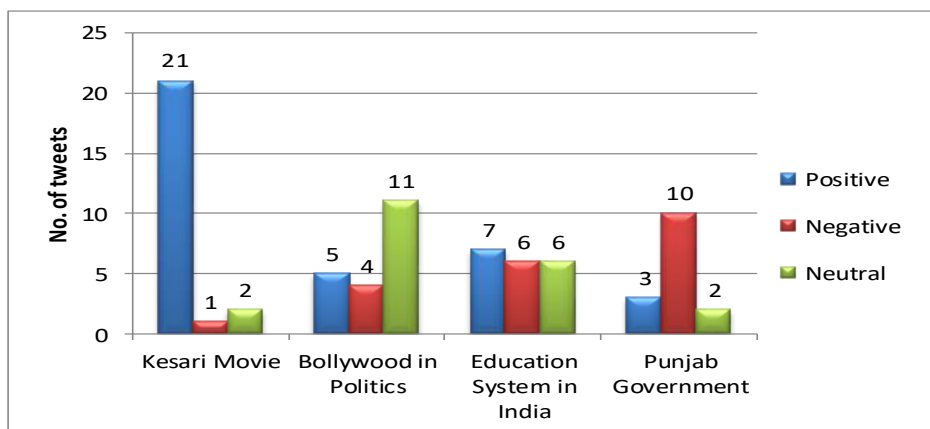


Figure: 12 Graphical representation of Results

6.7 Summary

In this chapter, output of the sentiment analysis algorithm is shown. The comments analysis based on the different datasets is graphically represented in the form of pie charts or histograms. The comparison of accuracy of different datasets is shown in table form.

7. CONCLUSION

Sentiment analysis is the emerging field that is mainly used in many application areas. Its scope is increasing. So a need arises to create or develop an algorithm that could properly find the sentiment of the public tweets or opinion. This work shows a new algorithm that is developed in Java language. The algorithm is applied on comments and efficiency is calculated based on the accuracy rate of the algorithm. The approximate efficiency of the algorithm is 86%.

7.1 Challenges

- Detection of spam comments.
- Recognize the fake comments.
- Recognize the co-reference between nouns and pronouns.

8. SCOPE FOR FURTHER RESEARCH

The accuracy of algorithm can be checked by taking the comments from other websites. Evaluation of two or more products or brands is also done for better performance. A rich lexicon dictionary is created for enhanced processing of the algorithm. Sentiment analysis can be applied to further more datasets for better analysis. The work can be extended by collecting the comments from different blogs and sites and apply different types of classifiers on the dataset and their accuracy can be compared to know which classifier is helpful for achieving better efficiency.

REFERENCES

- [1] Gupta, J. Pruthi, N. Sahu (2017), "Sentiment analysis of tweets using machine learning approach", International Journal of Computer Science and Mobile Computing, 6(4), pp. 444-458.
- [2] P. Jain, M. V. D. Katkar (2015), "Sentiments analysis of Twitter data using data mining", International Conference on Information Processing, 10, pp.807-810.

- [3] P. Rajan, S. P. Victor (2014), "Web sentiment analysis for scoring positive or negative words using tweeter data", International Journal of Computer Applications, 96(6), pp. 33-37.
- [4] Alvares, N. Thakur, S. Patil, D. Fernandes, K. Jain (2016), "Sentiment analysis using opinion mining", International Journal of Engineering Research & Technology, 5(4), pp.88-91.
- [5] M. Bandgar, D. S. Sheeja (2016), "Analysis of real time social tweets for opinion mining", International Journal of Applied Engineering Research, 11(2), pp. 1404-1407.
- [6] S. Dattu, P. Deipali, V.Gore (2015), "A survey on sentiment analysis on Twitter data using different techniques", International Journal of Computer Science and Technologies, 6(6), pp.5358-5362.
- [7] Chandni, N. Chndra, S. Gupta, R. Pahade (2015), "Sentiment analysis and its challenges", International Journal of Engineering Research & Technology, 4(3), pp.968-970.
- [8] E. Oleary (2015), "Twitter mining for discovery, prediction and causality : applications and methodologies", International Journal of Intelligent Systems in Accounting and Finance Management, 22(3), pp.222-247.
- [9] G. Hu, P. Bhargava, S. Fuhrmann, S. Ellinger, N. Spasojevic (2017), "Analyzing users' sentiment towards popular consumer industries and brands on twitter", International Conference on Data Mining Workshops, 381-388.
- [10] G. Sabarmathi, D. R. Chinnaiyan (2017), "Reliable data mining tasks and techniques for industrial applications", IAETSD Journal for Advanced Research in Applied Sciences, 4(7), pp. 138-142.
- [11] H. P. Rahmath (2014), "Opinion mining and sentiment analysis- challenges and applications", International Journal of Application or Innovation in Engineering & Management, 3(5), pp.401-403.
- [12] Smeureanu, C. Bucur (2012), "Applying supervised opinion mining techniques on online user reviews", Informatica Economică, 16(2), pp. 81-91.
- [13] Umar, F. Chiroma (2016), "Data mining for social media analysis: using twitter to predict the 2016 Us presidential election", International Journal of Scientific & Engineering Research, 7(10), pp.1972-1980.
- [14] Sutar, S. Kasab, S. Kindare, P. Dhule (2016), "Sentiment analysis: opinion mining of positive, negative or neutral twitter data using hadoop", International Journal of Computer Science and Network, 5(1), pp. 177-180.
- [15] J. Sheela (2016), "A review of sentiment analysis in twitter data using hadoop", International Journal of Database Theory and Application, 9(1), pp.77-86.