GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
SENTIMENT ANALYSIS OF REAL TIME SOCIAL TWEETS

Nancy Verma[*1] & Dr. Gaurav Gupta[2]
[*1]Research Scholar: Computer Engineering, University College of Engineering, Punjabi University Campus, Patiala
[2]Assistant Professor: Computer Engineering, University College of Engineering, Punjabi University Campus, Patiala

---

ABSTRACT

Social media like Twitter, facebook serves as a unified platform for users to express their thoughts on subjects ranging from their daily lives to their opinion on entertainment, social topic or customer products. In this work, we use an Opinion mining techniques for the purpose to perform sentiment analysis on the views people have shared in Twitter. We collect dataset, i.e. the tweets from twitter that are in natural language and applied opinion mining on the collected datasets for calculating accurate efficiency.

Keywords: Tweets, Sentiment Analysis, Opinion Minning, etc.

I. INTRODUCTION

Data Mining is the investigation periods of the "data discovery in documents" a method for deciding plans in enormous information accumulations including approaches at the association of reenacted knowledge, machine learning and record frameworks. The total goal of the data mining methodology is to mine data from information accumulations and change over it into a sensible setup for extra utilize. Data mining is a predominant new ability with extreme planned to help enterprises accentuation on the more fundamental material in their information stores. Data mining devices estimate up and coming inclinations and exhibitions, allowing enterprises to make data centered judgments. Data mining apparatuses would reply expert interrogations be able to that for the most part were excessively time killing, making it impossible to choose. They wash records for concealed examples, disclosure investigative data that authorities may preclude as it trickeries outside to their expectations.

II. PROCESS OF DATA MINING

Data mining includes of various stages. Data mining is an imperative stage in the technique of data finding. Following are the rundown of stages in the data recognition process: Set all these values using the "FILE" Menu. Select the Page **Data Integration:** Firstly every one of the information is made and consolidated from all the distinctive sources.

Data Selection: As every one of the information gathered by the client isn't completely required. Here we pick the information which we consider advantageous for data mining.

Data Cleaning: The information we have warehoused isn't spotless. This may contain blunders, lost esteems or dishonest information. So we need to put on various techniques to get free of such irregularities.

Data Transformation: Modification of the information into the shape that is required for mining operations is called information change.

Data Mining: Is comprises of different procedures that can be utilized to discover different in secret arrangements or likenesses in the given dataset.

212

Pattern Evaluation and Knowledge Presentation: This progression incorporates taking out or expelling the copy designs from the examples we created.

Decisions/Use of Discovered Knowledge: This progression causes the client to settle on choices on the data that is gathered.
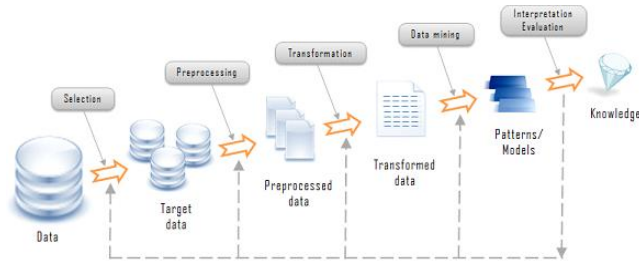

FIGURE 1.1: PROCESS OF DATA MINING

Data mining hierarchical model

Various capable ways are existing to store the gigantic volumes of information, computational procedures and models are required to separate the concealed examples and learning. These strategies and instruments are utilized to change the information into helpful data, to make advertise investigation, misrepresentation discovery and discover the client expectations and so forth. These methods are all things considered known as the Data Mining or at times perceived as Knowledge Discovery in Databases. An entire progressive model for information mining is appeared in Fig 1.2.
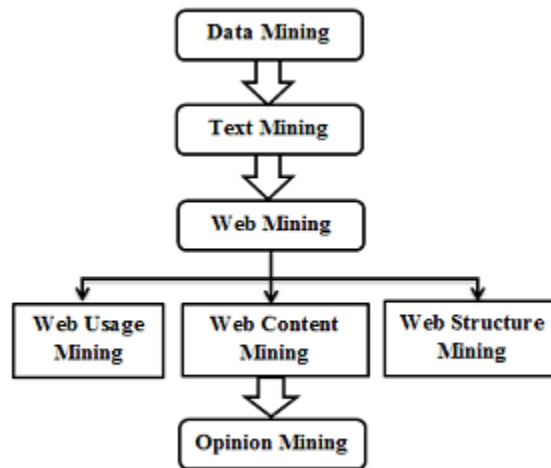

Figure 1.2: Data Mining Hierarchical Model

The Text mining is a ground that is utilized to distinguish the advantageous data in the literary archives or records. The content can be in any shape or in any dialect that can be English, Punjabi, Hindi and numerous others. Web mining is the technique to gather the useful information from the sites or online audits. It is difficult to gather or investigate the online data in light of the fact that a lot of data is accessible online to manage. Web mining is isolated into 3 sub parts. Web use mining is procedure to discover the use of any sites i.e. how as often as possible the clients utilize some specific site. Web structure mining is the strategy to discover the general structure of the online

destinations or web journals. Web content mining is the regularly utilized territory these days. It 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, audits of various clients. Web content mining additionally ordered into Opinion mining or slant examination. Opinion mining is the further advance in the Web content mining. The distinction between these two is that web content mining just gathers the information from the web destinations while the sentiment mining discover the point of view of open towards a particular subject or region.

III.   INTRODUCTION TO SENTIMENT ANALYSIS

Sentiment Analysis or Opinion mining is the method for finding or hauling out the sentiments and feelings of people to correct regions of consideration. It might be a thing or a film, surveys of people truly matters. These surveys additionally influence some other individual's approach making process. In the event that a purchaser desires to acquire another question, at first he would get a handle on the assessments or remarks of different people. Contingent on the extremity of surveys he chooses whether to purchase the item or not. Social collaborating sites, for example, Facebook, twitter are where characters put their status or sentiments. People tweet on their twitter account concerning any correct subject of their consideration. Conclusion examination is utilized to conjecture the share trading system, to anticipate the aftereffect of specific surveys, to distinguish the adequacy of any item or in some more.

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 nonpartisan. For instance, the tweet "I am exceptionally cheerful today since I bested in my interview" is a positive content and the content "I loathe this" is a negative content. Consider another case "robot is great film I recommend everyone to watch this motion picture", plainly client survey is absolutely positive towards the motion picture robot. Incidentally it is difficult to decipher whether the tweet is certain or negative, at that point we call the tweet as impartial. "Robot isn't terrible however I don't comprehend why individuals put it as number one film" these sorts of tweets considered as impartial. The tweets given above are about the specific theme which a motion picture is named Robot.

Twitter is a predominantly as often as possible utilized long range interpersonal communication site that gives its clients to refresh a 140 characters status. It stores a colossal measure of informational collection about the particular theme. WWW i.e. Internet made it less demanding for individuals to share their thoughts over the web. Assessment investigation basically makes utilization of common dialect taking care of and content preparing to finish the whole undertaking i.e. to recognize the supposition of the general population. For instance, in the event that one needs to know - if the elections of Punjab are doing the activity legitimately or not? The greatest technique to answer this is seeing any interpersonal interaction site. It is anything but difficult to get some answers concerning the work done by Punjab election by survey the tweets of client. In any case, the issue is that there countless how we perceive that what numbers of individuals are sure or negative towards the Punjab election. The overwhelming plausible answer is to utilize estimation investigation on the tweets and discover what individuals say in regards to Punjab election.

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

Sentiment Holder: It is the individual who is giving the conclusion about some subject. It might any association that is giving data or view point about something. In online audits conclusion holder is the individual which is composing the surveys, remarks.

Sentiment Object: It is the thing about which the supposition is given by some assessment or feeling holder.

Sentiment Orientation: It is the grouping or notion examination of the slant. It might be certain, negative or nonpartisan relying on the information in the supposition.

Levels of sentiment analysis
The procedure of assumption investigation should be possible in mostly 3 levels:

Document Level: The whole record or document is considered for slant investigation. The sentiment about the entire record is recognized whether it is certain, negative or impartial.

Sentence Level: Each sentence is independently regarded and delegated positive, negative or impartial.

Feature Level: It is otherwise called viewpoint level characterization. In this the supposition is improved the situation the some particular highlights from the record. This level manages specific highlights.

Classification of sentiment analysis
Up position investigation basically ordered into 3 categories which are as given beneath:

Positive Sentiment: It is the gathering of good or positive words in the supposition. On the off chance that the amount of good contentions more noteworthy than before it is referenced as a Positive assessment. For instance, if audits of an item have more positive remarks then it is certain to be purchased by numerous clients.

Negative Sentiment: If the negative words are available in the survey then the audit is called negative opinion. For instance, if the aggregate audits or tweets about any item have more adverse surveys then the item isn't so helpful then it is purchased by less number of individuals.

Neutral Sentiment: If the tweet is neither considered as negative nor positive tweet then it is dealt with as impartial feeling in the slant investigation process.

The supposition "Robot- The motion picture was great" contains a positive word marvelous so it is sure. "I watched this film" is a nonpartisan opinion and "This was the most exceedingly terrible motion picture ever" contain the negative word most exceedingly terrible, so it is negative notion as appeared in figure 1.3.

Movie was amazing   Movie was Worst   I watched movie
(POSITIVE)   (NEGATIVE)   (NEUTRAL)
Figure 1.3: Positive, Neutral and Negative sentiments

IV.   LITERATURE SURVEY

Data mining techniques offer a standard & great tool set to produce numerous data focused organization systems. This review of literature emphases on how data mining methods are used for different use regions for discovery out significant arrangement from the database.

Guoning Hu, PreetiBhargava, Saul Fuhrmann, Sarah Ellinger and NemanjaSpasojevic[1] Analyzing users' sentiment towards popular consumer industries and brands on Twitter, Online networking fills in as a brought together stage for clients to express their considerations on subjects running from their everyday lives to their conclusion on shopper brands and items. These clients use a huge impact in molding the suppositions of different customers also, impact mark observation, mark steadfastness and mark support. In this paper, we dissect the supposition of 19M Twitter clients towards 62 well known ventures, enveloping 12,898 undertaking and customer brands, as well as related topic subjects, by means of estimation examination of 330M tweets over a period crossing a month. We observe that clients have a tendency to be best towards fabricating and most negative towards benefit ventures. Furthermore, they have a tendency to be more positive or negative while collaborating with brands than by and large on Twitter. We likewise find that notion towards brands inside an industry changes enormously and we illustrate this utilizing two enterprises as utilize cases. What's more, we find that there is no solid relationship between theme estimations of various enterprises, illustrating that theme feelings are profoundly reliant on the setting of the business that they are specified in. We exhibit the estimation of such an investigation all together to evaluate the effect of brands via web-based networking media. We trust that this underlying examination will demonstrate profitable for both analysts and organizations in understanding clients' recognition of businesses, marks and related points and energize more research in this field.

Ankita Gupta, JyotikaPruthi , NehaSahu[2] Sentiment Analysis of Tweets using Machine Learning Approach , Slant Analysis goes under investigation inside Natural Language preparing. It helps in finding the conclusion or sentiment covered up inside content. This exploration concentrates on discovering conclusions for twitter information as it is all the more difficult because of its unstructured nature, constrained size, and utilization of slangs, incorrectly spells, shortened forms and so forth. The majority of the scientists managed different machine learning methodologies of slant examination and think about their results yet utilizing different machine learning approaches in mix have been underexplored in the writing. This exploration has discovered that different machine learning approaches in a half and half way gives better outcome when contrasted with utilizing these methodologies in disconnection. Besides as the tweets are exceptionally crude in nature, this examination makes utilization of different preprocessing steps so we get helpful information for contribution to machine learning classifiers. This examination essentially concentrates on two machine learning calculations K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) in a half and half way. The expository perception is acquired as far as order exactness and F-measure for every assumption class and their normal. The assessment investigation demonstrates that the proposed crossover approach is better both regarding exactness and F-measure when contrasted with singular classifiers.

L.JabaSheela[3] A Review of Sentiment Analysis in Twitter Data Using Hadoop, Twitter is an online interpersonal interaction website which contains rich measure of information that can be organized, semi-organized and unorganized information. In this work, a technique which performs grouping of tweet notion in Twitter is talked about. To enhance its versatility and proficiency, it is proposed to actualize the work on Hadoop Ecosystem, a generally received circulated preparing stage utilizing the Map Reduce parallel preparing worldview. At long last, broad tests will be directed on genuine informational collections, with a desire to accomplish practically identical or more prominent exactness than the proposed systems in writing.

KomalSutar, SnehalKasab , SnehaKindare, PoojaDhule[4] Sentiment Analysis: Opinion Mining of Positive, Negative or Neutral Twitter Data Using Hadoop, Person to person communication Service (SNS), is a stage to give social relations among people who share basic intrigue. Twitter has turned out to be exceptionally well known. Millions of clients post their remarks on twitter; they indicate their see on current issues. Day by day substantial measure of line information is accessible and which can be useful for mechanical or business reason. Consequently the twitter information can be investigated and utilized for various organizations which will accommodate for choice making. This paper gives a method for investigation of twitter information utilizing AFFIN, EMOTICON for regular dialect preparing. To store, classifications and process expansive assessments we are utilizing Hadoop an open source system.

216

B. M. Bandgar, Dr. S. Sheeja[5]   Analysis of real time social tweets for opinion mining, We built up the indigenous Windows based easy to understand application in Java to concentrate, process and group the genuine time informal organization tweet utilizing unstructured models. The significant continuous tweets are acquired and the same is utilized for nostalgic examination. The prepared significant tweets are ordered into three distinctive supposition mining classes positive, negative and unbiased by utilizing unstructured calculations, for example, EEC, IPC and SWNC demonstrate. The SWNC Model gave better comes about finished the EEC and IPC show. Their outcomes are thought about utilizing the perplexity framework, exactness and precision parameters. The outcomes are likewise envisioned utilizing pie diagram.

Syed Akib Anwar Hridoy, M. TahmidEkram, Mohammad Samiul Islam, Faysal Ahmed and Rashedur M. Rahman[6] Localized twitter opinion mining using sentiment analysis, Examination of open data from online networking could yield intriguing outcomes and experiences into the universe of general assessments about any item, administration or identity. Informal community information is a standout amongst the best and precise markers of open feeling. In this paper we have examined a procedure which permits use also, elucidation of twitter information to decide general suppositions. Examination was finished on tweets about the iPhone 6. Highlight particular popularities and male– female particular examination has been incorporated. Blended suppositions were found yet broad consistency with outside surveys and remarks was watched.

Emma Haddi[7] Sentiment Analysis: Text Pre-Processing, Reader Views And Cross Domains, Opinion investigation has developed as a field that has pulled in a huge sum of consideration since it has a wide assortment of uses that could profit by its comes about, for example, news examination, advertising, question replying, learning administration et cetera. This region, be that as it may, is still right off the bat in its improvement where earnest upgrades are required on many issues, especially on the execution of slant characterization. In this proposal, three key testing issues influencing slant characterization are plot and inventive methods for tending to these issues are displayed. To start with, content pre-preparing has been discovered essential on the slant grouping execution. Thusly, a blend of a few existing preprocessing techniques is proposed for the notion characterization process. Second, content properties of money related news are used to fabricate models to foresee opinion. Two unique models are proposed, one that utilizations money related occasions to foresee budgetary news notion, and alternate uses another intriguing point of view that considers the assessment peruser see, rather than the great approach that inspects the supposition holder see.

PrernaChikersal[8] Modeling Public Sentiment in Twitter, Individuals regularly utilize web-based social networking as an outlet for their feelings and sentiments. Breaking down web-based social networking content to separate feeling can help uncover the considerations and suppositions individuals have about the world they live in. This theory adds to the field of Sentiment Examination, which plans to see how individuals pass on opinion to eventually conclude their feelings and sentiments. While a few assessment arrangement techniques have been contrived, the expanding greatness and unpredictability of social information calls for examination what's more, headway of these strategies. The extent of this task is to enhance customary administered learning techniques for Twitter extremity recognition by utilizing principle based classifiers, etymological examples, and presence of mind learning based data.

PragyaTripathi, Santosh Kr Vishwakarma, Ajay Lala[9] Sentiment Analysis of English Tweets Using RapidMiner, Person to person communication locales nowadays are incredible wellspring of correspondence for web clients. So these are critical hotspot for understanding the feelings of individuals. In this paper, we utilize information digging methods for the motivation behind order to perform slant examination on the perspectives individuals have partaken in Twitter. We gather dataset, i.e. the tweets from twitter that are in natural dialect and apply content mining methods – tokenization, stemming and so forth to change over them into valuable shape and after that utilization it for building estimation classifier that can foresee upbeat, miserable and impartial slants for a specific tweet. Fast Miner instrument is being utilized, that aides in building the classifier and additionally ready to apply it to the testing dataset. We are utilizing two unique classifiers and furthermore contrast their outcomes all together with find which one gives better outcomes.

Ion Smeureanu ,CristianBucur (10) Applying Supervised Opinion Mining Techniques on Online User Reviews, As of late, the breathtaking advancement of web advances, prompt a tremendous amount of client produced data in online frameworks. This extensive measure of data on web stages make them suitable for use as information sources, in applications in light of supposition mining and conclusion examination. The paper proposes a calculation for identifying opinions on film client surveys, in view of gullible Bayes classifier. We make an investigation of the feeling mining area, procedures utilized as a part of conclusion examination and its appropriateness. We executed the proposed calculation and we tried its execution, and recommended bearings of improvement.

V.   METHODOLOGY

Sentence level classification is used to analyze the tweets. 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 Twitter assumption messages. It gathered its own set of data. For the preparation data, it gathered messages that contained the emojis :) furthermore, :(through the Twitter API. The test information was manually. An arrangement of 98 negative tweets and 78 positive tweets were manually checked. A web interface instrument was worked to help in the manual arrangement undertaking. The dictionary will be creating for the positive and Negative words. The tweets will be collect and store. Data pre-processing methods will be applied. The algorithm will be applied to the tweets to analyze their sentiment. Some of the devices have been tried and utilized by researchers over various years, and most by far of these predominantly handle information from Twitter. It is pleasant to have scholastic and social listening apparatuses to recover information from other online networking stages, for example, Facebook, Instagram, and Amazon, and furthermore dull web-based social networking stages, for example, WhatsApp. Be that as it may, this may not be conceivable in light of the fact that these applications are not liable to give the majority of their information to designers as Twitter does. Additionally, there might be moral ramifications of getting to information from dim web-based social networking stages.

It should start to make inquiries with respect to the kinds of research made conceivable by utilizing devices that don't require end clients to hold specialized learning. Besides, it should try to better comprehend the sorts of inquiries more specialized instruments can address. Therefore, engineers of apparatuses should look to liaise with social researchers at the advancement stage, to take into account the likelihood of new highlights in light of sociologies inquire about inquiries.
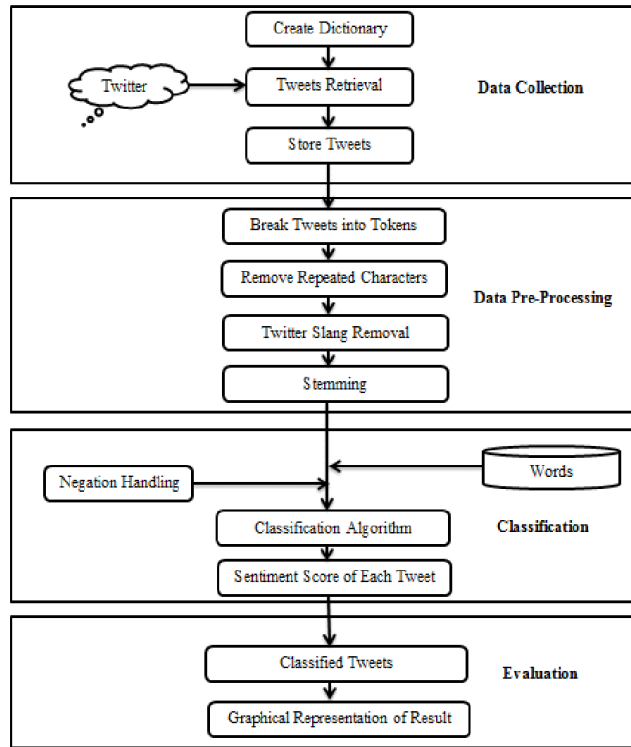
The research follows the steps:-
- The data will be collected from tweets about some specific topic.
- The tables of database are created; it contains the positive & negative words.
- The tweets will scored with some numbered values i.e.1 for positive tweet,-1 for negative tweets & 0 for neutral tweets.
- Data filtering will be performing to remove the unnecessary data from tweets e.g.URLs, usernames, duplicate & repeated characters.
- The slang words (e.g.lol means laughter out loud) will be changed into actual words.
- The words with Negation (never, not, nor etc) will be handle.
- The single tweets will perform the words which will analyze & compare with the database.
- Sentiments will be shown graphically.

The complete detail of the steps is given in following steps:
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 3.1: Database table

Table Name	Field Name	Data Type
NegWords	Nwords	Varchar
PosWords	Pwords	Varchar

Tweets Database	Tweet	Varchar
	Sentiment	int

Table 3.2: Positive words table

Pwords
awesome
gorgeous
happy
beautiful
good

Table 3.3: Negative words table

Nwords
hate
destroy
bad
damage
hurt

Tweets Collection: The tweets are collected from the twitter. Firstly one have to create a twitter account then login to that account to collect the tweets. SQL database is used to store the tweets. www.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 3.4: revolution sentiment score database table

Sentiment Source	Tweet	Sentiment Score
sentiment140	Scary that we are not yet out of the thoughtless decisions and poor execution #gst#revolution	-1
sentiment140	And someone says #revolution wasn't a good move by Modi! I will repeat it was the best step taken by Modi Government so far!	1
sentiment140	Revolution Happened in India.	0

If
  Tweet is positive, then Assign Sentiment Score=1

220

Tweet is Negative, then Assign Sentiment Score=-1

Tweet is Neutral, then Assign Sentiment Score=0

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

*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:

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

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.

Duplicate or Repeated characters: Users sometimes use casual language in tweets. For example, users mostly write 'baaaaaad' in place of bad word. But actually this the same word bad. Sometimes they write 'happpppppppppy" instead of happy. The more than two repeated characters in the document are replaced by only two character occurrences. Hence happpppppy 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.

Table 3.5: Data filtering

Tweets Having	Replaced By
https://t.co//Htxxx	Empty Space
@avneet	Empty Space
@rupinder	Empty space
hhhhaaaappppppy	happy
foooooooddddd	food

Twitter slag removal : There is less space offered for writing a tweet on twitter as tweet 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 3.6: Slang removal

Twitter Slang	Actual Word
Gud	good
Awsm	awesome
Fav	favorite
Thnx	thanks
Bff	best friends for ever

Tc	take care
Sd	sweet dreams

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

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.

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

Table 3.8: Example for tweets pre-processing

Actual Tweet	@avneet And someone says #revolution 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.raseerha.com
Change to Lowercase	@avneet and someone says #revolution 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.raseerha.com
Remove special characters	@ravneet and someone says revolution wasn't a good move by Modi! I will repeat it was the best step taken by modi government so farhappppylol checkout https://www.raseerha.com
Remove Usernames	and someone says revolution wasn't a good move by Modi! I will repeat it was the best step taken by modi government so farhappppylol checkout https://www.raseerha.com
Remove URLs	and someone says revolution wasn't a good move by Modi I will repeat it was the best step taken by modi government so farhappppylol! checkout
Remove extra space	and someone says revolution wasn't a good move by Modi I will repeat it was the best step taken by modi government so farhappppylol checkout

Remove more than 2 repeated characters	and someone says revolution 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 revolution wasn't a good move by Modi I will repeat it was the best step taken by modi government so farhappylaugh out loud checkout
Stop words removal	and says revolution was not good move by modi will repeat was best step taken by modi government so far happy laugh out loud checkout

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. For example,

- iPhone has a difficulty of chargers breaking.
- iPhones are the greatest phones all the time... i am happy to have an iPhone.
- iPhone is the most problematical phone.
- It must be really cool if someone works on iPhone.

These sentences show the tweets about the iPhone. Sentence (1) and (3) are negative sentence whereas (2) and (4) are positive sentence. As (1) & (2) sentence contain words like difficulty, breaking, problematical these are negative words so the sentiment score is negative. Similarly for the (2) & (4) sentence, both are positive.

3.2 Algorithm for sentiment analysis
Problem: A list of tweets collected from twitter, calculate sentiment score for each tweet.
Input: A tweet from twitter for analysis
Output: Sentiment for each tweet
Algorithm:
1. Select tweet from database.
2. Change tweet to lowercase
tweet.toLowerCase();
3. Replace URL in the tweet with empty space
tweet.replaceAll("https?://\\S+\\s?", "");
4. Replace special characters with empty space
tweet.replaceAll("[^a-zA-Z0-9@'\\s]"," ");
5. Replace extra space
tweet.trim();
6. Split the tweet into words
String words[]=tweet.split(" ");
7. Remove more than 2 repeated characters from string
- Add one space at end of word
- Add single unrepeated character to output
- Compare character with next character
- Store 2 similar characters to output

- Discard more than 2 similar characters

8. Repeat step 7 until words.length()

9. Create database connection

Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");

Connection con=DriverManager.getConnection ("jdbc:odbc:mydsn");

10. Replace slang word with its actual word from database

11. Remove unused words, prepositions, articles from tweet

12. If (negation==1)

If associated next word is positive then increment negative counter

Else if associated next word is negative then increment positive counter

13. PreparedStatementpstmt=con.prepareStatement ("select * from NegWords where Nwords=?");

ResultSetrs=pstmt.executeQuery();

if rs.next()

   Increment NegCounter

PreparedStatementpstmt=con.prepareStatement ("select * from PosWords where Pwords=?");

ResultSetrs=pstmt.executeQuery();

if rs.next()

   Increment PosCounter

14. Repeat step 12 and 13 until words.length()

15. End Loop

16. Result=PosCounter-NegCounter;

17. If result>0, then tweet is positive

else if result<0, then tweet is negative

else, tweet is neutral

17. Calculate Error & Accuracy of Algorithm.

Error=ActualValue-CalculatedValue

Accuracy%=((TotalTweets-Error)/TotalTweets)*100

Here, the actual value is human calculated value and calculated value is software predicted.

Tweets dataset

To check the accuracy of the algorithm 6 datasets are created collecting the tweets. Example screenshot of one is shown below:
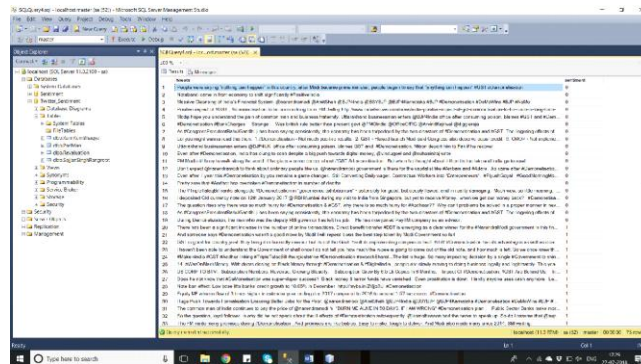

Fig 4.4 Sajjan Singh Rangroot tweets table

VI.   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 %.

Results for Revaluation Dataset

Total 220 tweets are collected. The Algorithm is applied on them. The software calculated the sentiment with the efficiency of 42%. Fig 5.1 shows the analysis of the revaluation tweets. Overall sentiment of tweets shows that the opinion of the public towards the Revaluation is positive. 43 tweets from the total tweets are calculated with wrong sentiment.
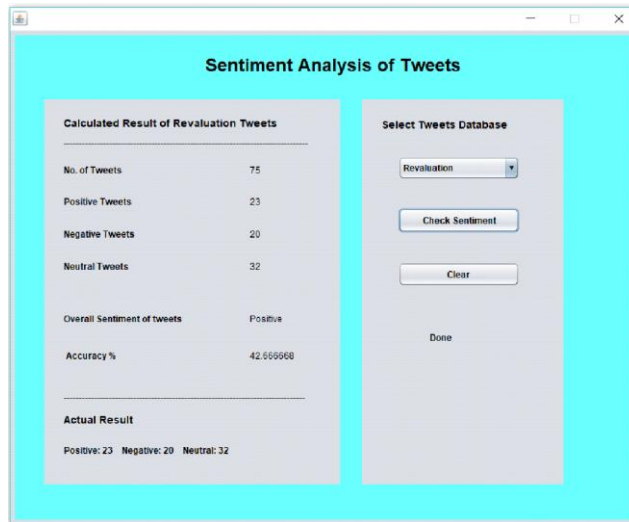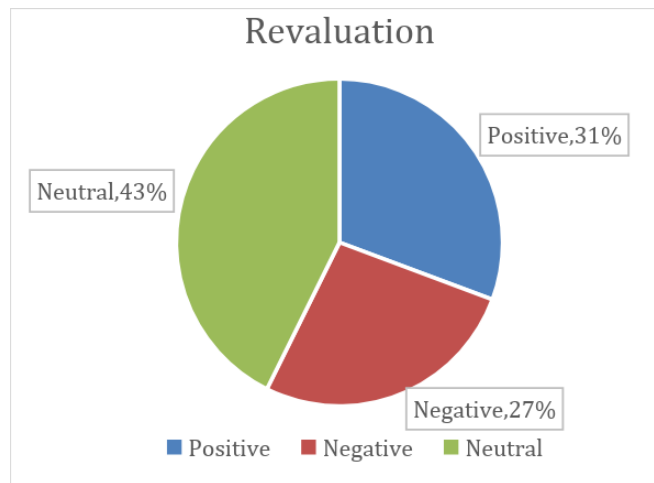

Fig 5.1 Result of Revaluation tweets

Fig 5.2 Pie chart for "Revaluation" tweets

VII.   CONCLUSION & FUTURE SCOPE

Sentiment Analysis is the emerging field that is mainly used in many application areas. Its scope is increasing. So a need arise to create or develop an algorithm that could properly find the sentiment of the public tweets or opinion.

This paper shows a new algorithm that is developed in Java language. The algorithm is applied on tweets and efficiency is calculated based on the accuracy rate of the algorithm. The approximate efficiency of the algorithm is 86%.

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 tweets 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. Guoning, Hu.,Bhargava, P., Fuhrmann, S., Ellinger, S., (2017), "Analyzing users' sentiment towards popular consumer industries and brands on Twitter" arXiv:1709.07434v1 [cs.CL] 21 Sep 2017.
2. Gupta, A., Pruthi, J., Sahu, N., (2017), "Sentiment Analysis of Tweets using Machine Learning Approach", Ankita Gupta et al, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April-2017, pg. 444-458.
3. Sheela, L.J., (2016), "A Review of Sentiment Analysis in Twitter Data Using Hadoop", International Journal of Database Theory and Application Vol.9, No.1, pp.77-86
4. Sutar, K., Kasab, S., Kindare, S., Dhule, P., (2016), "Sentiment Analysis: Opinion Mining of Positive, Negative or Neutral Twitter Data Using Hadoop", IJCSN International Journal of Computer Science and Network, Volume 5, Issue 1, February 2016 ISSN (Online): 2277-5420 www.IJCSN.org Impact Factor: 1.02.
5. Bandgar, B. M., Sheeja, S., (2016), " Analysis of real time social tweets for opinion mining", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 2 pp 1404-1407 © Research India Publications.
6. Hridoy,S.A.A., Ekram, M.T., Islam, M. S., Ahmed, F., and Rahman*, R., M.,(2015), "Localized twitter opinion mining using sentiment analysis", Anwar Hridoy et al. Decis. Anal. (2015) 2:8 DOI 10.1186/s40165-015-0016-4
7. HADDI, E., (2015), "Sentiment analysis: text preprocessing, reader views and cross domains, Brunel university London college of engineering, design and physical sciences department of computer science".
8. Chikersal, P., (2015), "Modeling Public Sentiment in Twitter", o the School of Computer Engineering, in partial fulfillment of the requirements of the degree of Bachelor of Engineering (B.Eng.) in Computer Science at Nanyang Technological University, Singapore.
9. Tripathi, T., Vishwakarma, S.,Kr., Lala, A., (2015), "Sentiment Analysis of English Tweets Using Rapid Miner", 2015 International Conference on Computational Intelligence and Communication Networks, 978-1-5090-0076-0/15 $31.00 © 2015 IEEE DOI 10.1109/CICN.2015.137.
10. Smeureanu, I., Bucur, C., (2012), "Applying Supervised Opinion Mining Techniques on Online User Reviews", InformaticaEconomică vol. 16, no. 2/2012.
11. Kaur, R., Gupta, G., Singh, G., (2017), "Sentiment Analysis and its Challenges", International Journal of Engineering Research in Computer Science & Engineering, Vol. 4, No. 2, pp. 97-102.
12. Ghai, A.S., Gupta, G., Bhathal, G.S., (2017), "Survey on the effects of Sports on Vocational Academics", International Journal for Multi Disciplinary Engineering & Business Management, Vol. 5, No. 3, pp. 7-9.
13. Sachdeva, A., Gupta, G., Bhathal, G.S., (2017), "Review of Data Mining in Contrast to Modern Medical Equipments", International Journal for Multi Disciplinary Engineering & Business Management, Vol. 5, No. 3, pp. 18-20.
14. Kaur, H., Gupta, G., Attwal, K.P.S., (2017), "Review of Electronic Library using Data Mining", International Journal for Multi Disciplinary Engineering & Business Management, Vol. 5, No. 3, pp. 10-13.
15. Kaur, H., Gupta, G., Attwal, K.P.S., (2017), "Review on Therapies and Medical Treatment using Data Mining", International Journal for Multi Disciplinary Engineering & Business Management, Vol. 5, No. 3, pp. 14-17.

16. *Kaur, K., Bhathal, G.S., Gupta, G., (2017), "An Analytic and Comparative Study of Map Reduce – A Systematic Review", International Journal of Advanced Research in Computer Science, Vol. 8, No. 5, pp. 2453-2459.*
17. *Kaur, J., Bhathal, G.S., Gupta, G., (2017), "Cloud Computing: Types, Topologies, Virtual Machines and VM Migration for Decrease in Power Consumption", International Journal of Advanced Research in Computer Science, Vol. 8, No. 5, pp. 2357-2361.*
18. *Kaur, D., Bhathal, G.S., Gupta, G., (2017), "Analysis of DDOS attacks in Cloud Networks", Asian Journal of Computer Science and Information Technology, Vol. 9, No. 14, pp. 9-14.*
19. *Kaur, A., Gupta, G., Singh, G., (2017), "Role of Virtualization in Cloud Computing", Global Journal of Engineering Science & Research, Vol. 4, No. 7, pp. 142-149.*
20. *Gupta, G., Aggarwal, H and Rani, R. (2015), "Mining the Customers Data for making Segments based on RFM Analysis in Building Successful and Profitable CRM", Ciência e TécnicaVitivinícola Journal, Vol. 30, No. 3, pp. 261-270.*
21. *Gupta, G. &Aggarwal, H. (2012), "Improving Customer Relationship Management Using Data Mining", International Journal of Machine Learning and Computing (IJMLC), Vol. 2, No. 6, pp. 874-877.*
22. *Gupta, G. &Aggarwal, H. (2016), "Analyzing Customer Responses to Migrate Strategies in Making Retailing and CRM Effective", Int. J. Indian Culture and Business Management (IJICBM). Vol. 12, No. 1, pp. 92-127.*
23. *Gupta, G., Aggarwal, H. & Rani, R. (2016), "Segmentation of Retail Customers based on Cluster Analysis in Building Successful CRM", Int. J. of Business Information Systems (IJBIS). Vol. 23, No. 2, pp. 212-228.*
24. *Gupta, G. &Kahlon, J.S. (2016), "PREDICTING THE CAUSE OF ABSENTEEISM AMONG PUBLIC VERSUS PRIVATE COLLEGE STUDENTS USING DATA MINING", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 4, No. 4, pp. 1-5.*
25. *Kaur, S. & Gupta, G. (2016), "DATA MINING APPROACH TO CRM IN BANKING SECTOR", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 4, No. 2, pp. 81-89.*
26. *Kaur, N. & Gupta, G. (2016), "EFFECTIVENESS OF CRM IN RETAIL SECTOR USING DATA MINING", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 4, No. 2, pp. 71-80.*
27. *Kaur, K. & Gupta, G. (2015), "PREDICTING THE USE OF INTERNET AMONG TEACHERS AND STUDENTS USING DATA MINING", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 3, No. 3, pp. 97-106.*
28. *Kaur, N.K. & Gupta, G. (2015), "PREDICTING THE VARIOUS RISK FACTORS OF LEPROSY USING DATA MINING TECHNIQUES", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 3, No. 3, pp. 79-87.*
29. *Kaur, H. & Gupta, G. (2014), "Optimizing Cart Algorithm by Enhancing Attributes", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 2, No. 3, pp. 34-38.*
30. *Kaur, J. & Gupta, G. (2014), "Hybrid of K-Means and Hierarchal Algorithms to Optimize Clustering", International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM). Vol. 2, No. 3, pp. 39-44.*

227