Journal of Applied Information Science
7 (2), December 2019, 10-18
http://www.publishingindia.com/jais

An Efficient Sentimental Analysis Mining for Unclassified Data in Big Data Analytics

Komarasamy G.[1*] and Vivek V.[2]
[1]Department of Computer Science and Engineering, School of Engineering and Technology, Jain (Deemed-to-be-University), Bangalore, Karnataka, India. Email: gkomarasamy@gmail.com
[2]Department of Computer Science and Engineering, School of Engineering and Technology, Jain (Deemed-to-be-University), Bangalore, Karnataka, India. Email: vivek544@hotmail.com
*Corresponding Author

Abstract: In the community networks, large numbers of users contribute to their opinions, for tracking and analyzing public response for making a valuable platform. Such tracking and analysis can provide critical information for decision-making in a variety of domains. In this paper, we generate a real-time suggestion for user comments among social network. To analyze the user comments, pattern recognition and data dictionary have been implemented. A raw dataset has been used as the input like posted by, date, time, posted topic, user name, and user comments. The data dictionary administrator can create a dataset for pattern matching, as it contains a group of all sentimental words. Sentimental words may be positive and negative words to be updated by the administrator. Instead of reading all the comments commented by the users, chart illustrates the graphical information of the discussion. Through the pattern recognition method, the user comments are categorized and the given input so that they could be separated into various sentiments like happy, sad, angry, etc. Moderate comments analyzed based on non-identified words from all the sentimental classifications. All the comments are grouped using global patterning reports and various charts generated. Ranking can be calculated using global patterning report. Through the generated chart, the admin can view a clarity report for the users' opinion on the comment posted.

Keywords: Classifications, Dictionary, Pattern, Recognition, Reports, Sentiment.

I. Introduction

This work on sentimental data analysis is also called opinion mining for unclassified data in big data analytics. The objective is to produce a real-time suggestion for public remarks among social networks in a graphical configuration. It comes under data mining and data engineering category because of text and data processing. In social networks, there are bunches of remarks and comments posted by the general people every day. In case of fewer comments, it is possible to read all; but in case comments are in thousands, it is really hard to read all the comments.

The proposed a sentimental data analysis model using pattern recognition. All sentiments like positive, negative and moderate feedback will be calculated. Taking an example from social networks, in any census-based process, there are always variables that are uncertain. Likes, comments, posts, feeds, and other variables may not be known with great precision. All the comments will be analyzed by the pattern recognition method and classification technique. As mentioned above, all the sentiments in the comments will be separated and not hitting sentence will be considered as moderate comment. These results will be stored in the database and represented in a graphical format.

A. Data Analysis Methodologies

Data analysis defines the method of finding the unprocessed data. Fig. 1 illustrates that the data analysis model is collected and analyzed to answer questions, test hypotheses, or disprove theories. There are many phases that can be distinguished. There are following phases of sentiment analysis.

Data Requirements: The data collected as inputs to the study are particular based on the requirements of those directing the study. The general type of entity upon which the data will be collected is referred to as an experimental unit.

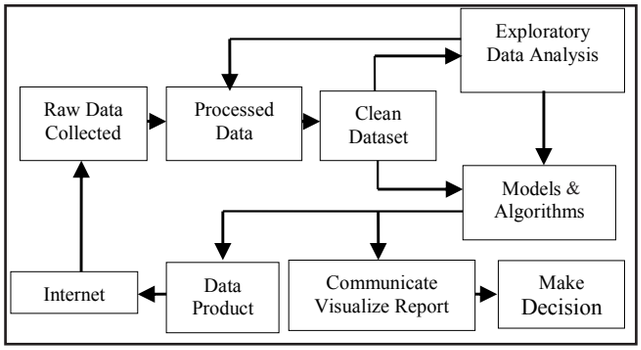

Fig. 1: Data Analysis Model

Data Collection: Data are collected from various sources like information technology personnel within an organization. The sensor data are collected from recording devices, satellites, and traffic cameras. The phases of the intelligence cycle used to convert raw information into actionable aptitude. Fig. 2 shows the relationship between intelligence and data information. Data initially obtained must be processed by an analysis.

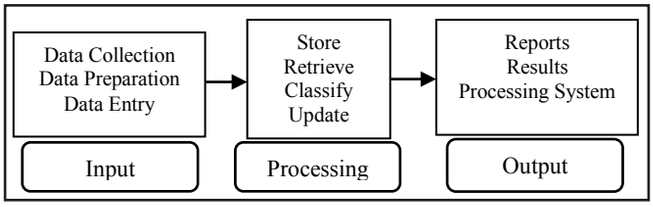

Fig. 2: Relationships of Data, Information, and Intelligence

Data Cleaning: Data cleaning is the process of preventing and correcting these errors. Once the data are processed and organized, the data may be incomplete.

Exploratory Data Analysis: Once the data are cleaned, it can be analyzed. Analysts may apply a range of techniques referred to as examining data analysis to begin understanding the messages controlled in the data.

Modeling Process: Arithmetical formulas can be used for identifying the relationships among variables, like causation or correlation.

II. Literature Review

Learning similarity metrics for event identification in social media [1-2] process is described in this article. Social media e-commerce sites are distribution outlets for users looking to share their experiences and interests on the internet. These sites host substantial amounts of user-contributed materials for different real-world measures of various types and scales. Twitter is mainly popular in micro blogging and social medium [3]. People from time to time comment post in Twitter which is called tweeting. The diversity of people on twitter makes the tweets more flexible and valuable. Therefore, Twitter becomes the most valuable place to find opinions on any subject. This allows computer scientists to perform believable sentiment analysis and develop pathways for data mining. This data can be used in marketing, sales or poll analysis. Timely feedback on products can be collected by evaluating peoples' tweets on Twitter [4-6].

Parameter assessment for text analysis [7] are starting with maximum possibility, Bayesian estimation, posterior process and central concepts like conjugate distributions.

Twitter sentiment classification using distant supervision [8] micro blogging services like Twitter is becoming more and more influential in today's globalized world. Its facets like sentiment analysis are being extensively studied. Authors describe the techniques to speed up the computation process for sentiment analysis. Another method that is used for sentiment analysis is the machine learning approach [9]. This method is effective for the classification of sentences and documents by training the classifier to determine positive, negative and neutral sentiments. Since manual labeling of large set of tweets is often time consuming and difficult, this approach is not easy to implement. Also, deep learning algorithms could provide the most accurate results, but these techniques are extremely computationally expensive to train. To optimize the large amount of matrix multiplication operations that deep learning involves, substantial investment is needed to upgrade the IT infrastructure for more processing power.

III. Existing System

A. Word Emotion Computation

Sentiment computing for the news event based on the social media big data [10] is discussed word emotion computation. People can publish views, opinions and attitudes for objects, individuals, events, or topics on social media whenever and wherever. These views, opinions, and attitudes contain users' emotion. They call it social media text emotion. Up to now, there are many researches on social media text emotion, and many multidimensional emotions have been reported from different perspectives. It is used for multidimensional emotion classification, in which emotions are divided as love, joy, anger, sad, fear, and surprise.

The proposed is the method to compute emotion of words in a news event. Given a news event, a word in different stage may have different emotions, it needs to compute the word emotion at a specific time. Due to the large scale of word number and the varied usage of words, computing word emotion is extremely difficult. What is worse, with the development of the web, the appearance of new cyber words makes word emotion computation even harder. According to Table I, all eight groups have reached an accuracy of more than 75%. At the same time, the line chart shows that the refined emotion is more accurate than before, but not very obvious. The reason for that acquire all microblogs about the Malaysia Airlines MH370 and the WEAN based on that is totally integrated.

Table I: Accuracy of Text Emotion Computation

Group No.	Before Refinement	After Refinement
1	0.785714286	0.787456446
2	0.801709402	0.803418803
3	0.804794521	0.808219178
4	0.7694974	0.776429809
5	0.75	0.755244755
6	0.752557732	0.7577319599
7	0.768439108	0.777015437
8	0.803108808	0.80656304
Average	0.779480106	0.784009928

Definition of Word Emotion Computation: Like association link network (ALN), WEAN is constructed by words and their association rules, which can be represented by,

$$\text{WEAN} = <N,W> \quad (1)$$

In Eq. (1), N is a set of nodes and W is a set of weighted links belonging to NXN.

The different scale and intension of word circumstance can affect word emotion. The larger the scale and the stronger the intension, the more intense the word emotion [11]. Based on these, we propose two assumptions, namely quantity assumption and intensity assumption.

Algorithm 1: Word Emotion Computation

Input: WEAN and $W^{e}_{i,j}$

Output: Word emotions randomly initialization

1: Tag = 1, it = 1;
2: for each k^{th} dimension of emotion do
3: for tag = 1 do
4: tag = 0;
5: for each word w_i in WEAN do
6: Update $V^{ek.}_{wi}$ (it)
7: end for
8: for each word w_i in WEAN do
9: if $|V^{ek.}_{wi}$ (it)- $V^{ek.}_{wi}$ (it-1)|>0 then
10: tag = 1;
11: end if
12: end for
13: end for
14: end for

Algorithm 1 refines the word emotion found by the words in standard sentiment thesaurus. For the words in both standard sentiment thesaurus of e^k and WEAN, define them as w_B. e^k of these words is approximate to 1, actually. However, the word emotion computed by Algorithm 1 may be less than 1. At this time, it needs to refine them. To set the maximum errors for words in w_B is δ, and every step is Δ, which is a positive number. After that, do the word emotion computation again using Algorithm 1. When all the words (in w_B) emotion values are close to 1, stop the word emotion computation. The whole procedure is summarized in Algorithm 2.

The author in [12] depicts many techniques used to find the divergence of the tweets. The algorithms used are Naïve Bayes, K-Nearest Neighbor, and Random Forest. This paper illustrates that the polarity of the reviews helps in various domains. Experts systems can be implemented which can assist the users with complete reviews of movies, products, and services without the users to go through individual reviews. Thus, it can directly take decisions based on the results given by the Experts systems.

B. Emotion Refinement by Standard Sentiment Thesaurus

The word emotion computation above only considers the emotions of emoticons in microblogs. However, in the early stage of one event, microblogs about this event are not many, especially the microblogs with emoticons. The WEAN constructed by few words and rules lacks credibility. At this point, the accuracy of using WEAN to compute the word emotion will be low.

The words in standard sentiment thesaurus have obvious and stable emotion, which seldom change with the different news events. Therefore, we planned to refine the word emotion obtained by Algorithm 1 according to the words in standard sentiment thesaurus. For the words in both standard sentiment thesaurus of e^k and WEAN, define them as w_b. e^k of these words is approximate to 1, actually. However, the word emotion computed by Algorithm 1 may be less than 1. At this time, it needs to refine them. To set the maximum error for words in w_B is δ, and every step is Δ, which is a positive number. Not surprisingly, the emotion values of all the words are convergent after a number of iterations.

The more interesting observation is that the different words will converge to different values even with different or same initializations. Therefore, it can draw the conclusion that the final emotion values for words from Algorithm 1 are not influenced by the initialization but determined by their structural portions in the WEAN.

When the emotion of one word in w_B computed by iteration is much less than 1, add the weights of the links with the word. After that, do the word emotion computation again using Algorithm 1. When all the words emotion values are close to 1, stop the word emotion computation. The whole

procedure is summarized in Algorithm 2. Texts are composed by words. Words' emotions reflect texts' emotions indirectly. After computing word emotion, obtain the text six dimensional emotion by adding the six-dimensional emotions of words that in the text, respectively, in Eq. (2).

$$es = \Sigma_i w_i \in S\, ew_i \tag{2}$$

where,

$$e_{wi} = < e_{wi}^{love},\ e_{wi}^{joy},\ e_{wi}^{anger}, e_{wi}^{sad},\ e_{wi}^{surprise} >$$

is a six-dimensional vector, which means the six-dimensional emotion of w_i; e_s is also a six-dimensional vector, which means the six-dimensional emotion of text.

Algorithm 2: Word Emotion Refinement

Input: WEAN, ***wi***,δ, Δ and standard sentiment thesaurus

Output: Word emotions

1: Word emotions obtained by Algorithm 1;
2: tag = 1;
3: for each k-th dimension of emotion do
4: for tag = 1 do
5: tag = 0;
6: for each word i both in standard sentiment thesaurus and WEAN do
7: if |1- ***Vwbek***| >δthen
8: tag = 1;
9: for each link weight ***wi***, ***je*** linked with word i do
10: ***wi***, = ***wi***, + Δ. frac |1- ***Vwbek*** | δ
11: end for
12: end if
13: end for
14: if tag = 1 then
15: for each ***wi*** in WEAN do
16: Update ***Vwbek***
17: end for
18: end if
19: end for

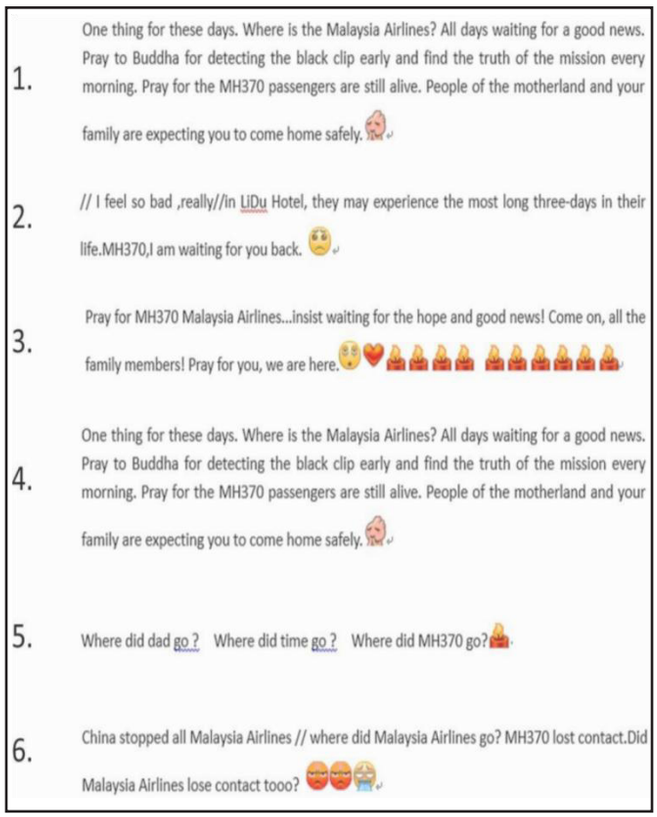

Fig. 3: Six Microblogs with Emoticons

In order to show the accuracy of our method, compare our method that sums up word emotions together as the text emotion. In Table II, we have shown six microblog (Fig. 3 from The Malaysia Airlines MH370) examples with their emotions from both, i.e., our method and the benchmark emotion, using developed evaluation metric in Eq. (1). The result shows that although the short text is not with single emotion, the main dimension's emotion is in keeping with the benchmarks.

C. Evaluation Metric

In order to evaluate performance of the method on the text affective computing [13] need to design an evaluation metric. The idea is that the emotion of a microblog should be consistent with its emoticons since microblogs are very short. Therefore, the emotion computed from emoticons could be seen as the benchmarks of the proposed method as follows:

$$e_b^d = \frac{1}{N_e^d} \sum_{i=1}^{N_e^d} e_{ei} \tag{3}$$

where N_e^d is the number of emoticons in microblog d; e_i is one emoticon in microblog d; e_{ei} is the emotion vector of an emoticon; $e_b{}^d$ is the emotion vector of microblog d. $e_b{}^d$ could be seen as the benchmark of our proposed method. This can

evaluate the performance of the proposed method by comparing e_b^d with the value from our method as shown in Eq. (3).

The evaluation metric needs the microblogs with emoticons, needs to split the data into two parts: one part as training data and the other as test data. The whole procedure is as follows:

- Pick the microblogs with emoticons of a news event.
- Split these microblogs with emoticons into two parts: training part and test part.
- Put the training part with other microblogs without emoticons into our proposed method.
- Predict the test part using the proposed well-trained method.
- Compute the benchmarks for the test part.
- Evaluate the performance of the proposed method.

The word emotion computation needs the initialization of the variables. A natural question may be asked: is this algorithm sensitive to the different initializations? The answer is 'no'. The empirical Algorithm 1 is not sensitive to the different initializations. To feed the Algorithm 1 in various initializations and run it many times, the values at each iteration are recorded and plotted. It can be seen that the many starting points will lead to same convergent value for the same word after a number of iterations. This conclusion improves our confidence about the output of Algorithm 1 even with random initialization. Furthermore, we plotted the convergent curves of 10 different words. Not surprisingly, the emotion values of all the words are convergent after a number of iterations.

The more interesting observation is that the different words will converge to different values even with different or same initializations. Therefore, to draw the conclusion that the final emotion values for words from Algorithm 1 are not influenced by the initialization but determined by their structural portions in the WEAN is not wrong. In order to show the effectiveness of word emotion refinement in Algorithm 2, we have eight group experiments to compare the word emotions before and after the refinement. All eight groups have reached an accuracy of more than 75%. At the same time, the line chart shows that the refined emotion is more accurate than before, but not very obvious. The reason for that is to acquire all microblogs about the Malaysia Airlines MH370 and the WEAN based on that is totally integrated.

TABLE II: TEXT EMOTION FOR THE SIX MICROBLOGS IN FIG. 3

Micro Blog Id	Straightforward Sum	Benchmark
1	< 1.0, 0.1, 0.1, 1.0, 0.2, 0.1 >	< 1.0, 0.0, 0.0, 1.0, 0.0, 0.0 >
2	< 0.8, 0.0, 0.0, 1.0, 0.0, 0.0 >	< 0.0, 0.0, 0.0, 1.0, 0.0, 0.0 >
3	< 1.0, 0.0, 0.0, 1.0, 0.3, 0.0 >	< 1.0, 0.0, 0.0, 1.0, 0.1, 0.0 >
4	< 1.0, 0.1, 0.1, 1.0, 0.1, 0.0 >	< 1.0, 0.0, 0.0, 1.0, 0.1, 0.0 >
5	< 1.0, 0.1, 0.2, 1.0, 0.3, 0.2 >	< 1.0, 0.0, 0.0, 0.9, 0.1, 0.0 >
6	< 1.0, 0.8, 1.0, 1.0, 1.0, 0.9 >	< 0.0, 0.0, 1.0, 0.3, 0.0, 0.1 >

The obtained results even more accuracy than the refined, the Table II, the plot word emotions of 5,363 words from news event in Malaysia Airlines MH370. Since each word emotion is a six-dimensional vector, there are six subfigures in 8 each of which corresponds to one dimension of word emotion.

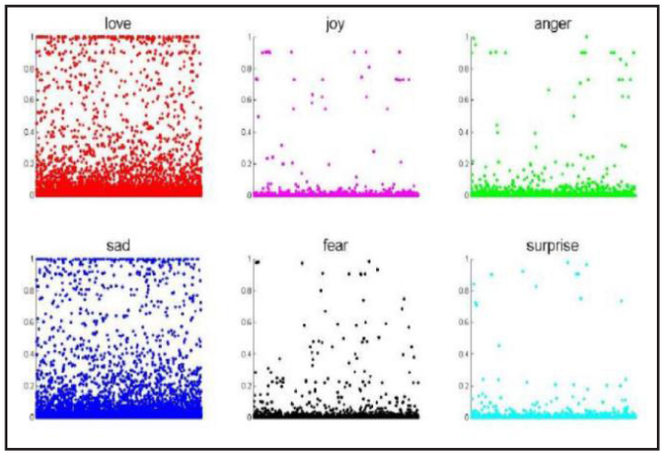

Fig. 4: Visualization of the Word Emotions for 5,363 Words of News Event the Malaysia Airlines MH370

Each point in the subfigure denotes a word and its corresponding value in y-axis is its final emotion value on this dimension. It can be seen that the emotions 'love' and 'sad' of these words are much stronger than the other four emotions because there are more words located in [0:51] in subfigures love and sad. The conclusion that people mainly feel 'love' and 'sad' on event the Malaysia Airlines MH370. In order to show the accuracy of our method, compare our method that sums up word emotions together as the text emotion.

Herewith shown six microblog examples with their emotions from both our method and the benchmark emotion using developed evaluation metric. Fig. 4 shows that although the short text is not with single emotion, the main dimension's emotion is in keeping with the benchmarks. Therefore, draw the conclusion that our method is effective in short text emotion computation for the news event.

IV. Proposed System

A. Social Network Pattern Creation

A social network pattern is a web application that is accessed over a network such as the internet or an intranet. The term may also mean a computer software application that is hosted in a browser-controlled environment. Web applications are accepted due to the ubiquity of web browsers, and the handiness of using a web browser as a client. Table III is an example for a dataset which consists of users review on GALAXY NOTE 4 features.

Table III: Dataset Which Consists of Users Review on GALAXY NOTE 4 Features

Posted By	Post	Comment	Commented Date and Time
Samsung Mobile India	What's your favorite Galaxy Note4 Feature.	The best phone I have used till now	March 3 at 11:24 pm
Samsung Mobile India	What's your favorite Galaxy Note4 Feature.	Excellent Floating Multiwindow	March 4 at 9 am
Samsung Mobile India	What's your favorite Galaxy Note4 Feature.	Quad HD super amoled display... it's the best display	March 3 at 11 pm

B. Centralizing the Data

Data analysis purposes all the data will be uploaded in the centralized server. Centralized data distribution systems defined here are systems that allow distributed end-user applications, databases, and data providers to be integrated with dedicated data sources.

C. Cluster Analysis

Cluster analysis describes the extracting a data from large amount of datasets. This can be very effective in identifying patterns and distinguishing objects.

D. Classification

Classification models calculate the categorical class labels and prediction models, to predict the constant valued functions.

i. Building the Classifier Model

In this step, the classification algorithms build the classifier. The classifier is construct from the training set completed in database tuple and their associated class labels. Each tuple that comprise of training set is referred to as a category or class.

ii. Classifier for Classification

In this step, the classifier is used for classification. Here, the test data are used to estimate the accuracy of classification rules. Table IV illustrates the building the classifier model.

Table IV: Building the Classifier Model

Fields	Data Type	Size	Con-strains	Description
Id	Int	20	Primary Key	Data identification number
Posted by	Varchar	40	Not Null	Comment posted by details
Posted date	Varchar	80	Not Null	Comment posted date
Posted time	Varchar	80	Not Null	Comment posted time
Shared to	Varchar	250	Not Null	Shared to details
Post	Varchar	50	Not Null	Post details
Name of the Member	Varchar	30	Not Null	Name details
Comment	Varchar	40	Not Null	Comment
Status	Varchar	40	Not Null	Status of comment
Date time	Date time	30	Not Null	Date and Time Details

E. Sentimental Data Analysis Model

The pattern recognition is implemented for sentimental data analysis. The pattern recognition method also hand shacked with clustering and classification methods. These methods will analyze the input data from the dataset; in the case of customized social networks, it will analyze online based data. Each sentence is analyzed with pattern recognition methods. So that all the sentimental words will be compared accordingly. Table V shows that the repeated comments will be omitted and clustered for fined tuned data report.

TABLE V: ANALYZED SENTIMENTS FROM THE DATASET

Data Set	Positive	Negative	Happy	Satisfied	Sad	Angry	Moderate
This product is excellent and good	Yes	No	Good	Excellent	No	No	No
I am more *disappointed* in this product	No	Yes	No	No	Disappointed	No	No
I am more *disappointed* in this product. I *won't buy* this anymore	No	Yes	No	No	Disappointed	Won't buy	No
I am *ok* with this product	Yes	No	No	No	No	No	Ok
This is *not worthy* for 25000 rupees	No	No	No	No	No	Not worthy	No
After buying the mobile phone, there are many scratches in the panel	No	No	No	No	No	No	No sentimental words found

A common example of a pattern-matching algorithm is regular expression matching, which looks for patterns of a given sort in textual data and is included in the search capabilities of many text editors and word processors. After separating the positive and negative comments here, the repeated comments are classified and deleted. According to this study, initially all the data will be considered as the input data and processing data. But, the proposed method needs to preprocess the data for a fine tuned result.

V. EXPERIMENTAL RESULTS

The comment description details are given in Table VI shows the result of classification of positive, negative and moderate comments from the dataset.

TABLE VI: COMMENT DESCRIPTION DETAILS

S. No.	Comment Description	Number
1.	Total Number of Comment	38
2.	Number of Positive Comment	21
3.	Number of Negative Comment	2
4.	Number of Moderate Comment	15
5.	Number of Redundant Data	0

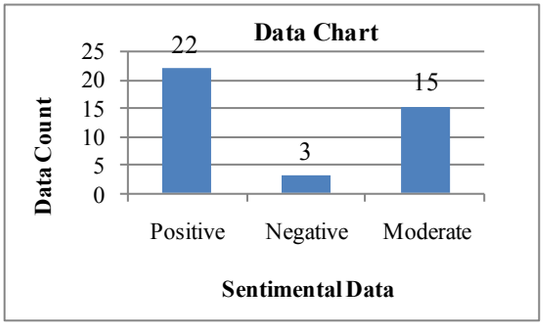

Fig. 5: Before Data Analysis Process

TABLE VII: COMMENT DESCRIPTION DETAILS

S. No.	Comment Description	Number
1.	Total Number of Comment	14
2.	Number of Positive Comment	5
3.	Number of Negative Comment	1
4.	Number of Moderate Comment	8
5.	Number of Redundant Data	24

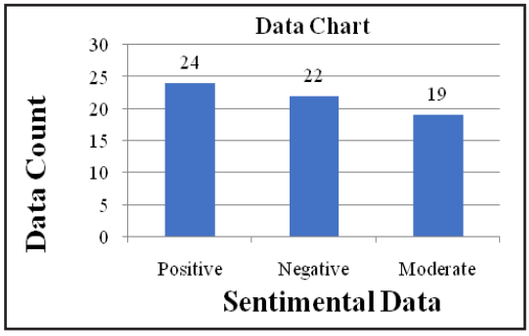

Fig. 6: After Data Analysis Process

Fig. 6 shows the actual result after the preprocessing. The preprocessing has been done using clustering and classification methods. It removes all the repeated users and repeated comments. After the removal of data, the analyses are executed again in the same execution process. Table VIII shows the comparison methods and the comparison of before and after data analysis chart is shown in Fig. 7.

TABLE VIII: COMPARISON METHODS

Data Description	Before Processing	After Processing
Total Number of Data	38	14
Number of Positive Data	21	05
Number of Negative Data	02	01
Number of Moderate Data	15	08

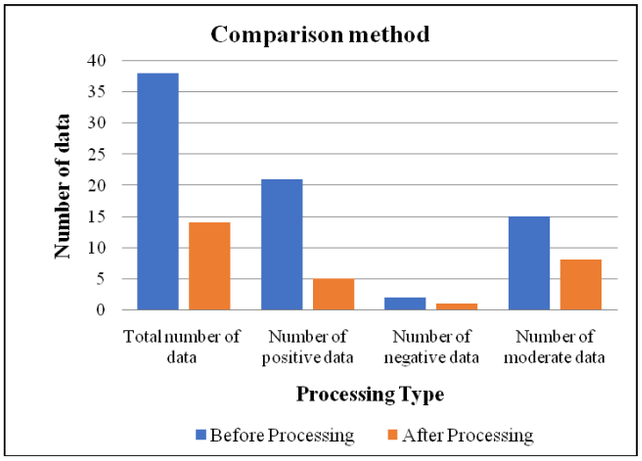

Fig. 7: Comparison Chart of Before and After Data Analysis

The timing process will calculate all the sentimental process like positive performance, negative performance, and moderate performance are illustrated in Fig. 8. Each and every process is calculated according to the query-execution process.

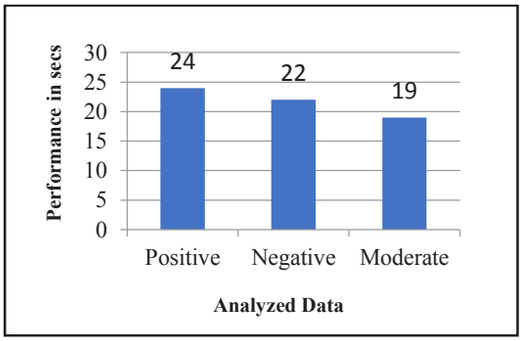

Fig. 8: Timing Chart

VI. Conclusion

Classification is very essential to organize data and retrieve information correctly and swiftly. Implementing machine learning technique to classify data is not easy given the large amount of heterogeneous data that're present in the web. Text categorization algorithm depends on the accuracy of the training dataset for building its decision trees. The text categorization algorithm learns by supervision. Due to this text categorization algorithm, it cannot successfully classify documents in the web. The data in the web are unpredictable, volatile, and most of it lacks metadata. The way forward for information retrieval in the web, in the future opinion, would be to advocate the creation of a semantic web where algorithms which are unsupervised and reinforcement learners are used to classify and retrieve data. Thus, the article illustrates the trends, threads, and process of the text-categorization algorithm, which was implemented for finding the sensitive data analysis.

The future enhancement discusses the issues related to the application of the text categorization algorithm, an important representative of the inductive learning family. A prototype workbench which has been developed to provide an integrated approach to the application of text categorization is presented. The design rationale and the potential use of the system are justified. The further enhancements of the workbench like implement for web based application, handshakes with inductive learning algorithm.

References

[1] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in *Proc. of 3rd ACM WSDM*, Macau, China, 2016.

[2] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc. of 5th International AAAI Conference Weblogs Social Media*, Barcelona, Spain, 2016.

[3] A. Pak, and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. of the International Conference on Language Resources and Evaluation, LREC 2010*, Valletta, Malta, 2010.

[4] E. M. Cody, A. J. Reagan, P. S. Dodds, and C. M. Danforth, "Public opinion polling with twitter," The University of Vermont, August 2016.

[5] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," HP Laboratories, 2011.

[6] P. B. Filho, L. Avanço, T. Pardo, and M. D. G. V. Nunes, "NILC_USP: An improved hybrid system for sentiment analysis in twitter messages," *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014),* 2014.

[7] G. Heinrich, "Parameter estimation for text analysis," Fraunhofer IGD, Darmstadt, Germany, Univ. Leipzig, Leipzig, Germany, Tech. Report, 2015.

[8] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford: 2014.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up," *Proc. of the ACL-02 Conference on Empirical Methods in Natural Language Processing - EMNLP'02*, 2002.

[10] D. Jiang, X. Luo, J. Xuan, and Z. Xu, "Sentiment computing for the news event based on the social media big data," Special section on Intelligent Sensing on Mobile and Social Media Analytics, *IEEE Access*, vol. 5, pp. 2373-2382, 2017.

[11] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," *Proc. ACL Conf. Empirical Methods Natural Lang. Process.,* pp. 79-86, 2002.

[12] P. Baid, A. Gupta, and N. Chaplot, "Sentiment analysis of movie reviews using machine learning techniques," *International Journal of Computer Application*, vol. 179, no. 7, pp. 45-49, December 2017.

[13] B. Pang, and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.