Sentiment Analysis of Face Book Statuses

K.Srividya, A. Mary Sowjanya

Abstract: Individuals share their encounters, suppositions or essentially talk pretty much whatever worries them on the web. The extensive measure of accessible information pulls in framework engineers, considering of mining and investigation. Sentiment analysis has popularized due to the availability of abundant opinions that recides in social networks such as Facebook, Twitter, etc. Sentiments are published on these media inform of texts for expressing social support, happiness, anger, friendship etc. A Sentiment is frequently expressed in subtle and complex ways. In addition, data collected from World Wide Web often contains high noise. Sentiment Analysis is treated as a characterization undertaking as it groups the introduction of a content into either positive or negative. In this paper we present sentiment analysis of Facebook statuses using Naive Bayes Classifier and Support Vector Machine (SVM).The essential and basic thought of the paper is that, realizing how individuals feel certain Facebook statuses can be utilized for classification

I.   INTRODUCTION

Sentiment Analysis is turning into a famous report nowadays, for the most part as a result of the way that person to person communication destinations incorporate online clients who are allowed to express their considerations, sentiments and impressions concerning an explicit point. Indeed, now a days, any sort of promoting business is at present drenching to the new patterns of organizations. Aside from composed overviews, the organizations likewise broaden their consumer loyalty investigation through the web, with the end goal to accumulate a lot of information. The interpersonal interaction web page Face book will be the focused on site for this paper. This is on the grounds that Face book, in contrast to Twitter, has 5000 characters for each notice[12]. Consequently, clarity in sentence development would be more expectable. Besides, the quantity of Facebook clients is totally rich, in particular it is a decent example for making a corpus.

Scarcely any investigations about analysis and sentiments have just been exhibited[13]. These investigations are focused on contents related to Twitter, for tweet refreshes about an explicit subject, generally on brands of items. In these systems from twitter raw data will be collected using hashtags, like sample topic, and utilize the data as a corpus for implementing the classifying method. But, utilizing and analyzing the social networking

Revised Manuscript Received on December 22, 2018..

K.Srividya,Assistant Professor,Department of cse, Gmrit,Rajam Email: srividya.k@gmrit.org

A. Mary Sowjanya,Asst. Professor, Department of CS&SE, Andhra University College of Engineering (A), Visakhapatnam, AP, India Email: sowmaa@yahoo.com

data twitter has very known downside[14,15]. Each twitter refresh is limited to 140 characters long. Thus, Twitter clients will in general utilize overwhelming shortened forms and divided articulations.

Sentiment Analysis is engaged with the investigation of opinion mining[16]. Moreover, Sentiment Analysis is regularly utilized by publicists, film makers and different associations that desire to obtain their clients' response on an explicit theme[17]. Despite the fact that the least complex approach to assemble sentiments is as studies, there are couple of disadvantages, which comprise of extraordinary impediment of the promoting research. The issues developing of this methodology are the lead of a review for every item or highlight, the organization, the circulation and timing of the overview, and the dependence on the positive attitude of individuals to take the study[18]. All the previously mentioned issues require high support for the advertising research gathering's perspective.

On account of assessments, only some word used the sentence have importance. Some words are named as clamor since they are of no utilization during the time spent grouping the extremity of the supposition. Additionally, there are words like "not" and at whatever point they are appended to a positive word, they will give a negative tweet to the existent assessment. Beside words, images like tragic face or an upbeat face " ☺ " gives importance in common dialect preparing.

II.   .BACKGROUND KNOWLEDGE

The fast development in the substance set on the Web, give an exceptionally vast accumulation of literary assets. Clients share their encounters, sentiments or basically talk about pretty much whatever worries them on the web. The enormous measure of put information pulls in framework designers, who are doing programmed mining and investigation. The essential hidden thought of the client is that of realizing how individuals feel in regards to specific themes, can be utilized for grouping task. The sentiments of individuals can be sure, negative or nonpartisan. A notion is regularly spoken to in inconspicuous or complex courses in content, portrayal of conclusion as a content is an unobtrusive or troublesome assignment, so online clients can lean toward wide varieties to express his or her feelings.

Here, we delineated how estimation investigation can help in language learning, in animating the instructive procedure by utilizing Naive Bayes Classifier [1]. Estimation examination is treated as a characterization task

as it sort content into positive or negative. This paper gives test results by utilizing Support Vector Machine (SVM) to prepare an assumption classifier on benchmark datasets. Supposition investigation is treated as a grouping task as it characterizes the introduction of a content into either positive or negative.

This work presents the investigation utilizing AI open source information mining programming apparatus. In writing overview explores no single device or procedure that dependably achieves the best result [2]. Social Network Sites (SNSs) play important roles in people's lives for sharing information. Face book becomes very common platforms for sharing. In this people have their own accounts to comments, like, express feelings and emotions via texts as well as emoticons. When a certain issue is discussed, monitoring such information becomes difficult since there are too many suggestions and the problems usually tend to be overlooked. This research is significant for the stakeholders such as administrators and businessmen to monitor any discussion to enhance their service [3]. The tremendous growth in social networking site web contributes vast amount of user data such as customer reviews, comments and opinions. This data contains issues related to products, people, events, etc. This information is very helpful for businesses, governments and individuals. While analyzing this huge used data is hectic and time consuming. Hence there is requirement to develop system which dynamically mine such large content and identify them as positive, negative and neutral category. The aim of this paper is to apply the concept of Sentiment Analysis in the field of Natural Language Processing, and depicts a comparative study of techniques in this field [4].

This paper manages conclusion examination in content records, particularly message valence discovery. The proposed strategy utilizes Support Vector Machine as classifier. This classifier is prepared with colossal measure of information and investigated with troublesome word examination. For this reason conveyed learning on 112 processors are utilized. Datasets utilized for preparing and testing are accumulated from genuine client criticism on items from different sites. The proposed arrangement has been assessed with wide assortment of dialects, for example, English, German, Czech and Spanish, and so on. This paper upgrades exactness accomplished with the Big Data approach by 11%. The best precision acquired in this work is 95.31% for acknowledgment of positive and negative content valence.

The depicted learning is completely versatile, can be connected to any language and no consistence preprocessing is required [5]. Online journals and interpersonal organizations have as of late turned into an imperative hotspot for mining notions in fields as assorted as client relationship the executives, general conclusion following and content sifting. Actually investigation accumulated from interpersonal organizations, for example, Twitter and Face book has been appeared truly significant to promoting research organizations, popular assessment associations and other content mining elements. We utilized a specialist predefined dictionary including around 6800

seed descriptive words with realized introduction to lead the examination. Our outcomes recognized positive customer estimation towards a few renowned brands. By utilizing both a subjective and quantitative strategies to break down brands' tweets, this data gives broadness and profundity to the dialogs over demeanors over cosmopolitan brands [6]. Late couple of years consider exhibited a couple of undertakings are submitted for mining appraisals and presumption thusly from customary language in online messages, news and business thing reviews.

In this paper, we have presented suspicion presentation by considering the positive and negative ends using film customer reviews. By applying the methodology Naive Bayes' classifier, we have associated the supposition examination on the reviews using the figurings like Naive Bayes, Linear SVM and Synthetic words. Our preliminary outcomes portrays that the Linear SVM has given the best exactness which is trailed by the Synthetic words approach [7]. The endeavor of file named by their substance is known as substance arrangement. Various investigations have done to improve content request by adding establishment figuring out how to the report using data stores like Word Net, Open Project Directory (OPD), Wikipedia and Wikitology. In past work, people have done exhaustive examinations by removing data from Wikitology and surveying the test on Support Vector Machine with 10-overlay cross-endorsements. The tried results clearly show Wikitology is far predominant than other data bases. In this paper we are differentiating Support Vector Machine (SVM) and Naïve Bayes (NB) classifiers under substance upgrade through Wikitology [8].Idea map is a realistic device which portrays a coherent structure of learning as associated ideas.

Numerous individuals create and apply idea maps for arranging, learning portrayal or assessment instrument, and store them in an open vault. In such circumstance substance and nature of these maps may differ. The most essential highlights are separated are utilized for learning and characterization. Preparing and arrangement are performed utilizing naive Bayes and SVM classifiers. Accomplished outcomes are promising, and with further information preprocessing and modification of the classifiers we think about that they can be improved [9]. This paper investigates a short audit on assumption examination. To mine the sentiment on the web, it is basic to play out an all around characterized task, to recover the data from the information dwells on the web. We have begun our exchange with the presentation on supposition investigation, which gives us knowledge into opinion examination.

The detail talk on different techniques proposed by various analysts is additionally displayed. Diverse sorts of estimation investigation procedures provide an exploration guidance in various ways. At last a strategy is proposed dependent on the naïve bayes classifier [10].

III.PROBLEM STATEMENT

Face book winds up celebrated long range informal communication site that shapes critical stages for association and correspondence. It enables individuals

to have their own records to express sentiments and pass on

feelings by means of writings and in addition emojis. In this paper we give proficient and enhanced characterization methods for slant investigation of Facebook statuses and characterizing the content as either positive or negative or impartial.

The creators gave a decent overview of different methods created in on the web sentiment analysis. It covers concept of emotion in written text (appraisal theory), various methodologies which can be broadly divided into two groups: (i) symbolic techniques that focuses on the force and direction of individual words (the so-called "bag of words" approach), and (ii) machine learning techniques that characterizes vocabularies in context.

In view of the study, the creators found that emblematic systems accomplishes exactness lower than 80% and are commonly poorer than machine learning strategies on motion picture survey supposition investigation. Among the machine learning strategies, they thought about three directed methodologies: Support Vector Machine (SVM), Naive Bayes Multinomial (NBM), and most extreme Entropy (Maxent). They discovered that every one of them convey similar outcomes on different element extraction (unigrams, bigrams, and so on) with high exactness at 80%~87%.

V. PROPOSED SYSTEM

The express focuses of our examination are:

- ➢ To develop our own one of a kind corpus through a Face book application.
- ➢ To genuinely set up the framework to acknowledge contributions to the type of notices from the corpus, insulting updates that don't contain words or face emoticons.
- ➢ The limit of the framework to arrange the furthest point of a supposition for every declaration premise, in the midst of the testing stage

IV.EXPERIMENTAL SETUP

The following steps are

- ➢ Preparing dataset for the Facebook statuses.
- ➢ Importing dataset to Eclipse through java.
- ➢ Analyzing the class label and train using Naive Bayes Classification technique.
- ➢ Calculating independent probability for different class labels.
- ➢ Testing the dataset.
- ➢ Measuring the efficiency using performance metrics.

Positive	0.5617978
Negative	0.2696629
Neutral	0.1685393

Fig 3: Independent Probability

A	B	C

To develop a framework that can be organize an inclination using sentence-level request whether it includes positive, negative or fair emotions. Suppositions are as declarations in the long range relational correspondence site Face book.
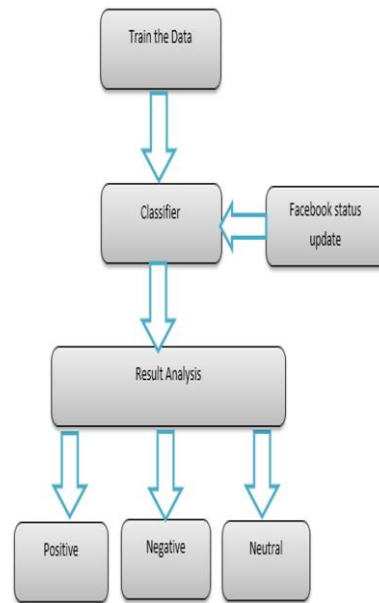

Fig 1: Flow diagram

- ➢ Training and testing the dataset using SVM classification technique
- ➢ Calculating the efficiency.
- ➢ Comparing the result with Naive Bayes classification technique

VI. RESULTS AND DISCUSSIONS

In this we have followed two approaches. The first approach is using the "Naive Bayes" method and the other approach using the "Support Vector Machine".

OUTPUT ANALYSIS OF NAIVE BAYES CLASSIFIER:

Fig2: Training Dataset

11	0	3
0	14	9
0	13	36

Fig 7: Performance Metrics

Fig 4: Probability of a Word

A	B	C
9	0	0

Fig 5: Confusion Matrix

Correctly Classified Instances	53	60.92%
Incorrectly Classified Instances	34	39.08%
Mean Absolute Error	0.34	
Root Mean Squared Error	0.438	

Fig 6: Measures
a= Neutral; b= Negative; c= Positive

PERCENTAGE OF CORRECT: 90.6

Output Analysis:

	Precision	Recall	F-Measure
Weighted Average	1	0.786	0.88
	0.519	0.609	0.56
	0.75	0.735	0.742
	0.729	0.709	0.716

Methods	Accuracy
Naïve Bayes Classifier	70.9
Support Vector Machine	**90.6**

Confusion matrix in Table shows the concept of TP, TN, FP and FN in mass localization.

Correctly Classified Instances	49	45.3704%
Incorrectly Classified Instances	59	54.6296%
Mean Absolute Error	0.4262	
Root Mean Squared Error	0.4617	

PERCENTAGE OF CORRECT: 70.9
Fig 8: Confusion Matrix

Output Analysis of Support Vector Machine

0	1	22
0	0	49

Fig 9: Measures

	Precision	Recall	F-Measure
Weighted Average	1	1	1
	1	0.043	0.083
	0.69	1	0.817
	0.813	0.728	0.629

Fig 10: Performance Metrics

Analysis is generally measured using following parameters.
TP: True Positive: Positive correctly recognized as Positive.
FP: False Positive: Negative incorrectly recognized as Positive.
TN: True Negative: Negative correctly recognized as Negative.
FN: False Negative: Positive incorrectly recognized as Negative.
Accuracy denotes the proportion of the correct result.
Accuracy $(AC) = (TP+TN)/ (TP+FP+TN+FN)$
Several standard terms have been defined for the 2 class confusion matrix:

Accuracy:
The Accuracy is used to calculate the overall correctness of the model and it is calculated as the ratio of the sum of correct classifications and the total number of classifications, as determined using the equation:

$$Accuracy\ (AC): \frac{TP+TN}{TP+FP+TN+FN}$$

Precision: The Precision is the proportion of the positive cases that were predicted correctly, and is calculated using the equation:

$$Precision = \frac{TP}{TP+FP}$$

Recall: Recall is the fraction of relevant instances that are retrieved.

$$Recall = \frac{TP}{TP+FN}$$

F-Measure: F-measure is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score:

VII.ANALYSIS AND CALCULATIONS

$$F-Measure = \frac{2*Precision*Recall}{Precision+Recall}$$

Fig 11: Precision Performance

Naïve Bayes	SVM
1	1
0.519	1
0.75	0.69

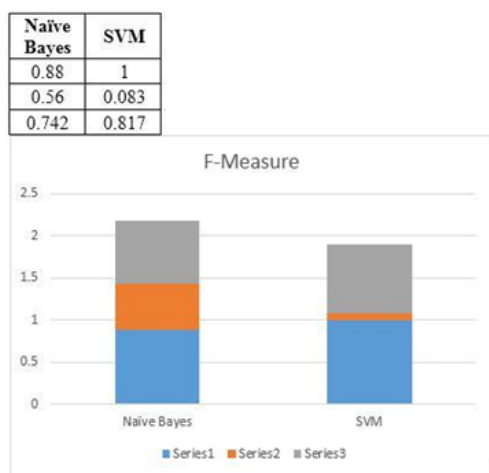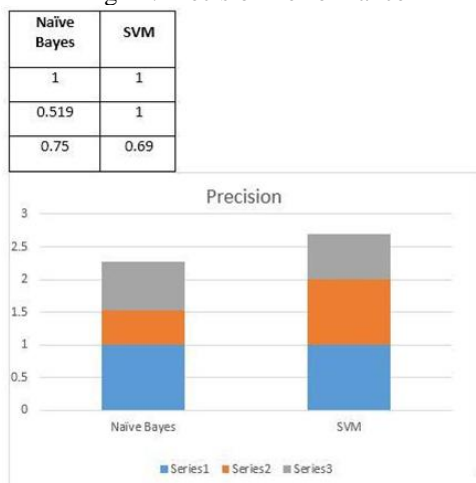| Naïve Bayes	SVM
0.88	1
0.56	0.083
0.742	0.817

Fig 13:F-Measure Performance

VIII.CONCLUSION & FUTURE SCOPE:

In this paper ,we present the possibility of assessment investigation of facebook statuses utilizing Naive Bayes Classifier and Support Vector Machine(SVM).We additionally exhibited a valuable component that give a critical increase over a unigram baseline. Further we investigated distinctive information portrayals and recorded enhancements over the unigram models. The information utilized is an irregular inspecting of spilling face book statuses, which are not gathered by explicit queries. The size of our manual marked information licenses us to perform cross approval and think about execution of the classifier crosswise over folds. Finally, we delineated exploratory outcomes which demonstrate the exactness in dissecting the wistful territory of Face book clients utilizing Support Vector Machine, is truly high.

Further our work can be improved by following changes of individuals' estimation on explicit subject and the time reliance of our information can be investigated to examine their patterns it might additionally gives intriguing results, if we think about the transient highlights on this examination and not to concentrate exclusively on past posts or exchanges.

REFERENCES:

1. Troussas, Christos, et al. "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning." Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on. IEEE, 2013.
2. Vohra, S. M., and Jay Teraiya. "A comparative study of sentiment analysis techniques." Journal JIKRCE 2.2 (2013): 313-317.
3. Zamani, N. AZMINA M., et al. "Sentiment Analysis: Determining People's Emotions in Facebook." (2013): 978-960.
4. Zainuddin, Nurulhuda, and Ali Selamat, "Sentiment analysis using support vector machine." Computer, Communications, and Control Technology (I4CT), 2014 International Conference on. IEEE, 2014.
5. Povoda, Lukas, Radim Burget, and Malay Kishore Dutta, "Sentiment analysis based on Support Vector Machine and Big Data", Telecommunications and Signal Processing (TSP), 2016 39th International Conference on, IEEE, 2016.
6. Hassan, Sundus, Muhammad Rafi, and Muhammad Shahid Shaikh. "Comparing svm and naive bayes classifiers for text categorization with wikitology as knowledge enrichment." Multitopic Conference (INMIC), 2011 IEEE 14th International. IEEE, 2011.
7. M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," Expert Systems with Applications, vol. 40, no. 10, pp. 4241 – 4251, 2013.
8. Rana, Shweta, and Archana Singh. "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques." Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on. IEEE, 2016.
9. B S Kumar., Karthik, S., & Arunachalam, V. P. (2018). Upkeeping secrecy in information extraction using 'k' division graph based postulates. *Cluster Computing*, 2018, Pages:1-7 https://doi.org/10.1007/s10586-018-1705-2
10. Zubrinic, Krunoslav, Mario Milicevic, and Ivona Zakarija. "Comparison of Naıve Bayes and SVM Classifiers in Categorization of Concept Maps." (2013).
11. Raghuvanshi, Neha, and J. M. Patil. "A brief review on sentiment analysis." Electrical, Electronics, and Optimization Techniques(ICEEOT),International Conference on. IEEE, 2016.
12. BalaAnand, M., Karthikeyan, N. & Karthik, S." Designing a Framework for Communal Software: Based on the Assessment Using Relation Modelling", Int J Parallel Prog (2018). https://doi.org/10.1007/s10766-018-0598-2
13. M.BalaAnand, S.Sankari, R.Sowmipriya, S.Sivaranjani "Identifying Fake User's in Social Networks Using Non Verbal Behavior", International Journal of Technology and Engineering System (IJTES), Vol.7(2), pg:157-161.
14. Maram, B., Gnanasekar, J.M., Manogaran, G. et al. SOCA (2018). https://doi.org/10.1007/s11761-018-0249-x
15. M. BalaAnand, N. Karthikeyan, S. Karthick and C. B. Sivaparthipan, "Demonetization: a Visual Exploration and Pattern Identification of People Opinion on Tweets," 2018
16. International Conference on Soft-computing and Network Security (ICSNS), Coimbatore, India, 2018, pp. 1-7. doi: 10.1109/ICSNS.2018.8573616
17. K. Anupriya, R. Gayathri, M. Balaanand and C. B. Sivaparthipan, "Eshopping Scam Identification using Machine Learning," 2018
18. International Conference on Soft-computing and Network Security (ICSNS), Coimbatore, India, 2018, pp. 1-7. doi: 10.1109/ICSNS.2018.8573687.
19. CB Sivaparthipan, N Karthikeyan, S Karthik "Designing statistical assessment healthcare information system for diabetics analysis using big data" Multimedia Tools and Applications, 2018
20. Zemedkun Solomon, C.B. Sivaparthipan, P. Punitha, M. BalaAnand, N. Karthikeyan "Certain Investigation on Power Preservation in Sensor Networks" ," 2018 International Conference on Soft-computing and Network Security (ICSNS), Coimbatore, India, 2018, doi: 10.1109/ICSNS.2018.8573688