PERFORMANCE COMPARISION OF IMPACT ON USERS OF SOCIAL MEDIA USING CLASSIFICATION ALGORITHM

Mrs. R. Srividhya
Assistant Professor

Ms. R. Suganya
Research Scholar

School of IT and Science
Dr. G. R. Damodaran College of Science

ABSTRACT

Social networking is a networking site that plays a vital role in modern life. Among them facebook plays an important role in modern world. Facebook a social networking sites that allow users to keep in contact with each other globally and helps to reconnect the old friends and allow them to share their views. Classification an supervised algorithm that is used to classify the comments. The comments may be positive, negative and neutral. The study is to investigate how people are accessing the Facebook and to review the views. In this study the articles posted are education, entertainment, sports, and current affairs. The number of users having shared a news article on Facebook, the number of likes given for the article, and comments for the articles are to be examined. The datasets have been taken for positive, negative and neutral comments. Linguistic analysis is applied to the comments of the users to reviews to know the different words are used to represent the positive aspects, negative aspects and neutral aspects. The system has the capacity to analyses the linguistic scenario with in a sentence, which has different meaning for a same word expressed in different situations with different word combinations. Depending upon on the age group the comments for the topics are gathered. The mode of accessing the facebook also taken as the attribute for doing the classification. With the continuation of the existing work, the attribute age and the mode of device are taken for performing the accuracy of the work. The accuracy is calculated using the naive bayes algorithm and the artificial neutral network algorithm.

KEYWORDS: Social Networking, Facebook, Classification Algorithms.

1. INTRODUCTION

Data mining is the process of extracting patterns from data. Data mining in general is the search for hidden patterns that may exist in large databases. Data mining is the process of extracting patterns from data. Data mining in general is the search for hidden patterns that may exist in large databases. Data mining is the process of extracting patterns from data. Data mining in general is the search for hidden patterns that may exist in large databases. In social network, Facebook is an important media which is used for fast communication between large numbers of users. Facebook is the world's fastest online growing social network, with 1.18 billion monthly users as of August 2015. Social networks are places where people with same interests can get together and

come to know one another on the internet. Web mining is the application of data mining techniques to discover patterns from the web. Web data mining is a technique used to crawl through various web resources to collect required information, which enables an individual or a company to promote business, understanding marketing dynamics, new promotions floating on the internet. Web mining is of three types. They are:

1. Web usage mining
2. Web structure mining
3. Web content mining.

A social network is used to construct the useful pattern which explains the relationship among groups, organizations and entire societies. The applications are such as education, health care, politics, entertainments ports and business.

2. LITERATURE SURVEY

Dudhat Ankitkumar et al., [8] defined Sentiment analysis (SA) is machine learning scheme in which machine examines and categorizes the humans reactions, feelings, and judgments about some topic which are expressed in the form of either manuscript or language. This scenario targets to decide the approach of a presenter with respect to any subject or in general background division of a text. Numerous modern algorithms and diverse sentiment analysis applications are discussed in this scenario. The most efficient classification algorithm of naive bayes is used to compute the label features based on the probability of class models. This algorithm selects the features or attributes which satisfy the conditions and it is further used to analyze the unsupervised features easily.

Narayanan et al., [18] involves that an extremely correct and speedy sentiment classifier which can be constructed by using a naive bayes model that has linear training and testing time complication. Thus this approach has an accuracy of 88.80% on the popular IMDB film reviews dataset. The proposed technique is universal to a number of text classification problems for increasing speed and accurateness. Naive Bayes is a very easy probabilistic model that aims to work well on content classifications and typically considers orders of magnitude less time.

Suruthi et al., [19] studies are typically achieved in machine learning methods which are less precise than neural network methodologies. This research scenario is depends on sentiment categorization using competitive layer neural networks and categorizes the polarity of a specified content whether the expressed judgment in the content is positive or negative or neutral. It discovers the general theme of the specified text. A hybrid scheme using n gram analysis and dynamic artificial neural network builds numerous offerings to twitter opinion analysis, established through application on an amount of tweets.

Soni, Sneha [20] has stated the Prediction stock price or economic marketplace is one of the major issues to the artificial intelligence community. Different industrial, essential, and numerical indicators is enhanced and utilized along with unreliable consequences. Though, none of these methods or mixture of methods is successfully sufficient. The purpose of

classification approach is mostly ahead of the potential of conventional artificial intelligence research which is mainly targeted on improving intelligent systems.

Facebook API [17] is used to describe the creation of application from scratch, Facebook functionalities, integration test and other test strategy. The Facebook app is easy for usage purpose and fast for communication around the world. Thus it simple to access, easy to store and efficient retrieval is occurred in this application. It maintains everything as simple in this application package. This application utilizes java script to make a facebook API calls. It has a web service which encloses all business logic. It is used to send all kinds of information to and from UI also to/from databases through http GET and POST calls. It uses automated integration tests to build assured and complete application works correctly.

3. PROBLEM DESCRIPTION

Social networking a wide variety of networking site that allow users to keep in contact with other users or individuals globally. Among many social networking sites, Facebook is used in this work. All the comments, likes, shares of the users for the posts posted are gathered. The posts are related to education, entertainment, sports and current affairs. Classification algorithm is used to predict the comments. The comments can either positive, negative and neutral. Using the classification algorithm the comments are classified. Depending upon the age group the comments have been classified according to the category of articles posted. Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing the set of attributes and the respective outcome, usually called prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. The algorithm analysis the input and produces a prediction. With the continuation from the existing work the attribute age and the mode of accessing the facebook (website, Smartphone or tablet) are included to improve the performance accuracy. To improve the accuracy, the classification algorithms namely naive bayes and artificial neural network algorithms are used. With the additional of two attribute (age and mode of accessing the facebook) a comparative study is done for two classification algorithms, for naive bayes and artificial neural networks.

4. IMPLEMENTATION

The below figure 1 is the model Facebook account created for my research work. The Facebook account is created with different steps provided by the Facebook. The account holder name is Suganya. R and the account have a unique id and password. Many people are made as friends, friends are both known and unknown through the created Facebook account. The friends are from different age group varying from 18-40. All the articles are posted in the Facebook account are related to education, entertainment, sports, and current affairs.

Figure 1 A Model Facebook Account

The below figure 2 is the opening page of Facebook developer. Facebook developer website was created for developers who are working on the Facebook platform. It is for third party developers to create their own applications and services that access data in facebook. It was launched in the year 2010. It provides documentation on services and features provided by Facebook. Using this can learn how to build, grow, and monetize the app or website with the Facebook. They are different types of developer with the Facebook. They are IOS developer, Android developer and the web developer.

Figure 2 Facebook Developer

The below figure 3 is getting the access token and UID number of the facebook account. One must know the access token and UID for the Facebook account to extract the data of the account. Data extraction is the process extracting the data that are required for the work. Give Me My Data is a Facebook application that helps users export their data out of Facebook for reuse in visualizations, archives, or any possible method of digital storytelling. Data can be exported in common formats like CSV, XML, and JSON as well as customized network graph formats.

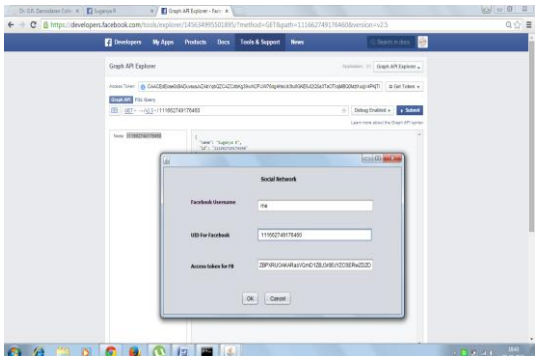

Figure 3 Access Token And UID Number

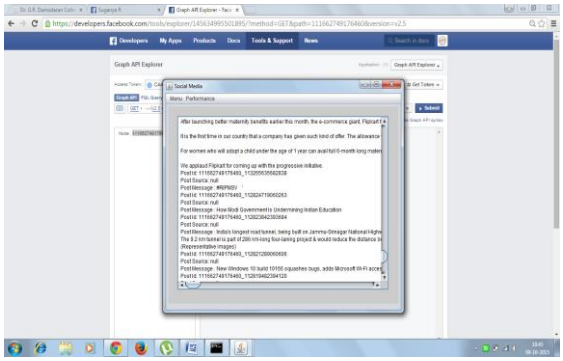

Figure 4 Data Extraction

The above figure 4 is extracting the data. Data extraction is the process or act of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage. Data processing is done, which involves adding metadata and other data integration, another process in the data workflow. The majority of data extraction comes from unstructured data sources and different data formats. This unstructured data can be in any form, such as tables, indexes, and analytics.

From the figure 5 and 6 the comparisons of the existing and proposed neural networks are shown. The articles based on education, entertainment, sports and current affairs are posted. All the comments are classified into positive, negative, and neutral comments based on the age group. For classification of the comments the classification algorithm are used.

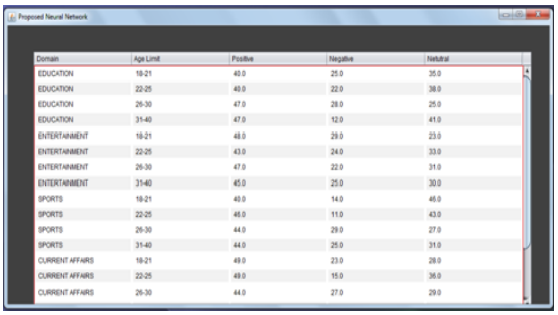

Figure 5 Proposed Neural Network

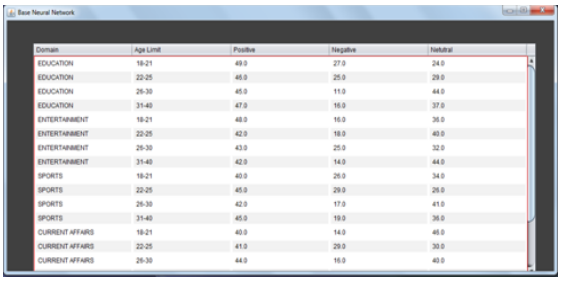

Figure 6 Base Neural Network

The calculation of the precision value is done as follows

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

The below graph figure 7 describes that the existing and proposed systems are analyzed using naive bayes as well as ANN algorithms. In x axis the methods are plotted and in y axis the precision ratio is plotted. The existing naive bayes and ANN algorithm have shown the lower precision values for dataset reviews. The proposed naïve bayes and ANN algorithm shown the higher precision values. But the proposed ANN algorithm is superior to proposed naïve bayes algorithms. Finally the conclusion decides that the proposed ANN provides higher performance than other algorithms.

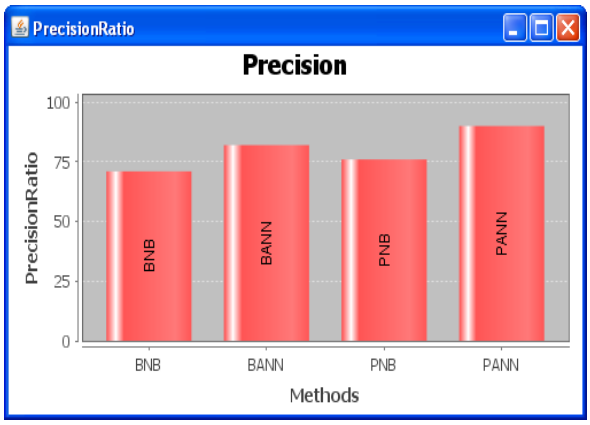

Figure 7 Precision Comparison

The calculation of the recall value is done as follows

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

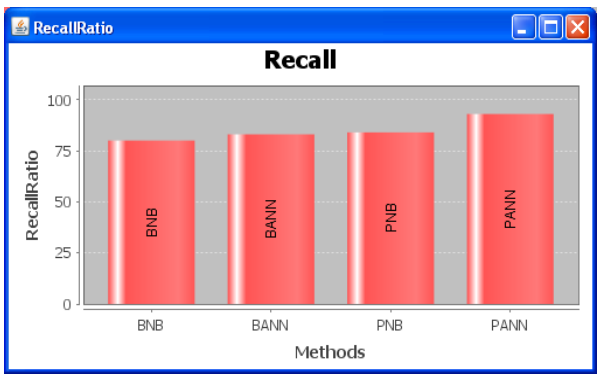

Figure 8 Recall Comparison

The above graph figure 8 the recall ratio values are compared with the naive bayes and artificial neural networks algorithms. In X axis the methods are plotted and in Y axis the recall ratio is plotted. When compared with the naive

bayes and artificial neural networks, the PANN algorithms shows the higher recall ratio of about 90%. The higher the ratio, the higher the performance results.

From the below graph figure 9 describes that analysis of post methods using website, mobile and tab. In x axis the methods are taken and for y axis the number of posts taken. The website has shown the 80% of posts, mobile shows the 20% of posts and tab shows 3% of posts.

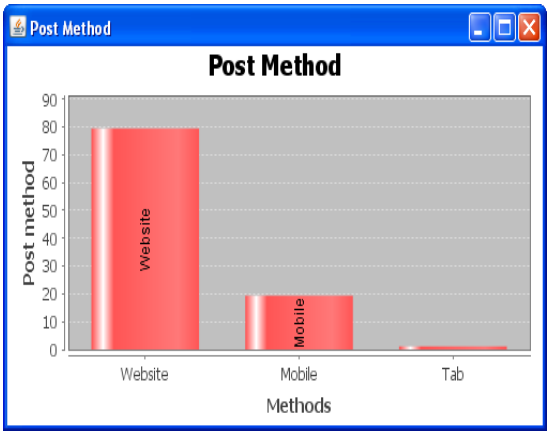

Figure 9 Post Method

The harmonic mean of precision and recall is F-Measure, the traditional F-measure or balanced F-score

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

There are several reasons that the F-score can be criticized in particular circumstances due to its bias as an evaluation metric. This is also known as the F_1 measure, because recall and precision are evenly weighted.

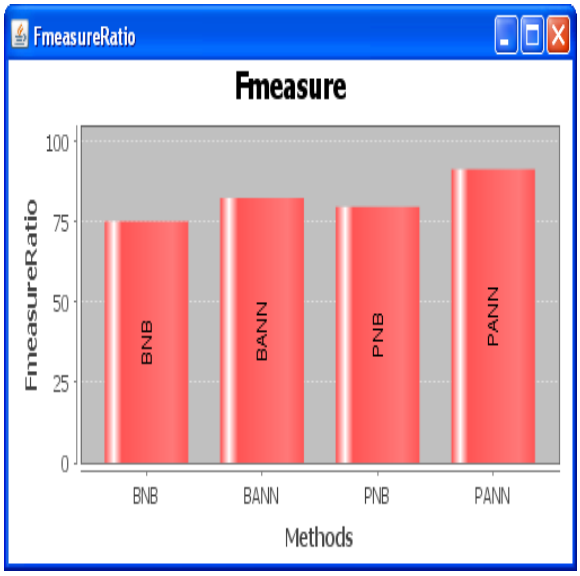

Figure 10 F-Measure comparison

The above graph figure 10 describes that the existing and proposed systems are analyzed using naive bayes as well as ANN algorithms. In X axis the methods are plotted and in Y axis the f-measure ratio is plotted. The existing naive bayes and ANN algorithm have shown the lower f-measure values for dataset reviews. The proposed naive bayes and ANN algorithm have shown the higher f-measure values. From the result, the proposed ANN algorithm is superior to proposed naive bayes algorithms.

CONCLUSION

Analyzing the social networking in day to day life is a tedious work. Because every seconds the amount of data goes on increasing. Although, the different kinds of domain such as education, entertainment, healthcare, sports are considered for evaluation. In the existing scenario, the algorithms are introduced named the naive bayes algorithm and artificial neural network algorithm to analyze and classify the review comments. The above mentioned algorithms analyze all the input

samples and predict the positive, negative and neutral reviews. However it has issue with huge number of attributes, age and advanced devices. To overcome this scenario, the proposed system encloses the tablets, websites and smart phones as devices. It analyzes and shows the number of posts for each device. The articles based on education, entertainment, sports and current affairs are posted. The people give their own comments for the articles posted. All the comments are classified into positive, negative, and neutral comments based on the age group.

It is observed from the experimental results the naive bayes and ANN algorithms are shown the higher precision, recall and f-measure values after adding the two attributes with the existing system. The proposed ANN is superior performance for all metrics than other algorithms. From the result, the proposed ANN (PANN) is better for providing efficient system performance.

FUTURE WORK

In future, the advanced algorithms will be developed such as optimization algorithms for executing the reviews. It handles extremely high dimensional dataset and analyzes all the large input samples with less execution time. The optimization algorithms are such as particle swarm optimization and multi objective genetic algorithm to analyze the extremely high dimensional dataset. It will select the global optimal features to model the label for huge volume of social network. It also captures the heterogeneous data sources and analyzes optimally the positive, negative as well as neutral comments. Furthermore, security is a major issue which can be considered to research as future work. Since several communities stimulate user for this kind of inappropriate actions and these communities has almost spoiled the image of social networks.

5. REFERENCES

1. Baruah, Trisha Dowerah. "Effectiveness of Social media as a tool of communication and its potential for technology enabled connections: A micro level study." International journal of scientific and research publication 2 (2012): 5.

2. Bennur, z. behavioral consequences of the third-person effect on turkish voters. diss. middle east technical university, 2008.

3. Corritore, Cynthia L., Beverly Kracher, and Susan Wiedenbeck. "On-line trust: concepts, evolving themes, a model." International Journal of Human-Computer Studies 58.6 (2003): 737-758.

4. Crespo, R. G., Martínez, O. S., Lovelle, J. M. C., García-Bustelo, B. C. P., Gayo, J. E. L., &Pablos, P. O. d. (2011). Recommendation system based on user interaction dataapplied to intelligent electronic books. Computers in Human Behavior, 27(4),1445–1449.

5. Cyr, D., & Head, M. (2013). Website design in an international context: The role ofgender in masculine versus feminine oriented countries. Computers in

HumanBehavior, 29(4), 1358–1367.

6. DeLorme, D. E., Huh, J., & Reid, L. N. (2006). Perceived effects of direct-to-consumer(DTC) prescription drug advertising on self and others: A third-person effectstudy of older consumers. Journal of Advertising, 35(3), 47–65.

7. Dhande, Lina L., and Girish K. Patnaik. "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier."

8. Dudhat Ankitkumar, M., R. R. Badre, and Mayura Kinikar. "A Survey on Machine Learning Techniques in Sentiment Analysis."

9. Gurkhe, Dhiraj, Niraj Pal, and Rishit Bhatia. "Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification." International Journal of Computer Applications 99.13 (2014): 1-4.

10. Hamouda, Safa Ben, and Jalel Akaichi. "Social networks' text mining for sentiment classification: The case of Facebook'statuses updates in the 'Arabic Spring'era." International Journal Application or Innovation in Engineering and Management 2.5 (2013): 470-478.

11. https://www.redhat.com/sumit/2012/pdf/2012-DevDay-ZeroToFacebookIn60Min.pdf

12. Jagtap, Balaji, and Virendrakumar Dhotre. "SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment."

13. Kamruzzaman, S. M. "Web page categorization using artificial neural networks." arXiv preprint arXiv:1009.4991 (2010).

14. Kotsakis, R., Kalliris, G., &Dimoulas, C. (2012). Investigation of broadcast-audiosemantic analysis scenarios employing radio-programme-adaptive patternclassification. Speech Communication, 54(6), 743–762.

15. Lance, P., & Guy, J. G. (2006). From subservient chickens to brawny men. Journal of Interactive Advertising, 6(2), 4–33.

16. Meyrick, J. (2001). Forget the ''blood and gore'': An alternative message strategy tohelp adolescents avoid cigarette smoking. Health Education, 101(3), 99–107.

17. Ms. Gaurangi Patil, Ms. Varsha Galande, Mr. Vedant Kekan, Ms. Kalpana Dange, Sentiment Analysis Using Support Vector Machine, in International Journal of Innovative Research in

Computer and Communication Engineering, 2014.

18. Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." Intelligent Data Engineering and Automated Learning–IDEAL 2013. Springer Berlin Heidelberg, 2013. 194-201.

19. S.Suruthi, M. Pradeeba, A.Sumiya "Neural Network Based Context Sensitive Sentiment Analysis" ,International Journal of Computer Applications Technology and Research Volume 4– Issue 3, 188 - 191, 2015, ISSN:- 2319–8656.

20. Soni, Sneha. "Applications of ANNs in stock market prediction: a survey."International Journal of Computer Science & Engineering Technology 2.3 (2011): 71-83.

