

# Social Media Analysis in Sentiment Instance Selection and Classification Based on FFNN

Manpreet Kaur

M.Tech(Research Scholar) Department of Computer Science and Engineering  
Sachdeva Engineering College for girls,Gharuan  
mniechahal@yahoo.in

Rakesh Kumar

Principal & Professor Department of Computer Science and Engineering  
Sachdeva Engineering College for girls,Gharuan  
rakesh77kumar@gmail.com

Nidhi Bhatla

Head of Department Department of computer Science and Engineering  
Sachdeva Engineering College for girls,Gharuan  
engineernidhi@yahoo.com

**Abstract – The sentimental analysis is the major concept of the analysis emerged with the advance of social media. Our purpose is to define the sentiment analysis of a Twitter and Facebook comment Positive, Negative and Neutral by using the Vectorization and instance selection approach. In the main sentiment analysis applications using these methods, the sentiment keywords, sentence plays a main role. It is most important to make a sentence and feature based covering various sentiment words. For the reason, we analyze the paper issues how to divide and list-words, sentences present on the focus into binary dictionaries. We have implemented feature extraction method, instance selection and a novel involuntary method to make the positive, negative and neutral dictionaries that search the sentiment keywords present in the comments. More significantly, our idea allows increasing these dictionaries with an en-richest phase. Lastly, by using these prepared datasets, we identify the category of the sentiments of the comments and reviews in social media networks. We consider our method by comparing to human authentication and classification. Our consequences are also real and consistent.**

**Keywords:** Sentimental Analysis, Classification, vectorization used for feature extraction, Instance selection based on the extracted features.

## 1. INTRODUCTION

Data mining, artificial intelligence and machine learning approaches were famous, consumer applied the use of a range spread of topics, ranging from classification of sounds to the identification of pen stroke on a hand-written assigned to employees. Easy to access the internet and computers, data are generated every day for investigators and engineers to study. In this area of natural language processing, while some use it to perform tasks such as extracted information, some use it for classification of sentiment alignment [1]. With the rapid development of the intranet, day by day people prefer to use network as the platform of expression of reviews, viewpoints and attitudes. The huge number of reviews or comments about people, products, movies and events appears on the internet [2]. If we could collect information about the reviews or comments of citizens and examine their sentiment analysis is called classification of sentiments.[3] The sentimental classification is diverse from the newest texture classification, which is based on the attributes of texture topics. Although the sentiment classification evaluate the sentiment analysis based on data in texture about viewpoints, preferences and attitudes. Both types of classification techniques use artificial intelligence and machine learning methods, which are based on static feature opinions, the opinion classification appears like a main branch of the sentence classification. This method normally needs lots explanation corpuses to teach a class. According to the recent scenarios of Social Media, Twitter and Facebook sentiment classification research, there is almost no authority resource of opinion researches [4].

In this paper, based on the Social Media i.e. Facebook and Twitter Review classification technology, we implement a Social reviews classify techniques by using English-sentiment resources, feature extraction, instance selection and classification[5]. This method evaluates and implement sentiment feature extraction in order to complete the training of the social classifier model. It is confirmed that when doing the Social sentiment classification, this type of classifier model has verify in some degree [6].

In Section 2 We discuss prior works on sentiment analysis. After that section 3 defines the classification of sentiment analysis at dissimilar levels. In Section 4 Highlights the existing issues in sentiment analysis . Next section 5 describes the proposed work with the help of flow charts. In Section 6 Represent the result of proposed work. In Section 7 Conclusion defines the percentage of accuracy and future scope.

## 2. RELATED WORK

Chu et al (2016) [1] the author did their research with a machine learning algorithm to find sentiment from movie reviews. As per previous researches in this field are related to support vector machine, but the author did their work on the basis of neural network and some vector representation. Proposed approach raises the accuracy to find sentiments from the input review. Author train their system with the help of some standard keywords. The proposed approach design a vector representation for classifier as a testing set. The proposed approach performs better than some traditional approaches like naïve bayes' and SVM. Ma et al (2015) [7] Author worked on sentiment analysis on cross-language datasets. Here the proposed approach used SVM (support vector machine) for classification of the sentiments. Proposed model uses some statistical method to find the uniqueness of input data. The processing extracts the features of English works and creates a dataset for training and testing with SVM. The word selection is based on some standards which use some weights to configure the dataset with classification process. The proposed method shows effective performance in the form of detection accuracy as compare to other existing approaches. Sahayak et al (2015) [8] discussed with existing study of twitter database with data mining method such as use of sentiment analysis method using artificial intelligence algorithms. Kiritchenko et al (2014) [9] Detecting motive behind political tweets, for feelings in text and toward recover the sentiment lexicons by producing them from larger quantities of data, and from dissimilar kinds of statistics, such as tweets, blogs, and Facebook posts. Jandail (2014) [10] to demonstrate the plausibility of the main idea through them using clustering and classifying opinion mining experiment on study of twitter blog posts on current product policy and services reviews . Virmani et al (2014) [11] implemented on online politics reviews and books the efficiency through kappa measure, which has a large accuracy of 97% and less error rate. Asghar et al (2014) [12] described previous methods used for extracting feature in sentiment analysis and opinion mining. They tried to verify the most and least normally used instance selection methods to search for future work. Pontik et al (2014) [13] ABSA problem has been formalized into a righteous unified framework in which all the well-known constituents of the conveyed opinions meet a set of conditions and are linked to each other within the tuple.

Fong S et al (2013) [14] author uses some machine learning algorithms to speed up the performance of the existing system for the sentiment analysis of online news using some existing development toolkit features as MALLET. The proposed approach based on some fact and rule based approaches which able to classify the news article as a lead role in this research. Author creates training sets for some classifiers to check the efficiency and compare their experimental results. Lin et al (2013) [15] the author chooses their research area in online book review classification based on their sentiments. Author used data clustering to enhance the experimental results. Proposed approach extracts features as uniqueness of input data and store in a dataset as training set for the classification. All the classification process is based on data clustering and some mining processes. The proposed approach uses some factors like similarity and raring to classify the sentiment of the online review. All over the results of this process provide enhanced results in the form some similarity factors .

Table 1: Summarized prior reviews

Year	Authors	Proposed algorithms	Result parameters
2016	Chun-Han Chu	Feature vector representation and neural network	Precision, Recall, Accuracy
2015	Hongxia Ma	Support Vector machine with some statistical feature extraction for English words	Accuracy
2013	Simon Fong, Yan Zhuang	Naïve Bayes, Maximum Entropy, Decision Tree, C4.5 Decision Tree, Winnow Balanced Winnow	ROC curve
2013	Eric Lin	hierarchical clustering, human domain knowledge	Similarity, Rating

### 3. CLASSIFICATION OF SENTIMENTAL ANALYSIS

Sentiment analysis can define at dissimilar levels [8,16]:

- Document level
- Sentence level
- Feature level.

**3.1 Document Level Classification:** In this procedure, feeling is extracted from the complete review and sentiment is classified based on the overall sentiment of opinion holder. The main goal is to classify a review negative, neutral and positive.

**3.2 Sentence Level Classification:** This approach has two types: (i) Subjective classification of a sentence into one of binary classes: objective and subjective. (ii) Sentiment classification of particular sentences into binary follows: Negative and Positive.

**3.3 Feature Level Classification:** This process goal is to search and extract entity features that have been commented on by the emotion holder and define whether the opinion is negative, positive and neutral. Feature similar meanings are grouped, and a featured based summary of multiple reviews is formed.

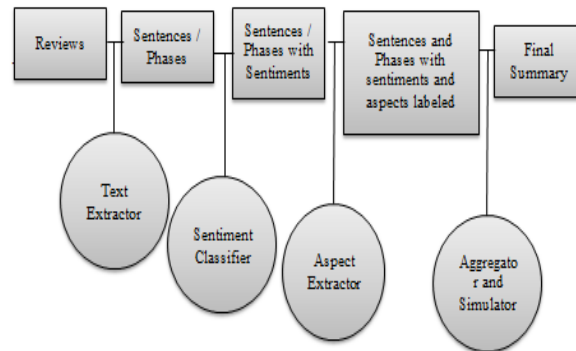


Fig.1 Types of sentiment Analysis

### 4. EXISTING ISSUES

The issue is unstructured, un-grammatical text. Because social messages are limited to 140 characters length, clients might have a tendency to use acronyms, emotions to shorten the text. Issues could lead to un-usual message. The second issue is the fact that tweet texts aren't always accurate. During speed typing, using phones is input devices, the client might have miss-typed message and make analysis step header [4].

The next issues are ambiguous due to less quality of information. It is difficult to verify the corresponding entity of interest.

The last issue we survey, concerns which concrete analyzing since human emotion is very diverse.

The trouble lies in the fact that there could be mixed feelings in a document, and with the inspired nature of natural languages, people might express the similar sentiment in vastly dissimilar ways [17].

### 5. PROPOSED WORK

In this section, we implement the proposed algorithm, i.e. feature extraction using vectorization, instance selection using genetic algorithm and classification using a feed forward neural network.

**Step1: Dataset:** Upload the social media three types of categories, i.e. Neutral, Negative and Positive. We create two databases, In first, we implement the single word stored in the database and In the second we store one complete word file or text file in the main database.

**Step2: Feature Extraction:** We implement the feature extraction approach by using vectorization. The vectorization process performs the binary conversion to read the machine learning. In feature extraction approach to identify the unique properties of the sentence formation in sentimental analysis. Feature extraction technique is used to recover most revealing terms from amount of matrix. This study used Component Analysis technique to calculate and study.

**Step3: Instance Selection:** Instance selection is a data optimization approach. Main concentrate of instance selection is to eliminate some malicious characteristics of a given data set. In transactions with selection of instances to optimize the size of the matrix and would easily process to deal with the further proceeding input. Genetic algorithm optimization is an instance based method which is used to optimize the instances of the sentiment words.

Step4: Classification: Network of Neural is a computational scheme inspired by the arrangement, dispensation technique, and knowledge ability of an organic brain. The essential dispensation rudiments of neural systems are named artificial neurons. It is a simplify arithmetical meld of the neuron. The effect of the synapses is represented by association weights that adjust the consequence of the linked input signal and amp; the nonlinear feature exhibit by neurons is represented by a transport occupation. The learn ability of an artificial neuron is attained by regulating the loads in agreement to the selected learning method.

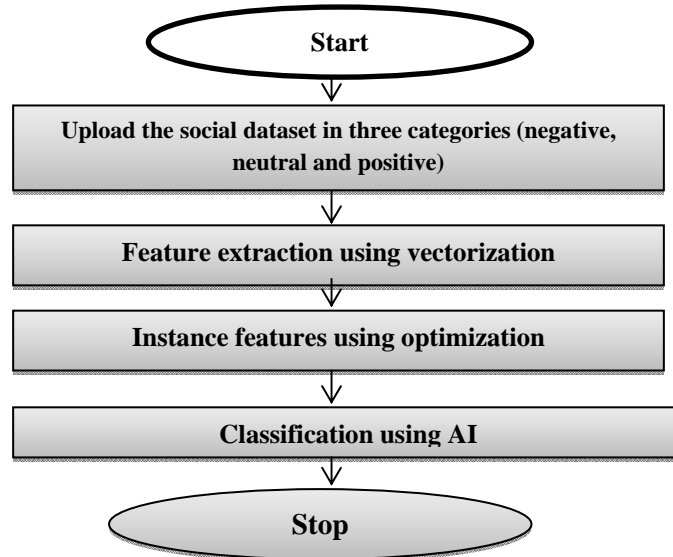


Fig 2. Proposed Flow chart

## 6. RESULTS AND DISCUSSIONS

In this section, we discussed the result explanation and comparison between proposed and previous methods.

Table 2. Word List containing Positive and Negative Adjectives

Type	Proposed Word Lists
Human 1	Positive: dazzling, brilliant, phenomenal, outstanding, fantastic negative: suck, terrible, dreadful, un watchable, hideous 12
Human 2	Positive: gripping, mesmerizing, riveting, remarkable, cool, awesome, exciting, excellent, moving, exciting negative: bad, clichéd, sucks, boring, stupid, slow

All the parameters show the efficiency of the system. FRR and FAR is used to find the rejection rate and acceptance rate that how correctly the system is accepting the right authentication and false authentication accurately and efficiently. The one other parameter, Accuracy is used to check the detection accuracy and processing the textual data over different knowledge base. The values of FRR and FAR need to be less and accuracy will be high in best solutions systems.

Table 3. False Acceptance Rate (Proposed Work)

Iterations	False Acceptance Rate
20	0.0003664
40	0.00077
60	0.001118
80	0.001523
100	0.00190

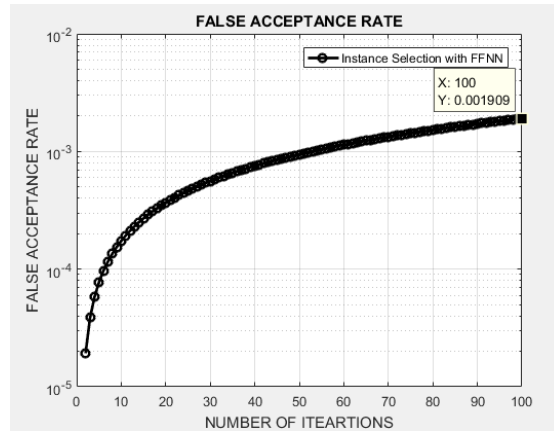


Fig 3. False Acceptance Rate (Proposed)

Fig 3. Shows that the false acceptance rate means wrong data will acceptable bases of number of iterations change output. We implement the number of iterations vs FAR performance parameters evaluated.

Table 4. False Rejection Rate (Proposed Work)

Iterations	False Rejection Rate
20	1.407
40	2.743
60	4.15
80	5.55
100	6.96

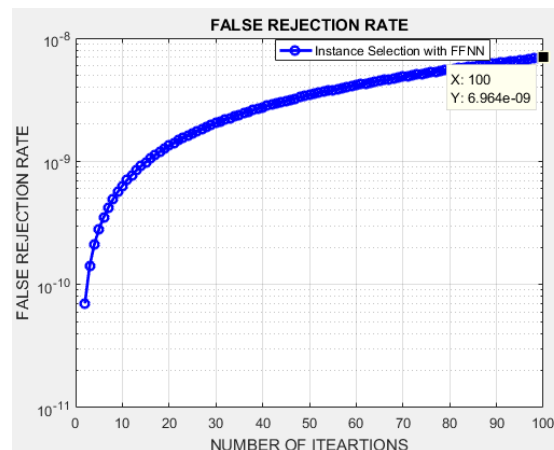


Fig 4. False Rejection Rate (proposed)

Fig 4. Shows the performance parameters means wrong data is reject able case to enhance the performance parameters based on percentage. We calculate the false rejection rate based on number of iterations. We decrease the rejection rate as well as acceptance rate.

Table 5. Accuracy in Proposed

Iterations	Accuracy
20	19
40	40
60	59
80	70
100	99

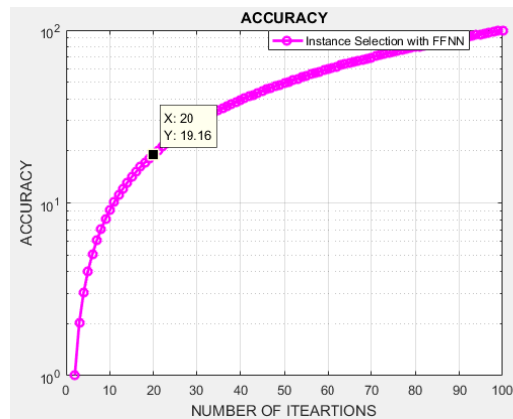


Fig 5. Accuracy in Proposed (%)

Fig 5. Represents that the decrease in rejection rate and acceptance rate improve the performance of accuracy in the proposed work.

Table 6. Performance parameters

Sentimental	Negative	Neutral	Positive	Accuracy
1	0.4805	0.5030	0.4978	0.4938
2	0.7773	0.7513	0.7850	0.7712
3	0.6755	0.6601	0.6623	0.6660
Base Work	0.76	0.780	0.7810	0.7552
Proposed Work	0.81	0.89	0.99	0.98

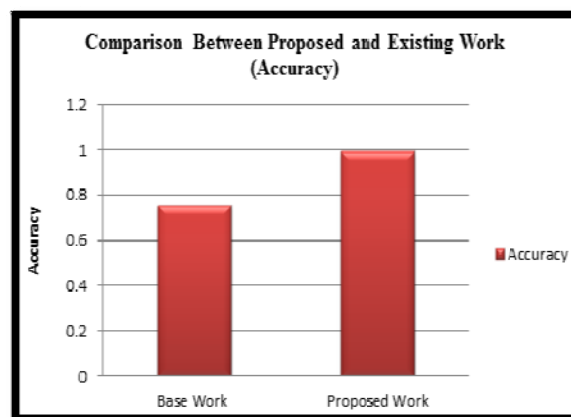


Fig 6. Comparison between Existing and Proposed Work Accuracy

Fig 6. Shows the comparisons between the proposed and existing work (Accuracy). We improve the accuracy, performance as compared to existing one, i.e. support vector machine.

### 7. CONCLUSION

In this research work, we implemented the feature extraction, instance selection and classification algorithms to resolve the issues like unstructured data, to analysis an online social media i.e comments, reviews and opinions. The basic definition of the sentiment analysis is to evaluate user vocabulary, sentence and feeling based feature extraction. Then laws are described as the normal constraints of human sentiments based on sentiment area. In this paper , we apply the matrix based technique for feature extraction means obtained the data type of the sentiment category and calculate the percentage of the sentiment . After feature extraction we implement the instance selection using genetic algorithm. The classification approach used to analyze the sentiment has shown the performance according to the iterations. This network creates three layers, i.e. input, hidden and output layer. Hidden layer transfer the data for activation function which has scanned or filter the information in the output layer and give them results. In testing section to evaluate the performance parameters like false acceptance and false rejection rate these are the errors to find the testing section the major focus to decrease the error rate as well as increase the accuracy. We detect the sentiment category with the help of classification approach FFNN. Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. From the above work, discussed with

the feature extraction, selection and classification algorithms achieve the 98% accuracy, as compared to existing ones ANN value was 75% accurate. The performance of the proposed approach is increased as compared to other approaches in the sentiment analysis. The fake reviews are decreased with the help of optimization and classification feature vector. In future, the performance also can be optimized with minimization of fake reviews using some other techniques like clustering, matrix based pre-processing. The clustered and prioritized vector can enhance the working of the proposed approach in future.

#### REFERENCES

- [1] Chu, C. H., Wang, C. A., Chang, Y. C., Wu, Y. W., Hsieh, Y. L., and Hsu, W. L. (2016). Sentiment analysis on Chinese movie review with distributed keyword vector representation. In technologies and applications of Artificial Intelligence.
- [2] Ma, H., Zhang, Y., And Du, Z. (2015). Cross language sentiment classification based on support vector machine. In Natural Computation.
- [3] Chatterjee, A., and Perrizo, W.(2016). Investor classification and sentiment analysis. In Advances in Social Networks Analysis and Mining.
- [4] Hussein, D. M. E. D. M. (2016). A survey on sentiment analysis challenges. Journal of King Saud University Engineering Sciences.
- [5] Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. International Journal of Computer Applications.
- [6] Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S. M., Ritter, A., & Stoyanov, V. (2015, June). Semeval-2015 task 10: Sentiment analysis in twitter. In Proceedings of the 9th international workshop on semantic evaluation
- [7] Ma, H., Zhang, Y., & Du, Z. (2015, August). Cross-language sentiment classification based on Support Vector Machine. In Natural Computation .
- [8] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment Analysis on Twitter Data. International Journal of Innovative Research in Advanced Engineering .
- [9] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research.
- [10] Jandail, R. R. S. (2014). A proposed Novel Approach for Sentiment Analysis and Opinion Mining. International Journal of UbiComp.
- [11] Virmani, D., Malhotra, V., & Tyagi, R. (2014). Sentiment Analysis Using Collaborated Opinion Mining. arXiv preprint arXiv.
- [12] Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A review of feature extraction in sentiment analysis. Journal of Basic and Applied Scientific Research.
- [13] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. Proceedings of SemEval.
- [14] Fong, S., Zhuang, Y., Li, J., and Khoury, R. (2013). Sentiment analysis of online news using MALLET. In Computational and Business Intelligence.
- [15] Lin, E., Fang, S., & Wang, J. (2013, March). Mining Online Book Reviews for Sentimental Clustering. In Advanced Information Networking and Applications Workshops IEEE.
- [16] Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. Wiley Interdisciplinary Reviews: Computational Statistics.
- [17] David, O. E., van den Herik, H. J., Koppel, M., & Netanyahu, N. S. (2014). Genetic algorithms for evolving computer chess programs. IEEE Transactions on Evolutionary Computation.