



## Sentiment Analysis of Technical Article Review Using Hybrid Approach

Babaljeet Kaur\*, Naveen Kumari

Computer Science Engineering, PURCITM,  
Mohali, India

---

**Abstract**— *Sentiment analysis is a major branch of natural language processing (NLP). It is a big problem to review on a product and technology due to large numerous new features comes every day in the technology. The sentiment analysis is the similar task to the opinion mining, which provides different reviews or opinions on text, product, article and news that can be positive and negative. The classification is done with the help of different NLP classification tools. People use social networking sites Twitter, Facebook, and other expert technology websites express their views on the wide variety of topics. In this paper, present work proposing a combined approach of SVM and K-Nearest Neighbor classification techniques to reviews of the technical article. This approach used to classify the reviews as positive and negative.*

**Keywords**— *Sentiment analysis, K-Nearest Neighbor, Support Vector Machine, Classification, NLP.*

---

### I. INTRODUCTION

The sentiment analysis is the process of identifying people's emotions and opinions. In natural language processing different machine learning approaches are used to determine the sentiments of a large amount of products, services, and text. Sentiment analysis provides a lot of facilities to implement new applications which have a major role in medical, data mining and evaluate the feedback of different products.

The reviews of people in sentimental analysis can be classified in various ways like a positive and negative opinion. Positive reviews have large polarity than the negative reviews. According to the reviews by experts, one can find out the quality of technique. The analysis of sentiment work up with few difficulties: Among other things, it must be resolved whether document or segment thereof is subjective or objective and whether the sentiment communicated is positive or negative. The sentiment of the content may be vital in fewer applications:

- Mining, customer, summarizing customer, book and film reviews.
- Analyzing political opinions.
- Classifying sites posts and comments.

The main steps included in the research are Pre-processing, analysis of data and comparison and evaluation of result. On websites, people share their views in the form of 'comments'. These reviews and comments are the main components, which determine the sentiment of people as these reviews as original. The sentiment analysis on the website is done in two steps: (a) Determine the sentiment expression. (b) Determining the polarity of reviews (positive, negative and neutral). (c) Combined approach is used to classify the sentiments on the website [15].

It has increased much consideration as of late and studies individual's emotions towards certain substances [6]. Sentiment analysis deal with many challenges. The description of element considered in this area. It is not a cup of tea. For example sentiments about any product can be determined by positive and negative opinions of people on the product. Sentiment analysis can help in defining the features of different web services, technologies and social sites. The information available on the blogs, social networks, and forums provide many opportunities to the developers. The Sentiment analysis used widely in the industrial field.

Sentiment analysis aims to determine the feelings of speaker and writer with respect to the same document. The polarity of the document can be checked with different methods and give some range to a word in order to determine polarity, the sentence that is positively considered as greatest polarity as compared to the negative words.

In the present work discovers the reviews of a specialized article. Determine the positive and negative survey of individual of these articles. According to the reviews of the expert define how much that article technically sound or not. The positive review regarding article gives high polarity.

### II. RELATED WORK

The Support Vector Machine and K-Nearest Neighbor are two classification techniques used in an area of research. Different researches have been considered in the field of sentiment analysis, upon which our research is based.

**C. Bhadane et al.** Proposed that various methods used for classifying the text according to the opinion. The aspect classification followed by polarity classification. Described the lexical approach and machine learning approach. Maximum accuracy achieved by using kernel for all classes. Machine learning (SVM) and specific lexicon techniques combined for classification. It indicated about 78% accuracy [3].

**A.Tripathy et al.** Paper managed the text, grouping with a specific end goal to decide the aim of the creator of the text. It could be negative and positive. Identified the comparison of result by Naive Bayes and Support vector machine algorithms and these determined the positive and negative reviews. The data set created for the testing of the model is Polarity movie data set. The maximum accuracy achieved by the support vector machine classifier is 0.9406 [7].

**Xing Fang et al.** Paper tracked the problem of sentiment polarity categorization in which the different phases constitute such as sentiment sentence extraction and sentiment score calculation. The reviews regarding product collected from Amazon, sentence level categorization and review level categorization were conducted. Naive Bayesian, Random forest and Support vector machine methods were implemented [6].

**Emma Haddi et al.** Paper implemented the role of text processing in sentiment analysis. Experimental results produced appropriate representation. The SVM technique produced to more accurate result. FRN (feature relation network selection) was proposed to select relative features and improve accuracy by SVM utilization. The accuracy of FRN (89.65%) achieved greater than another utilizing Chi squared method (85.5%) [1].

**A.Hridoy et al.** Paper have developed methodology for an explanation of twitter data. Determines public opinion about the twitter data. Different opinions from the people were found, but generally consistent with outside reviews. Standard university developed tool SNLP (which is an open source natural language processing tool used for grammatical relations) between the words in a sentence as an output. Twitter was processed utilizing SentiWord after POS tagging [2].

**R. Gouveia Rodrigues et al.** SentiHealth-cancer(SHC-pt) tool implemented and SHC-pt was most appropriate to indicate the text. The worst accuracy (58%) achieved by tool greater than the highest accuracy by the other tool [14].

Table I Existing Approaches And Techniques

S.No.	Authors	Research paper	Approaches, Techniques and Results
1.	E. Guzman et al.	Sentiment analysis of commit comments in GitHub: An empirical study	Lexical sentiment analysis used in emotions of different open source projects and for commit comments of Git repository.
2.	C. Musto et al.	CrowdPulse: A framework for real time semantic analysis of social streams	CrowdPulse, a domain-agnostic framework for text analytics of social streams.
3.	Hongwei hao et al.	Recursive learning for sentiment analysis over social data.	The accuracy obtained by Recursive Neural Deep model (RNDM) is the maximum (90.8%) compared to NB (78.65%), ME (87.46%) and SVM (84.9%).
4.	Nagamma P* et al.	An improved sentiment analysis of online movie reviews based on clustering for box office prediction.	Clustering method used and SVM classifier has accuracy from 62% to 89.65% with clustering and NB has 72.4%.
5.	J.Serrano-Guerrero et al.	Sentiment analysis: A review and comparative analysis of web services.	Machine learning and Lexicon based techniques are utilized.
6.	K. Denecke et al.	Using SentiWordNet for multilingual sentiment analysis.	The method leverages on lexical resources for sentiment analysis obtainable in English (SentiWordNet).
7.	U. Grandi et al	A Borda count for collective sentiment analysis	The paper proposed notion of Borda count, which joins peoples' sentiment, according to similar comparative preference information .

In present work hybrid technique is applied to provide efficient analysis. In previous researches described the support vector machine provided about 78% accuracy [3] and it is very promising performing their tasks.

### III. PROPOSED APPROACH

The proposed work “sentiment analysis of technical article review using hybrid approach”, presenting a method to analyze the technology reviews, so the one can understand that how much present technology is technically sound. The Present approach is using the combination of K-Nearest Neighbor and Support Vector Machine classification techniques.

### A. Support Vector Machine:

The main idea behind the support vector machine is the hyper plane is produced, which can be helped in the classification. The hyper plane constructs in infinite dimensional space that separates the items of two categories. The hyper plane with maximum margin, generalization errors would be lower for the classifier. The vector at the margin help to classification of items is called support vectors. The support vector machine method takes a decision surface to convert the training data into different classes and support vector makes a decision, which efficient element selected as training set [15].

### B. K-Nearest Neighbor (KNN):

K-Nearest Neighbor is a simple classifier, but depends on the category labels attached to the training document similar to the text document. After that, the labels of most K-Nearest similar documents are considered and the final decision of the label of the new document is considered using a weighted average of the labels of this K-Nearest Neighbor.

The SuperFetch dataset has been taken in the consideration which consists of positive, negative and neutral reviews. The dataset cleaned using preprocessing method. The hybrid techniques are used for classification of reviews.

The steps followed for classification:

1.) *First step:* SuperFetch data are collected from a website which may consist of positive, negative and neutral reviews.

2.) *Second step:* The reviews contain vague information, which is eliminated using preprocessing. Stop words are removed (a, and, the, he, she etc.) [3] because these words do not contribute any information about the sentiments.

3.) *Third step:* Combination of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) used as hybrid approach for classification of reviews.

4.) *Fourth step:* Precision, Recall, F-measure and Accuracy as performance measurement parameters are evaluating the results.

5.) *Fifth step:* Finally, compared the result with existing techniques.

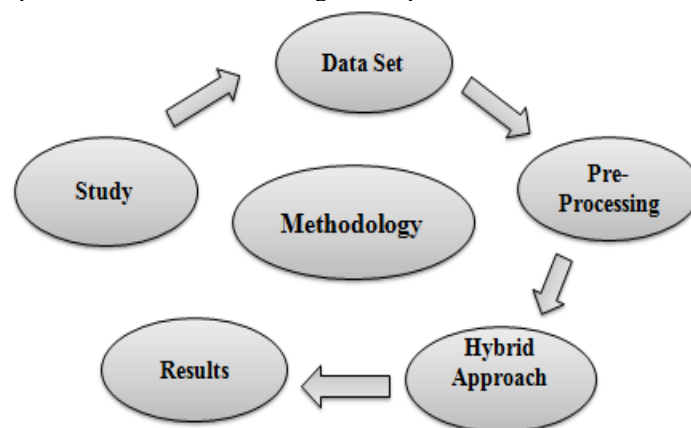


Fig.1 Methodology

## IV. RESULTS AND DISCUSSION

The implementation of hybrid technique is carried out on the SuperFetch review dataset. In Present work results are described based on the following parameters:

- Accuracy
- Precision
- F-Measure
- Recall
- Positivity
- Negativity

The present work divided the results into three cases:

Case1: Compute the described parameters using 50 reviews.

Case2: Compute the described parameters using 100 reviews.

Case3: Compute the described parameters using 120 reviews.

The calculated results for these cases are given below:

#### Case1:

Number of reviews : 50  
Positivity : 62.00%  
Negativity : 8.00%

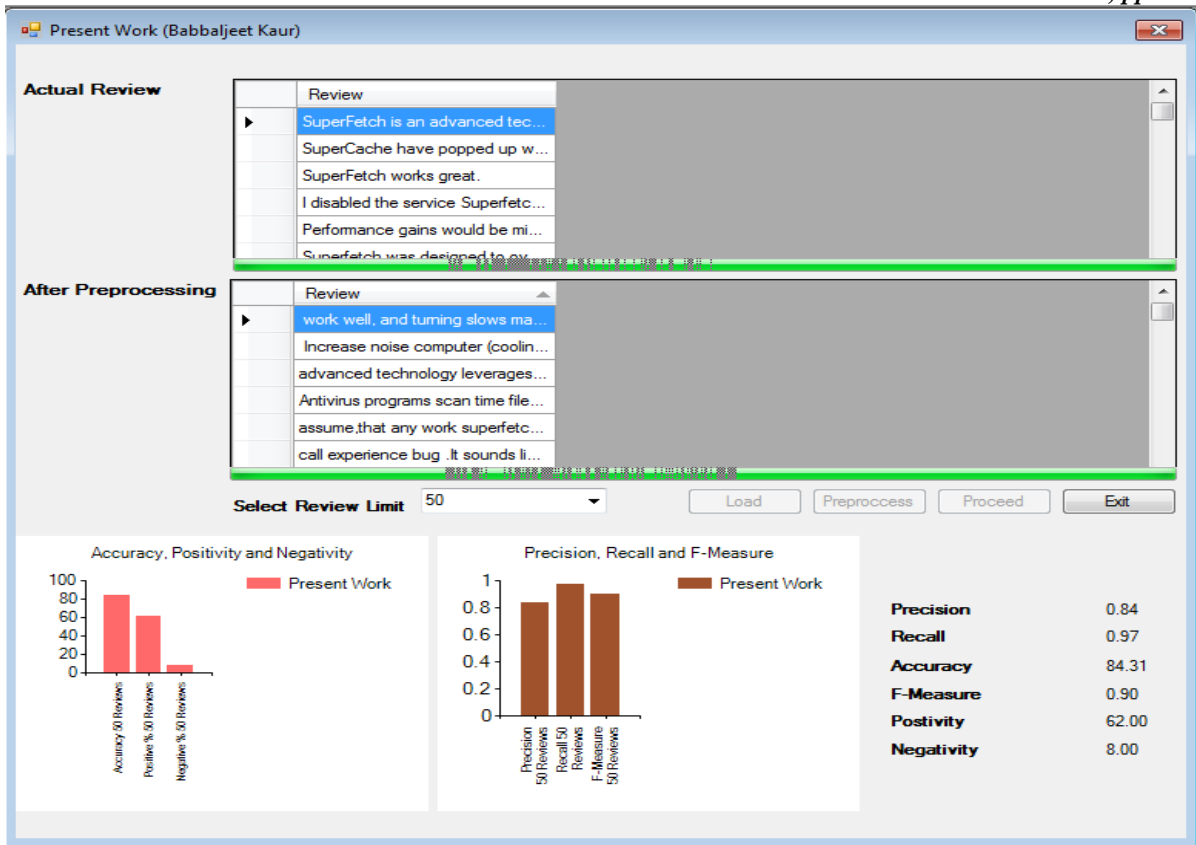


Fig.2 Result evaluations of 50 reviews.

**Case2:**

Number of reviews : 100  
 Positivity : 70.00%  
 Negativity : 5.00%

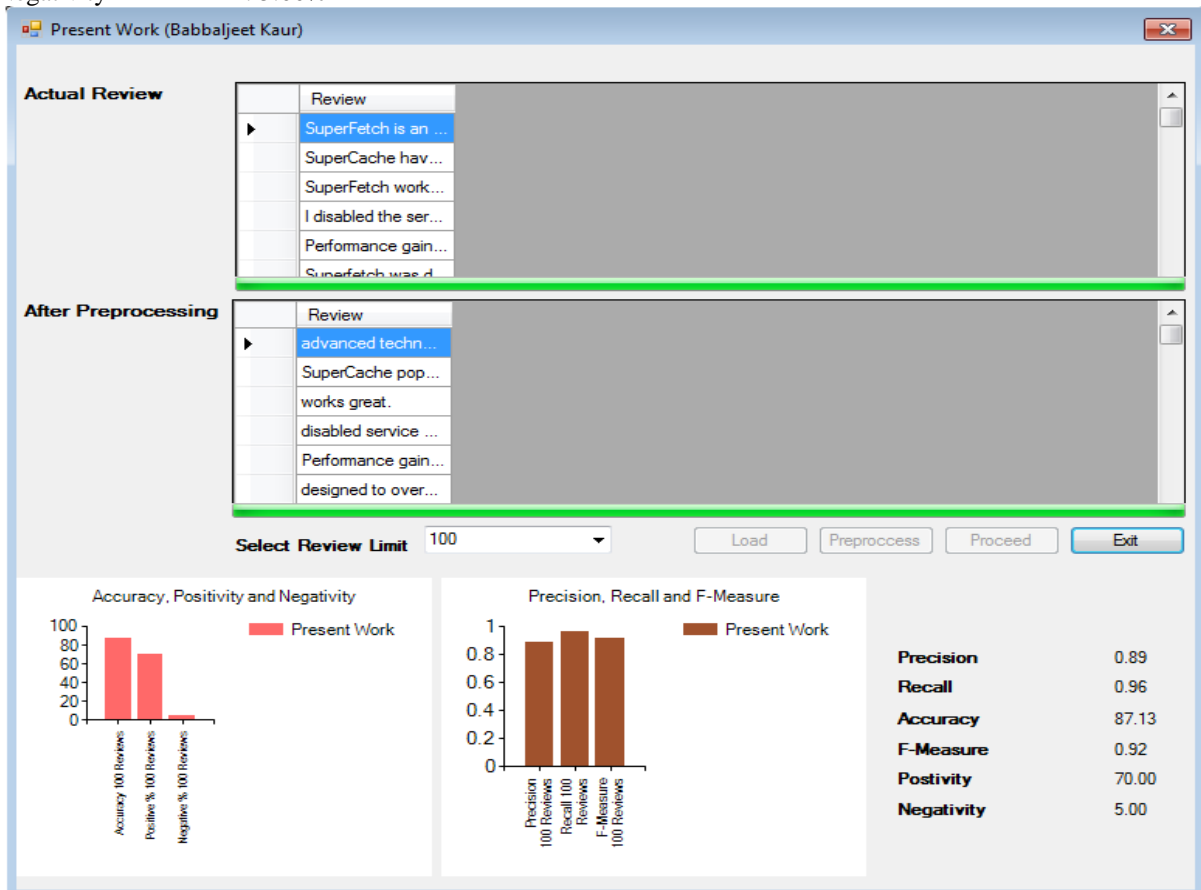


Fig.3 Result evaluations of 100 reviews.

**Case3:**

Number of reviews : 120  
 Positivity : 68.33%  
 Negativity : 5.00%

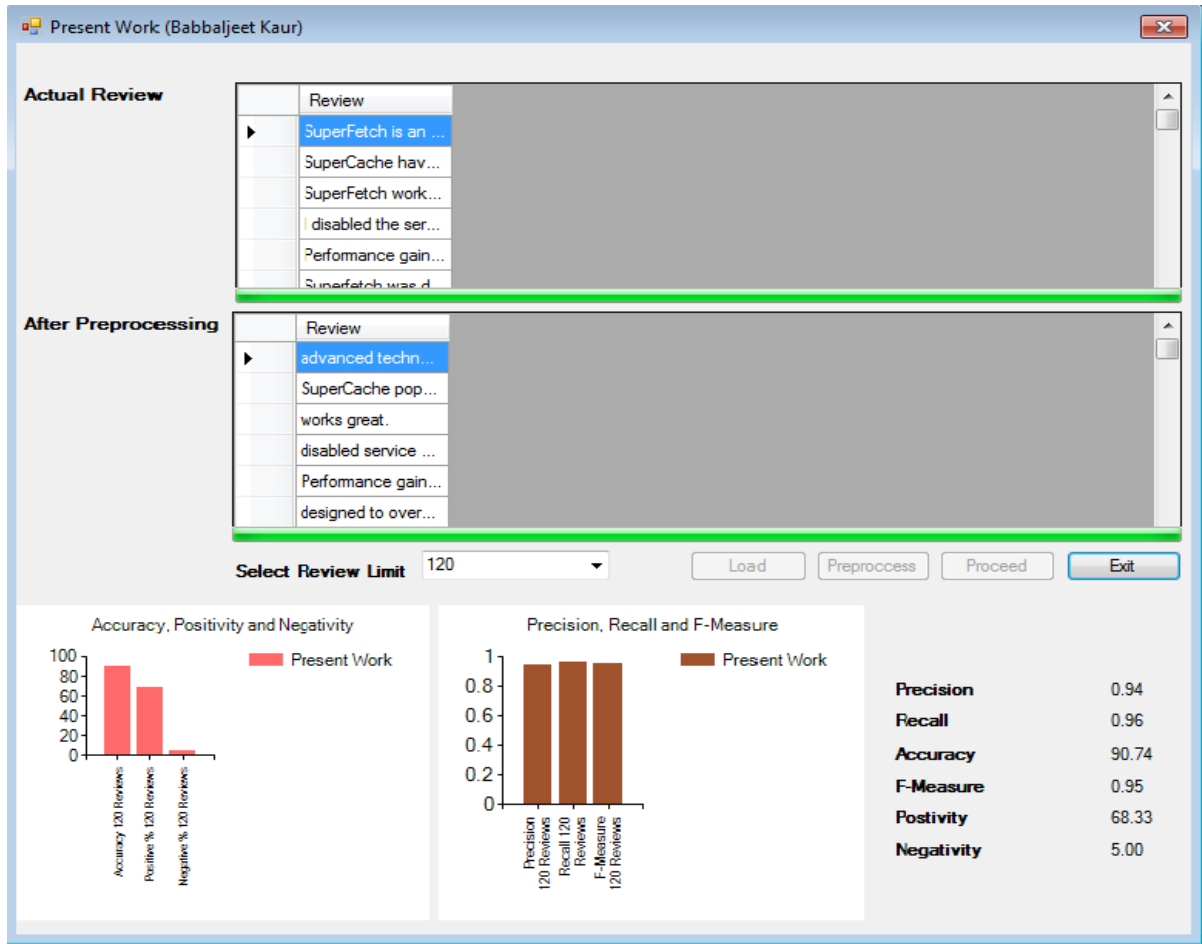


Fig.4 Result evaluations of 120 reviews.

Table II Result Table For All Cases

	Present Work		
	50 Reviews	100 Reviews	120 Reviews
<b>Precision</b>	0.84	0.89	0.94
<b>Recall</b>	0.97	0.96	0.96
<b>Accuracy</b>	84.31	87.13	90.74
<b>F-Measure</b>	0.90	0.92	0.95
<b>Positive %</b>	62.00	70.00	68.33
<b>Negative %</b>	8.00	5.00	5.00

**V. CONCLUSION AND FUTURE WORK**

Present work concluded that the combination of KNN and SVM produced better results on the basis of Accuracy, Precision, Recall and F-Measure. KNN improved the performance in the case of small reviews and SVM improved the performance in case of large reviews are working as a single hybrid approach.

There are two more parameters positivity and negativity are evaluated that shows most of the reviewers have positive thoughts about SuperFetch and the negativity percentage is very less, the remaining reviews are considered as neutral. So the results show that SuperFetch is a good feature in the memory management system. Future work includes the comparison of the present work with existing techniques.

**ACKNOWLEDGEMENT**

I am grateful to my guide Assistant Professor Mrs. Naveen Kumari for all help and valuable suggestion provided by her during the study and special thanks to Associate Professor Mr. Michael Swift who help me in collection of reviews.

## REFERENCES

- [1] Emma Haddi, Xiaohui Liu, Yong Shi, “The Role of text pre-processing in sentiment analysis”, *Information technology and quantitative management*, pp.26-32, 2013.
- [2] A. Hridoy, M.S.Islam, F. Ahmed and M.Rahman, “Localized twitter opinion mining using sentimental analysis”, *decision analytics*, 2015.
- [3] C. Bhadane, H. Dalal, H. Doshi, “Sentiment analysis: Measuring opinions”, *International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)*, pp.808 – 814, 2015.
- [4] Nagamma P\*, Pruthvi H.R, Nisha K.K and Shwetha N H, “An improved sentiment analysis of online movie reviews based on clustering for box office prediction”, *International conference on computing, communication and automation(IEEE)*, 2015.
- [5] S.Modha, S.Pandi, J.Modha, “Automatic sentiment analysis for unstructured data”, *International journal of advanced research in computer science and software engineering*, vol.3, pp.91-97, December -2013.
- [6] Xing Fang and Justin Zhan, “Sentiment analysis using product review data”, *Springer*, pp.1-14, 2015.
- [7] A.Tripathi, A.Agrawal, K. Rath, “Classification of Sentimental Reviews Using Machine Learning Techniques”, *International Conference on Recent Trends in Computing*, pp.821 – 829, 2015.
- [8] U. Grandi, A. Loreggia, F. Rossi, V. Saraswat, “A Borda count for collective sentiment analysis”, *Springer*, 22 October 2015.
- [9] J.S-Guerrero, JA.Olivas,FP.Romero,EH-Viedma, “Sentiment analysis: A review and comparative analysis of web services”, *Information Sciences*, pp. 18-38, 2015.
- [10] Kerstin Denecke, “Using SentiWordNet for Multilingual Sentiment Analysis”, *IEEE-2008*.
- [11] E. Guzman, D. Azocar, Yang Li, “Sentiment analysis of commit comments in GitHub: An Empirical Study”, *ACM* June-1, 2014.
- [12] C. Li, Bo Xu, G.Wu, S. He, G. Tian, H.Hao, “Recursive learning for sentiment analysis over social data”, *IEEE/WIC/ACM*, 2014.
- [13] C.Musto, G.Semeraro, P.Lops, M.De Gemmis, “A framework for real-time semantic analysis of social streams”, *information systems*, pp.127-146, 2015.
- [14] R. G. Rodrigues, R. M. das Dores, C. G. Camilo-Junior, T. C. Rosa, “SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks”, *International Journal of Medical Informatics (Elsevier)*, IJB-3243, pp.16.
- [15] Y. Sharma, V. Mangat, Mandeep Kaur, “Sentiment analysis and Opinion mining”, *IRF International conference*, pp. 35-38, March-8, 2015