# Sentimental Analysis of Product Reviews for e-Commerce Websites

## Akshatha M.P[1], Chandana S Urs[2], Javeria J[3], Meghana C[4], Mrs. Anandhi G[5]

[1,2,3,4]*Dept. of CSE,   GSSSIETW, Mysuru*
[5]*Assistant Professor, Dept. of CSE,   GSSSIETW, Mysuru*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract —** In today's world people tend to buy product online rather than purchasing manually. There are many e-shopping websites to purchase the product. People look through many e-Commerce websites about the product review before buying the product. It is a tedious task for the consumer to look through all the reviews in all the websites. The review may be present as positive, negative or neutral. It is possible to analyze huge reviews by using sentiment analysis. Sentimental Analysis is a machine learning technique to analyze the review as positive, negative or neutral. By using Sentimental Analysis, the sentiment of the single product can be compared between many e-Commerce websites. The aim of this paper is to suggest the user in which website the product can be bought based on positive reviews.

*Keywords—Sentimental Analysis, Product Review, Natural Language Processing, Machine Learning*

## I. INTRODUCTION

As in the days goes on, people purchasing through online has been increasing. Many people express their opinion about the product in e-Commerce websites they purchase. When consumer wants to buy the product, he will look on others opinion about the product before buying the product. The reviews are expressed in natural language. The reviews of the product are much larger to analyse by the consumer to make a decision. It is also difficult to look through reviews on multiple e-Commerce website for the product. To solve this problem Machine Learning technique can be used. Sentimental Analysis is an opinion mining which can classify the review as positive, negative or neutral. The analysis of the review consists of natural language processing and sentimental analysis of the reviews. It reduces the time taken by the user to view the product review from different website.

This paper aims at developing a system based on clustering and classification, to opinion mining of the product review. Clustering technique like mean shift algorithm and classification technique like Support Vector Machine (SVM) can be used. Then sentimental analysis of multiple e-Commerce website is compared. Based on the result obtained, it guides the user about the e-Commerce website. This approach adds additional features to the existing method.
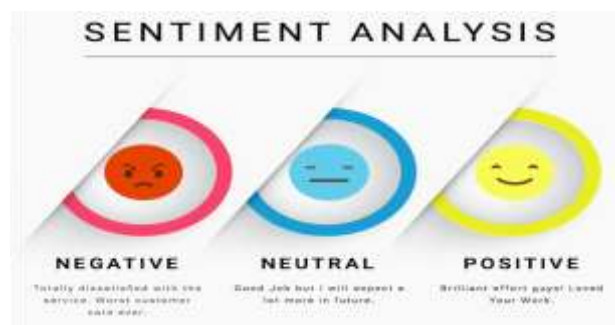


## II. RELATED WORK

Currently there is a website that can compare the product from different e-Commerce website based on price. But the greatest number of people search review before buying the product. Many surveys have shown that the review plays significant role in product purchase.

By the literature survey, we know that the opinion of the text can be found by many methods. To classify opinion there is NLP approach, lexical resources and hybrid approach. Machine learning technique is most efficient technique for opinion mining. It has the ability to deal with large amount of web data. Machine learning classification is subjected to three type: supervised, semi-supervised and unsupervised learning [8].

By paper [2], it is observed that by doing data pre-processing, the accuracy and efficiency of the system can be increased.

Paper [3], explains various steps in performing sentiment analysis using machine learning. The steps include data collection, data pre-processing, feature extraction, training and testing machine learning classifier. It also includes algorithm for sentiment analysis.

From [1], we came to know that SVM is the best classifier that provide accurate result for classifying problem. SVM can deal with large dataset.

Paper [5], suggests that data collection for the sentimental analysis of twitter can be obtained by two method, API search key twitter API streaming. Live tweets can be downloaded from API streaming.

## III. PROPOSED METHEDOLOGY

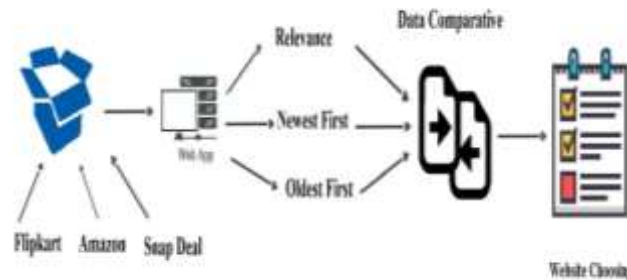The system architecture of the proposed system is shown below.



**Figure 1**: System Architecture

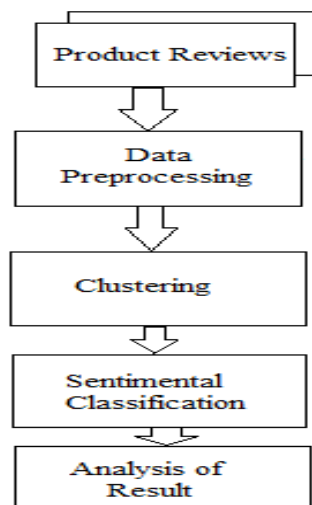The working process of the system is depicted in the below figure.



**Figure 2:** Work Flow of the system

To do sentimental analysis of the reviews of the product, the following steps need to be done.

*A. Data Gathering*

The reviews of the product from multiple e-Commerce website is gathered by using API streaming. By using API streaming live data can be obtained. Reviews are present in the form of natural language.

*B. Natural Language Processing (NLP)*

Preprocessing of the data can be done using data mining techniques. Data preprocessing increases the performance.

- Tokenisation

    It involves dividing a stream of sentences into word, term, symbol or into some element called tokens. This token act as an input to the further process.

- Stop word Removal

    Stop words are the useless words which occurs frequently such as a, an, the etc. It is a waste of processing time to analyse these words. So, such words are removed.

- Case Folding

    It refers to conversion of uppercase words to lowercase words.

- Stemming

    It is a process of reducing the derived words into their root or base form. It takes word as a input and returns stem of that word as an output. It is the important step in NLP.

*C. Sentimental Analysis*

The words are pre-processed and now available in the root form. The corpus of word should undergo some processes for the Sentimental Analysis.

- *Clustering Technique*

    Clustering is a method of grouping objects or data points into a number of groups such that similar data points form single group. The main aim is to segregate groups with similar characteristic.

    Clustering can be done using mean shift clustering. Mean Shift is a nonparametric clustering algorithm. It does not require prior knowledge on the number of clusters.

    Mean Shift works based on the concept of Kernel Density estimation. It is a method of underlying distribution for data. It places a kernel on each point in the dataset. And add all kernels up to generate a probability surface. Output of it doesn't depend on any explicit assumption on shape of data point distribution or the number of clusters. It is iterative for each point. In each iteration the point will move near to closer KDE.

- *Classification Technique*

    Classification is a technique that assigns set of items to a predefined class or category. In classification, the software is developed such that can learn how to classify the data items into groups.

    SVM is a supervised machine learning algorithm for classification problem. It is generally used for text categorisation. It is highly efficient and accurate compared to other classification algorithm. It tries to find a hyperplane which separates the data into classes as specified. By using high dimensional feature space, it can achieve good performance.

    For classification, it uses large margin. It separates the review using hyperplane. Discriminative function defined by SVM is

    $$g(X) = w^T \Phi (X) + b$$

**System Modules**

1. Admin

    Admin can login to the home and its credential details will be stored in the database. Admin can manage the products by adding or deleting. He can also view the product in the home page. Admin can manage users who has registered.

2. User

    User must register if he is new, otherwise can login by giving credentials like email, password. These details will be stored in the database. User will be able to view the products added by the admin and can also post comments for that particular product.

## IV. RESULT
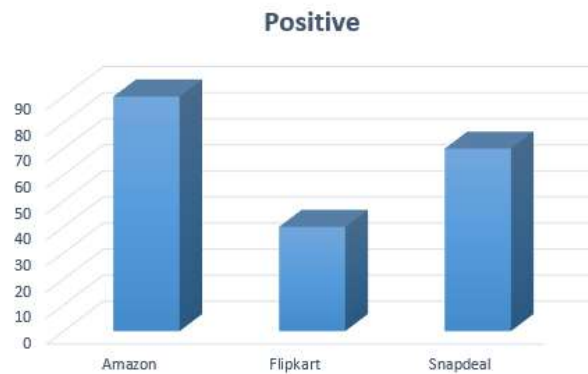


By comparing the websites it will show the website which will have maximum positive review. Based on the analysis the customer can choose the website. There will be a link which will redirect the user for the original e-Shopping website. The user can purchase the desired product.

## V. CONCLUSION AND FUTURE ENHANCEMENT

There are different ways to find sentiment of the text. Machine learning technique is easy, simple and efficient to find the opinion. Our system efficiently shows the most positively reviewed e-Commerce website for the purchase of the product. The main aim of this system is to compare different e-Commerce website. This technique can be applied in the field of transportation, in selecting movies and much more. The main advantage of the system is to reduce time wasted by the user in viewing the product review from different website.

Fake comments detection technique can be added as a feature to increase efficiency and accuracy in the future. The reason for the negative comment about the product in the e-Shopping website can be added.

## ACKNOWLEDGEMENT

## REFERENCES

[1] "Comparative Study of Machine Learning Techniques in Sentimental Analysis" by Bhavitha BK, Anisha P Rodrigues, dr. Niranjan N Chiplunkar, Dept. of CSE, NMAMIT, Nitte, ICICCT-2017.

[2] "Data Preprocessing for Efficient Sentimental Analysis" by Shreyas Vankhede, Ranjit Patil, Sagar Sonawane, Prof.Ashwini Save, VIVA Institute of Technology, Mumbai, ICICCT-2018.

[3] "Application of Machine Learning Techniques to Sentiment Analysis" by Anuja P Jain, Prof. Padma Dandannavar, 2 International Conference on Applied and theoritical Computing and Communication Technology, 2016.

[4] "A Novel approach for setimental analysis and opnion mining based on sentiwordnet using web data" by Shoiab Ahemd, Ajit Danti, IEEE-2015.

[5] "Real Time Sentiment Analysis of e-Commerse website using Machine Learning Algorithm" by K Sudheer, B Valarmathi, International Journal 2018.

[6] "A Model to Analyse Social media data to gain a Comapatitive Edge" by Kirtida Naik, Abhijith Joshi, Preeti Khanna, Narendra Shekokar, IEEE-2017.

[7] " Opinion Mining from Multi-domain User Reviews Using Sentiment Analysis", Ashish A. Bhalerao, Sachin N. Deshmukh, Asst.Professor, MGM Dr. G.Y.Pathrikar College of CS and IT,  Aurangabad, India.

[8] "Research on the Impact of Consumer Review Sentiments from Different Websites on Product Sales" by Jiangning Wu, Lingyu Du, Yanzhong Dang, Faculty of Management and Economics Dalian University of Technology Dalian, China. IEEE-2018.

[9] "Sentimental Analysis of Product Reviews for e-Commerce Recommendation" by D. Mali, M. Abhyankar, P. Bhavarthi, K. Gaidhar, M.Bangare, Department of: Information Technology, Smt. Kashibai Navale College of Engineering, Pune, International Journal of Management and Applied Science, 2016.

[10] "Sentiment Analysis in Twitter using Machine Learning Techniques" by Neethu M S, Rajasree R, Dept. of CSE, College of Engineering Trivandrum, IEEE-2013.

[11] "Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites" by Ting Bai , Wanye Xin Zha, Yulan He, Jian-Yun Nie, Ji-Rong Wen, IEEE-2018.

[12] "Combining Classification and clustering for tweet Sentimental Analysis" by Luiz F.S.Coletta, Nadia F.F da Silva, Eduardo R Hruschka, Estevam R Hruschka, Brazilian Conference on Intelligent System, 2014.

[13] "Product Ranking System in E-Commerce Website for Validation using Sentimental Analysis" by Jayasree J, Sri Amitha Mathi M, Malini S, Bavithra E, Dr. P. Boobalan, IJSTE - International Journal of Science Technology & Engineering, 2017