

# PRODUCT REVIEW USING FEATURED BASED SENTIMENTAL ANALYSIS

## Shubham Trivedi

Student, Department of  
Information technology, Shah &  
Anchor Kutchhi Engineering  
College, Chembur, Mumbai-  
400088

## Smit Shah

Student, Department of  
Information technology, Shah &  
Anchor Kutchhi Engineering  
College, Chembur, Mumbai-  
400088

## Ansh Sharma

Student, Department of  
Information technology, Shah &  
Anchor Kutchhi Engineering  
College, Chembur, Mumbai-  
400088

## Shubham Shah

Student, Department of  
Information technology, Shah &  
Anchor Kutchhi Engineering  
College, Chembur,  
Mumbai-400088

## Panjab Mane

Assistant Professor, Department  
of Information technology, Shah  
& Anchor Kutchhi Engineering  
College, Chembur, Mumbai-  
400088

**Abstract—** Today E-commerce has become an important part of our day to day life and people are getting dependent on these website products. The user reviews too, are becoming important for customers. So, through this project we are building a software, which rates the E-commerce products based on feature-based sentiment analysis of user reviews. With this shopping for customers will become very convenient as well as time saving from reviewing large set of user reviews. Semantic orientation is a measure of subjectivity and opinion in text. It usually captures an evaluative factor (positive or negative) and potency or strength (degree to which the word, phrase, sentence, or document in question is positive or negative) towards a subject topic, person, or idea.

**Keywords—**reviews, sentiments, analysis, learning, polarity

## I. INTRODUCTION

Online shopping is the current trend in the field of buying and selling products. It has been an integral form of E-Commerce in the current era. Product review is an important feature of these sites. Product reviews are used on shopping sites to give customers an opportunity to rate and comment on products they have purchased, right on the product page. Other consumers can read this when making a purchase decision. Often the company will include a URL on printed literature or e-mail marketing to invite customers to review their services after transaction has been completed.

Sentiment Analysis or Opinion mining refers to the use of natural language processing, text analysis, computational linguistics and machine learning to systematically identify, extract, quantify and study affective states and subjective information. An important research direction in sentiment analysis is to identify the sentiment polarity of individual words, known as words semantic orientation (referred to below as WSO). Semantic orientation (SO) is a measure of subjectivity and opinion in text. It usually captures an evaluative factor (positive or negative) and potency or strength (degree to which the word, phrase, sentence, or document in question is positive or negative) towards a subject topic, person, or idea. By using sentimental analysis in product review, the products will be rated based on the comments given by the buyers. The algorithm will allow the system to understand the language in comments and rate the product based on negative or positive comments. The product will be ranked according to rating it receives.

## II. LITERATURE REVIEW

In this section, we will only review some previous work, upon which our research is essentially based.

Hui and Liu summarized a list of positive words and a list of negative words, respectively, based on customer reviews. The positive list contains 2006 words and the negative list has 4783 words. Both lists also include some misspelled words that are frequently present in social media content. Sentiment categorization is essentially a classification problem, where features that contain opinions or sentiment information should be identified before the classification.

For feature selection, Pang and Lee suggested to remove objective sentences by extracting subjective ones. They proposed a text-categorization technique that is able to identify subjective content using minimum cut.

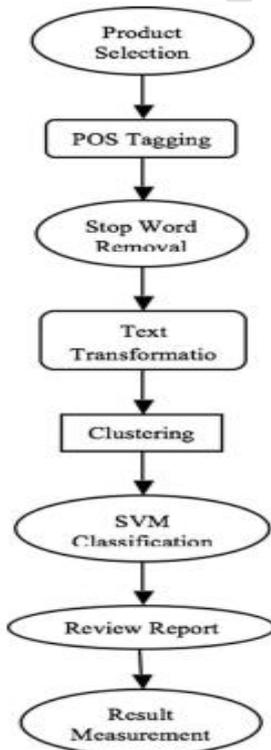
Gann et al selected 6,799 tokens based on Twitter data, where each token is assigned a sentiment score, namely TSI (Total Sentiment Index), featuring itself as a positive token or a negative token.

Shuai Bai, Jiangchun Ren performed analysis on products of JD Mall in which they used Grey prediction mode. The experimental result showed grey prediction model having accuracy of 79.48%.

However, they faced challenges in the form of computational complexity, which requires design effort of accurate modelling.

Athira U. and Sabu M. Thampi performed analysis on Posts from different social media groups in which they used Linguistic featured based method. The experimental result showed linguistic feature-based model having accuracy of 60%. However, they faced challenges in identifying multilingual conversations.

Robert E. Schapire used adaboosting approach to Machine Learning.



Flow of proposed work.

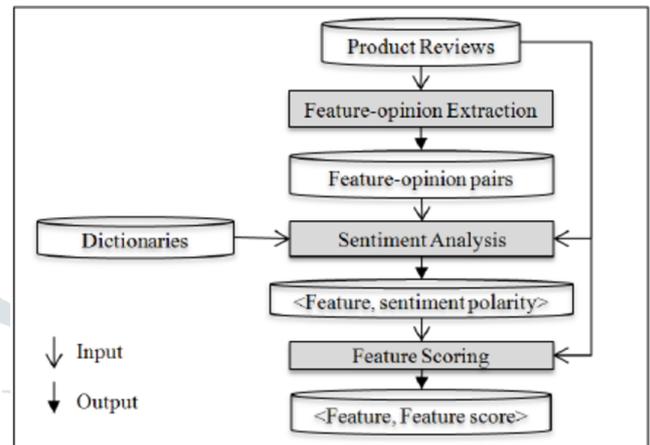
Fig 1 :

### III. MODELLING

The following design outline displays the work flow of the system.

The flow of the project is as follows:

1. In this the file is uploaded manually by the user/customer.
2. User login and fetching data by plugin and sorting it based on proposed algorithm.



Minimum hardware requirements: -

1. Processor: Pentium 4 or Higher
2. RAM: 512MB or more
3. Hard disk: 100MB or more

Software Interfaces.

Database management system will be required for a strong connectivity between the dataset, which includes the information of Captured data packets upon which the response is made and the algorithm. The basic windows upload option for taking the dataset file input are created. The algorithm is coded in Python.

Software:

1. Windows Operating System.
2. Python 3.0 or above
3. MySQL
4. Modern Web Browser
5. Amazon API
6. WebPy Framework Package

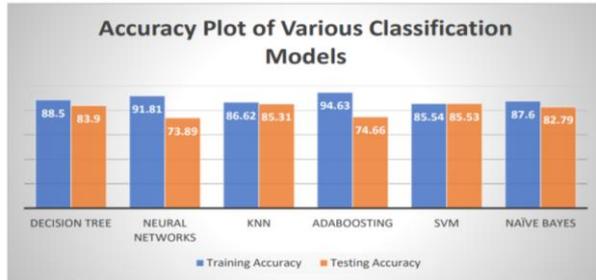
### IV. RESULT AND EVALUATION

The below table consists of all the results of all the classification models that we have implemented till now. Also, the precision, recall, and F1-score of the Testing dataset are calculated. Also, sample complexity is used as performance measures to compare the results. Colour coding depicts which are the best model and worse models on Amazon Dataset.

Classification Model	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score
Decision Tree	88.5	83.9	0.77	0.84	0.79
Neural Networks	91.81	73.89	0.66	0.74	0.67
KNN	86.62	85.31	0.79	0.86	0.8
Adaboosting	94.63	74.66	0.68	0.75	0.69
SVM	85.54	85.53	0.79	0.86	0.8
Naive Bayes	87.6	82.79	0.76	0.83	0.77

Table Demonstrating various values of performance of each model.

The plot for accuracy comparison of each classification models:



From the above results and tables, one of the best classification models that perfectly classifies the given Amazon Review dataset is both the K-Nearest Neighbour classification model which gives a training accuracy of 86.62% and Testing Accuracy of 85.31%. Also, Support Vector Machines with Poly Kernel is also giving a better accuracy with degree 3 having training accuracy of 85.54% and Testing accuracy of 85.53%. Also, these two classification models contain the highest Precision, Recall and F1-score values which are some of the other measures for estimating the performance of the dataset. Hence from the above values, we can conclude KNN and SVM as the best classification model for the Amazon Sentiment Review Analysis.

## V. CONCLUSION

Sentiment analysis or opinion mining is a field of study that analyses people's sentiments, attitudes, or emotions

towards certain entities. The main aim of this project tackles a fundamental problem of sentiment analysis and

sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study.

We have proposed a sentiment polarity categorization process along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed.

## VI. FUTURE WORK

- Increase accuracy and decrease processing time.
- Sarcasm detection
- Different browser support.
- Support multiple language for processing.

## REFERENCES

- [1] Hui Song and Ying Xiang Fan, Extracting Product features from Online Reviews. College of Computer Science and Technology, Donghua University, China. 2016.
- [2] Paitoon Portrakoon, A Model for Multi-dimensional Trust Measurement. Chayapol Moemeng Vincent Mary School of Science and Technology Assumption University, Thailand. 2017.
- [3] Shuai Bai and Jiangchun Ren, Dynamic Evaluation of Consumer Satisfaction. National University of Defense Technology, China. 2017.
- [4] Athira U. and Sabu M. Thampi, Linguistic Feature Based Filtering Mechanism for Recommending Posts in a Social Networking Group. Indian Institute of Information Technology and Management, India. 2018