

## Review Analyzing System using Lexicon Based Approach

Sunit Khavale<sup>1</sup>, Jagdish Korde<sup>2</sup>, Vaibhav Patil<sup>3</sup>, Shubham Gund<sup>4</sup>

INFORMATION TECHNOLOGY & JAYAWANTRAO SAWANT COLLEGE OF ENGINEERING, PUNE

\*\*\*

**Abstract** - E-Commerce allows us to interact with various product suppliers and buy various products from all over the world. There are various sites providing online facilities to both buyer and seller of product. In today's world quality of the product play major role and user must be satisfied with the product. Advantage of using this approach in the E-Commerce site is to provide the user guarantee about the product before buying it. This will help the customer to identify right product and buy it after verifying the product. Reviewer of the product are important who will help to identify the customer as well as site owner to manage the buying or visibility of the product. By applying technique will help to determine the opinion of the previous buyer as well as aware new customer while purchasing the product by providing rating of the product. Through this customer, product seller and web manager all will get analysis of product.

**Keywords:** Prediction, semi-supervised learning, lexicon based approach, expert system, data analytics,

### 1. INTRODUCTION

In traditional way, one who tend to buy the product go through reviews manually by reading some of them which doesn't help to understand the quality of product accurately. Sometimes rating given by previous buyers doesn't help both new customers, product managers. Due to this buyer doesn't get required guarantee or satisfaction and product manager not able to manage productivity. In the proposed system the customer will get accurate info regarding the product through rating or any other visual format.

As well as the product manager will get help by understanding product sales through by visualizing the obtained result. The research paper analyses the algorithms and methods used to implement prediction of prices. It is directed towards mining and also to provide a detailed analysis of products based on several factors that may have a direct impact on the seller and buyer. While attempts have been made to predict the class of products, however they have not yet been implemented for the products of ecommerce.

### 2. LITERATURE SURVEY

In the research paper named Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language.

[1]Franco Chiavetta, Giosue Lo Bosco and Giovanni Pilato. The system has been prototyped by using Python language and it has been tested on a dataset of reviews crawled from Amazon.it, the Italian Amazon website. Experiments show that the proposed system is able to automatically classify both positive and negative reviews, with an average accuracy of above 76%.

In the research paper named Extracting Sentiments from Reviews: A LexiconBased Approach

[2] Sujata L. Sonawane, Pallavi V. Kulkarni. This paper used lexicon based approach for classifying a review document as positive, negative or neutral. This paper extracts the sentiments from customer reviews and SentiWordNet is used to assign the polarity to each sentiment. The classification of review document is predicted by sentimental score. It provides the accuracy upto 76% In the research paper named Sentiment Analysis On Facebook Group Using Lexicon Based Approach

[3]Sunidhi Dwivedi proposed an unsupervised dictionary based approach. Document level opinion mining system performs.

1) data preprocessing- remove unwanted data, sentence split, tokenization, POS tagging and stemming.

2) Extraction of opinion words- adjectives are extracted

using Apache Open NLP tool and extracted adjective stored in text file.

3) Seed list preparation- extracted adjectives are matched with the seed list words to find their polarity.

4) Polarity detection.

5) calculation of sentimental score. It compares the sentimenta.

score of different mobiles finally determining whether the overall product has positive, negative or neutral opinion. It introduces a lexicon method. It focuses on single document unigram level sentiment analysis with an average of 70% accuracy.

Sr. no.	Title Of Paper	Technique used	Efficiency	Future Scope
1	Sentiment Analysis of News Articles: A Lexicon based Approach	Lexicon and corpus based approach	77.00%	Can improve using good data preprocessing techniques and
2	ML in prediction of stock market based on historical data and data from twitter sentiment analysis	Corpus based system with generative methods	64.10%	Can improve accuracy using a lexicon based approach
3	Predicting stock market trends using random forests	Random forest algorithm	65.00%	Hybrid model can be used to improve accuracy
4	Stock price prediction using regression based sentiment analysis	Opinion word extraction	72%	Can improve accuracy using POS tagging and word weighting
5	Predicting math test scores using k-nearest	K-nearest algorithm	72%	Can improve accuracy using categorization algorithms
6	Syntactic and semantic approach for data mining	Word sentiment polarity identification	70%	Can trained using corpus based approach and heuristic approach

Ding et al., [4] Proposed an all encompassing vocabulary based approach which utilizes outer signs and phonetic traditions of normal dialect articulations to decide the semantic introductions of conclusions. Favorable position of this approach is that sentiment words which are setting subordinate are effectively dealt with. The calculation utilized utilizations semantic examples to manage exceptional words, phrases. Analysts manufactured a framework called Opinion Observer in view of this method. Examinations utilizing item audit dataset was profoundly compelling. It was demonstrated that numerous clashing conclusion words in sentences are likewise managed effectively. This framework indicates better execution when contrasted with existing strategies.

[5] Anna Jurek, Maurice D. Mulvenna and Yaxin Bi, this shows another dictionary based slant investigation calculation that has been composed with the principle concentrate on ongoing Twitter content examination. The calculation

comprises of two key parts, to be specific opinion standardization and proof based mix work, which have been utilized as a part of request to evaluate the force of the supposition as opposed to positive/negative mark and to help the blended slant grouping process. At long last, they represent a contextual analysis looking at the connection between negative suppositions of twitter presents related on English Defense League and the level of confusion amid the association's connected occasions.

[6] Mao et al. explained In the experiment, method to determine text polarity classification has many advantages (1) The method automatic extracts the semantic features without semantic word dictionary.(2) The accuracy of the method will be higher when more prior knowledge is added. In the future research, the sentiment features in the Sentence Similarity Computing will be applied. Sentence Similarity Computing is the most critical technology in QA system, and the sentiment of sentences influences the similarity of sentences.

In the Research paper named

[7] Sentiment Analysis Based on Dictionary Approach Reshma Bhonde, Binita Bhagwat, Sayali Ingulkar, Apeksha Pande The naive bayes approach is used to interpret each sentence as positive and negative on the bases of useful word level feature. SVM classifier is trained on the interpreted sentences for the positive and negative classification. Contextual data is used to calculate the polarity of sentence and mark it as either negative or positive.

We are using "Lexicon and corpus based approach" which are giving 77% accuracy and also we can improve it by using good data cleaning techniques and semantic approach.

### 3. PROPOSED SYSTEM

#### A. System Architecture

1) Web Interface: It is the Front End of the system through which user can interact with the system. The Web Interface contains the following parts :-

a) Login Portal

b) User Details and Data Entry

### c) User Reviews

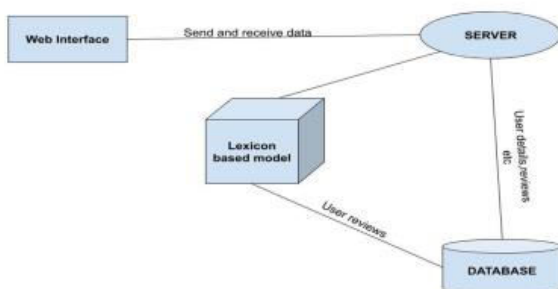
2) Web Server: It is the Middleware of the system which contains the actual raw data provided by the user and helps process that

data for further use.

It performs the following tasks :-

- a) Web Hosting
- b) Database Services
- c) Store the ML Model

3) Lexicon based Model: It is the Back End of the system through which the raw data provided by the user will be used to predict the product rating and analyzer the given data for product manager.



### 1) Problem Definition

- a) To take user reviews as input of particular product and generate rating for customer as well as make graphical representation for product manager.

### 2) Hardware and Software Requirements

- a) **Hardware Requirements:** 4GB Ram, Intel/AMD Processor, 5GB Storage Space.
- b) **Software Requirements:** Php, Latest Web Browser, Android 4.0.

### B. Analysis

As we are performing semi-supervised learning hence there is no requirement of previous data that are labelled or training the

model, we are using sentiword or vader dictionary to perform the classification of positive and negative words and as well as the words fall in neutral.

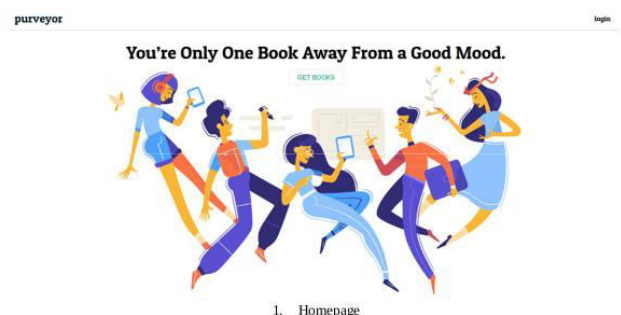
Analysis of Work needed to be done in stages:-

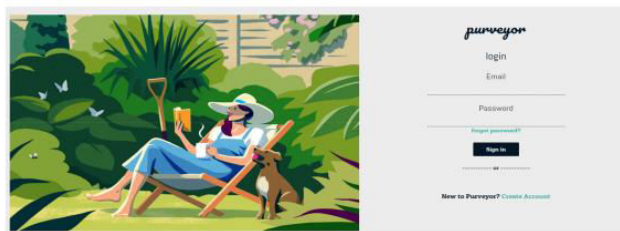
- 1) Identify the positive and negative reviews
- 2) Create a lexicon based algorithm which can provide the most accurate rating and visualization.

### C. Design Phase

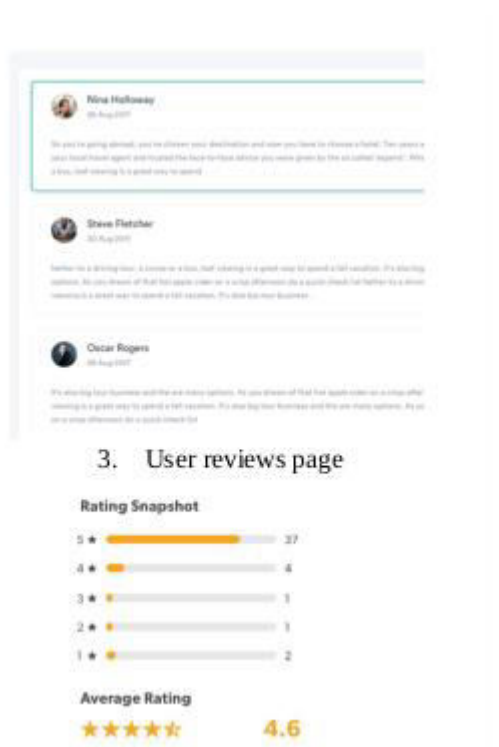
## 4. ANALYSIS AND DESIGN PHASE

### A. Requirements





2. Login page



3. User reviews page

## 5. CONCLUSION

A. We will develop the proposed expert system by using different data analysis algorithms to determine the quality, user perception predicted by each algorithm. The proposed expert system is to deploy on a web-application which will also be extended to an android application which will encompass the same functionality as that of the web-application itself.

B. We will also work to add some new features in the algorithm and try to increase accuracy of lexicon based model. We are going to get data from wordnet or vader that will use to predict the category of the review and will try to add some manual features to improve result.

## 6. ACKNOWLEDGEMENT

I would like to express gratitude to my guide Prof. P.D. Lambhate for valuable suggestions and direction towards the execution of this project. I convey my heartfelt thanks to Prof. M.K.Gawali for her dynamic support being the project coordinator. I'm very thankful to Dr. J.S.Patil, Head of the Department, Computer Engineering, who has extended support and valuable suggestions towards achieving success in this project.

## 7. REFERENCES

[1] Geetanjali Shirole,Savita sangam, in "Lexicon Based Sentiment Analysis For Product Review Data ©2018 IEEE

I would like to express gratitude to my guide Prof. darshana patil for valuable suggestions and direction towards the execution of this project. I convey my heartfelt thanks to Prof. Swati Patil for her dynamic support being the project



5. Review visualization for product manager

coordinator. I'm very thankful to Dr. Arati Dandavate, Head of the Department, Computer Engineering, who has extended support and valuable suggestions towards achieving success in this project.

[2] Shailendra Kumar Singh, Sanchita Paul, Dhananjay Kumar, Sentiment Analysis Approaches on Different Data set Domain: Survey, IJDTA, Vol. 7, No. 5,

2017, pp. 39-50.

[3] M. Devika, C. Sunitha, and A. Ganesh, "Sentiment analysis: a comparative study on different approaches," Procedia Computer Science, vol. 87, pp. 44–49, 2018.

[4] Xiaowen Ding, Bing Liu, and Philip S Yu. A holistic lexicon-based approach to opinion mining. In Proceedings of the 2017 International Conference on Web Search and Data Mining, pages 231–240. ACM, 2019

[5] Jessica Maikhanh Brown in "Predicting math test scores using k-nearest".978-1-5090-5379-7/17/\$31.00 (c) 2017 IEEE

[6] Itani M., Roast C., and Al-Khayatt S. "Corpora for Sentiment Analysis of Arabic Text in Social Media", In: Proceedings of the 2017 8th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 64-69, (2017)

[7]Neti Sriwulan Sutiawan and Baskara Nugraha in "Online Price Prediction system of consumption commodities" . 2017, IEEE International Conference on Information Technology systems and Innovation(ICITSI)

[8] Priti Saxena, Bhaskar Pant , R.H. Goudar , Smriti Srivastav , Varsha Garg and Shreela Pareek in "Future Predictions in Indian Stock Market through Linguistic-

[9] F. Bond and R. Foster, "Linking and extending an open multilingual wordnet," in Proceedings of the 51st Annual Meeting of the Association for

[10] ] D.M.E.D.M. Hussein, "A survey on sentiment analysis challenges," Journal of King Saud University - Engineering Sciences, vol. 30, no. 4, pp. 330–338, 2018.