CHARACTERIZING AND PREDICTING REVIEWS FOR EFFECTIVE PRODUCT MARKETING AND ADVANCEMENT

A Thesis

Submitted

In Partial Fulfillment of the Requirements

For The Degree of

MASTER OF TECHNOLOGY

In

Computer Science & Engineering

Submitted by:

Aihsan Suhail

Enroll No. 1900100926
Roll No. 1901621003

Under the Supervision of:

Mrs. Halima Sadia

(Assistant Professor)

Department of Computer Science & Engineering

Faculty of Engineering

INTEGRAL UNIVERSITY, LUCKNOW, INDIA

August, 2020

CERTIFICATE

This is to certify that **Aihsan Suhail** (Enroll. No. 1900100926) ha**s** carried out the research work presented in the dissertation titled **"Characterizing and Predicting Reviews for Effective Product Marketing and Advancement"** submitted for partial fulfillment for the award of the **Masters of Technology in Computer Science & Engineering** from **Integral University, Lucknow** under my supervision.

It is also certified that:

(i) This dissertation embodies the original work of the candidate and has not been earlier submitted elsewhere for the award of any degree/diploma/certificate.

(ii) The candidate has worked under my supervision for the prescribed period.

(iii) The dissertation fulfills the requirements of the norms and standards prescribed by the University Grants Commission and Integral University, Lucknow, India.

(iv) No published work (figure, data, table etc) has been reproduced in the dissertation without express permission of the copyright owner(s).

Therefore, I deem this work fit and recommend for submission for the award of the aforesaid degree.

Mrs. Halima Sadia
Dissertation Guide
(Assistant Professor)
Department of CSE,
Integral University, Lucknow

Date:

Place: Lucknow

DECLARATION

I hereby declare that the dissertation titled "**Characterizing and Predicting Reviews for Effective Product Marketing and Advancement**" is an authentic record of the research work carried out by me under the supervision of Mrs. Halima Sadia, Department of Computer Science & Engineering, Integral University, Lucknow. No part of this dissertation has been presented elsewhere for any other degree or diploma earlier.

I declare that I have faithfully acknowledged and referred to the works of other researchers wherever their published works have been cited in the dissertation. I further certify that I have not willfully taken other's work, para, text, data, results, tables, figures etc. reported in the journals, books, magazines, reports, dissertations, theses, etc., or available at web-sites without their permission, and have not included those in this MTech dissertation citing as my own work.

Date:

Signature
Aihsan Suhail
1900100926

RECOMMENDATION

On the basis of the declaration submitted by "**Aihsan Suhail**", a student of MTech CSE (FT/Evening), successful completion of Pre presentation on 20/07/2021 and the certificate issued by the supervisor Mrs. Halima Sadia, Assistant Professor Computer Science and Engineering Department, Integral University, the work entitled "**Characterizing and Predicting Reviews for Effective Product Marketing and Advancement**" , submitted to department of CSE, in partial fulfillment ofthe requirement for award of the degree of Master of Technology in Computer Science & Engineering, is recommended for examination.

_____	_____
Program Coordinator Signature	**HOD Signature**
Dr. Faiyaz Ahmad	Dr. M Akheela Khanum
Dept. of Computer Science & Engineering	Dept. of Computer Science & Engineering
Date:	Date:

COPYRIGHT TRANSFER CERTIFICATE

Title of the Dissertation: **Characterizing and Predicting Reviews for Effective Product Marketing and Advancement**

Candidate Name: **Aihsan Suhail**

The undersigned hereby assigns to Integral University all rights under copyright that may exist in and for the above dissertation, authored by the undersigned and submitted to the University for the Award of the M.Tech degree.

The Candidate may reproduce or authorize others to reproduce material extracted verbatim from the dissertation or derivative of the dissertation for personal and/or publication purpose(s) provided that the source and the University's copyright notices are indicated.

AIHSAN SUHAIL

ACKNOWLEDGEMENT

I am highly grateful to the Head of Department of Computer Science and Engineering for giving me proper guidance and advice and facility for the successful completion of my dissertation.

It gives me a great pleasure to express my deep sense of gratitude and indebtedness to my guide **Mrs. Halima Sadia, Assistant Professor, Department of Computer Science and Engineering,** for her valuable support and encouraging mentality throughout the project. I am highly obliged to her for providing me this opportunity to carry out the ideas and work during my project period and helping me to gain the successful completion of my Project.

I am also highly obliged to the Head of department, **Dr. Mohammadi Akheela Khanum (Associate Professor, Department Of Computer Science and Engineering)** and PG Program Coordinator **Dr. Faiyaz Ahamad, Assistant Professor, Department of Computer Science and Engineering,** for providing me all the facilities in all activities and for his support and valuable encouragement throughout my project.

My special thanks are going to all of the faculties for encouraging me constantly to work hard in this project. I pay my respect and love to my parents and all other family members and friends for their help and encouragement throughout this course of project work.

Date:

Place: Lucknow

INDEX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

SVM	Support Vector Machine
ACK	Acknowledgement
LR	Logistic Regression
OFM	Optical Flow method
AI	Artificial Intelligence
POS	Part-of-speech
WDE	Weakly-supervised Deep Embedding
BOG	Bag Of Words
SRS	Software Requirement Specification

ABSTRACT

Now days shopping on internet increases day by day so online surveys have become a important part wellspring of data for clients prior to settling on an smart buy choice. Early reviews of an item will in general exceptionally affect the ensuing item deals. In this thesis,we strengthened and studied the behavioral quality of the first batch of reviewers through reviews published on our procurement portal. We clearly divide the life cycle of the item into three stages after , especially the initial lion part. Customers who publish surveys in the initial stage of are considered early analysts. Based on their scoring practices, the support scores of obtained from other people, and the relationship between their surveys and the popularity of the item , we give a quantitative description of the early reviewers.We have tracked down that (1) an early analyst will in general relegate a higher normal rating score; and (2) an early reviewer will in general post more supportive audits. Our examination of item surveys additionally demonstrates that early reviewers appraisals and their got support scores are probably going to impact item prominence. As a survey audit release measure for competitive multiplayer games, we propose a novel implemented model based on Edge for early analyst forecasts. A comprehensive investigation of two different web-based business data sets showed that our proposed method defeated several cruel baselines.

CHAPTER - 1

INTRODUCTION

1.1 INTRODUCTION

The development of web based shopping websites has empowered clients to distribute or share buy encounters by uploading item reviews, which for the most part contain helpful conclusions, remarks and criticism towards an item. Item surveys, particularly the early audits (i.e., the surveys posted in the beginning phase of an item), profoundly affect ensuing item deals. We call the clients who posted the early surveys early commentators.

Albeit early analysts contribute just a little extent of audits, their sentiments can decide the achievement or disappointment of new items and administrations. It is significant for organizations to distinguish early commentators since their criticisms can assist organizations with changing promoting techniques and improve item plans, which can in the long run lead to the achievement of their new items.

Thus, early commentators become the accentuation to screen and draw in at the early advancement phase of an organization. The essential job of early audits has drawn in broad consideration from promoting experts to actuate shopper buy expectations . For instance, Amazon, one of the biggest online business organization on the planet, has supported the Early Reviewer Program1, which assists with getting early surveys on items that have not many or no audits.

Past examinations have profoundly stressed the wonder that people are emphatically impacted by the choices of others, which can be clarified by group conduct.

The impact of early audits on later purchases can be seen as a unique example of the impact of congestion. Early audits include evaluations of major projects from previous adopters, and is an important reference asset for the final purchase choice. As shown in , when shoppers use other people's article ratings to measure the quality of articles on the Internet, crowd behavior occurs in web-based purchase measures. Unlike the existing checks on crowd behavior, we revolved around using a large-scale real-world

data set to conduct a quantitative survey of the overall quality of the early commenters.. Moreover, we formalize the early analyst expectation task as a contest issue and propose a novel installing based positioning way to deal with this undertaking. As far as anyone is concerned, the assignment of early commentator forecast itself has gotten almost no consideration in the writing. Our commitments are summed up as follows:
The emergence of e-commerce websites has enabled users to publish or share purchase experiences by posting product reviews, which usually contain useful opinions, comments and feedback towards a product. It is important for companies to identify and analyze reviews since their feedbacks can help companies to adjust marketing strategies and improve product designs, which can eventually lead to the success of their new products.

In this project, we take the initiative to study the behaviour characteristics of posted reviews on representative e-commerce platforms. We aim to conduct effective analysis and make accurate prediction towards product improvement.

With the thriving development of e-commerce, people have gotten used to online consumption and writing reviews about their shopping experience on merchants / review sites. These moralistic content are valuable resources for future customers to make decisions and merchants to improve their products and / or services. However, with the rapid increase in the number of comments, people have to face a serious problem of information overload. To alleviate this problem, many opinion extraction techniques have been proposed, such as opinion summaries, surveys, and comparative analysis . The key challenge is how to accurately predict the sentiment orientation of sentences from comments. The popular sentiment classification methods generally fall into two categories: (1) dictionary-based methods, and (2) machine learning methods. The lexical-based method usually adopts a strategy that first constructs an emotional

dictionary of opinion words (for example, "wonderful", "disgusting"), and then designs classification rules based on the opinion words appearing in and prior syntactic knowledge. Although it is effective, this method requires a lot of effort in dictionary construction and rule design. In addition, lexical-based methods cannot handle implicit opinions, such as "I bought a mattress a week ago, and a valley appeared today" and other objective statements. As mentioned here , it is also an important form of opinion. Factual information is often more helpful than subjective perception. The lexical-based method can only temporarily handle implicit feedback. The first machine learning-based sentiment classification work applied popular machine learning algorithms (such as Naive Bayes) to this problem. After that, most of the research in this address was carried out around engineering features to obtain better sorting performance. has explored different types of functions, page . For example, ngrams, Partofspeech (POS) information and syntactic relations, etc.

Features engineering also works with many human efforts, and a suitable feature set for a domain may not produce good performance for other domains. In recent years, deep learning has arisen as an effective way to solve emotional classification problems. Deep neuronal networks are essentially learning high-level data expressions, which avoid problematic work such as functional engineering. The second advantage is that the exponential function is more expensive than the models with shallow models. However, deep learning success depends to a large extent on the availability of Largescale training data. It is very problematic to show a series of statements. Fortunately, on most traders' websites / reviews, customers can summarize their opinions with the general rating score (usually at 5 stars scale). The evaluation reflects the general emotions of customer reviews, and is already used for emotional analysis. However, the review evaluation is not a reliable label for the components. 5stars

reviews may contain negative prayers, and we will also see a positive word in a 1 star review. Therefore, dealing with binarized rating as an emotional label could be confused with emotional classifiers for examination statements. Despite the promising performance of deep learning about emotional classification, previous work tried to take advantage of unwanted evaluations to train serious models. This work suggests a new Deep Learning Framework for the Emotional Classification of the Revision Declaration. This framework deals with the review assessment as a weak tag and trains a deep neural network. For example, in 5stars scale, you can qualify 3 stars up / down as a positive / negative weak tag. The frames are generally composed of two steps. In the first step, instead of predicting the tag of the centimeter, it will learn an embedded space (high-level layer in the neuronal network) that reflects the general distribution of the distribution of sentences of a large amount of declaration of weak expression. That is, although prayers with the same weak tag are forced to each other, the statements with different weak tags are separated from each other. Proposition to reduce the impact of sentences with rating orientation (hereinafter referred to as illegal sentences) and approve the relative distance between sentences in embedded spaces through the loss of ranking. The second step adds a rating layer in the built-in layer and uses stagnated statements to use a deep network for fineeetune. The frame is called weak deep embedding (WDE). With regard to the structures of the network, it is adopted to learn to extract vectors of feature films of fixed length of the review declaration, that is, convolutionary function extractors and short short memory (LSTM). See the previous model as a neuronal network based on the WDECNN in a slight abuse of the concept. The latter is called WDE based on LSTM (WDELSTM). Then, it synthesize the extracted function to calculate the high level function (embedding) and calculate the product context aspects information (for example, the

mobile phone screen). The aspect entry represents the prior knowledge of the orientation of the declaration. Use of two huge certifiable data sets, we will announce the first report to draw an initial analyst at your web-based business website. We discussed quantitatively the attributes of the initial analyst and its impact on the morbidity rate. Our accurate inspection helps promote social science hypotheses and financial aspects. We trained the survey publications as a multiplayer contest game, and we encourage a positioning model based on the installation for the initial prognosis of the analyst. Our model can manage cold boot problems by integrating the data from the element side.

1.2 PROBLEM STATEMENT

Now days many techniques is available to predict about review, these techniques is made for to predict review, but accuracy is less to predict for all review with high accuracy.

1.3 OBJECTIVE

Objective of this thesis is to predict review for that study the conduct qualities of early reviewer through their posted audits on our shopping gateway. In explicit, we partition item lifetime into three back to back stages, in particular early, lion's share. A client who has posted a survey in the beginning phase is considered as an early analyst.

1.4 MOTIVATION

In this thesis proposed that significant for organizations to distinguish early commentators since their criticisms can assist organizations with changing promoting techniques and improve item plans, which can in the long run lead to the achievement of their new items. Thus, early commentators become the accentuation to screen and

draw in at the early advancement phase of an organization. The essential job of early audits has drawn in broad consideration from promoting experts to actuate shopper buy expectations.

1.5 SCOPE OF WORK

In this thesis proposed that an evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. In this, technically advanced decade, the significance of decision making of market strategy depends highly on the analysis of marketing surveys and product reviews. Thus we have tried to compare and learn some of the best models of sentiment analysis. Maximum researchers have tried to find out the overall analysis of the reviews but hardly anyone used that analysis for product marketing and enhancement.

1.6 THESIS ORGANIZATION

The organization of rest of the dissertation is as follows:

Chapter 2

In this chapter, there are reviews from various national and international journals and publications. It is done to identify the real problem statement for doing appropriate research.

Chapter 3

This chapter includes material and methodoligies used in our project.

Chapter 4

In this chapter the metrics that were utilized to measure the performance of proposed work along with diagrams that illustrate the performance measurements, the implementation details and results of the detection mechanism is discussed.

Chapter 5

In this chapter conclusion and some of the future scopes is discussed of this work.

CHAPTER – 2

LITERATURE SURVEY

2.1 LITERATURE SUMMARY

Najma Sultana et al [2019] Sentiment research is defined as the way to extract data, opinions, surveys or sentences to predict the feeling of the sentence through common language handling (NLP). "Positive" "Negative" "Unbiased . It analyses the data and marks the 'better' and the 'more regrettable' supposition as sure and negative individually. Hence, in the previous years, the World Wide Web (WWW) has become an immense wellspring of crude information produced custom or client. Utilizing web-based media, online business site, films audits, for example, Facebook, reviewer, Amazon, Flipkart and so on client share their perspectives, emotions in an advantageous way. Feeling examination is text based investigation, however there are sure difficulties to locate the precise extremity of the sentence.**[1]**

Alpna Patel [2019] E-business locations are being acclaimed and tremendous progress has been made in ads. In online business applications, customer devotion and persuasive fulfilment of customer needs are more relevant. To achieve effective results and contribution to online business applications, the necessity for suitably amassing customer reviews and analysis is required. The proposed work sufficiently crushes the issue of considering the customer lead from their online thing reviews. The experts are assembled into three classes explicitly unrefined, standard and present. This is sensible for both thing and expert portrayal. This portrayal is performed and associated among various investigators and their bit of leeway. The proposed work can in like manner choose the thing reputation and the thing use by different customers from web shopping destinations. In this paper to perform sentiment analysis, researcher has used IMDB movie review dataset and RNN as its machine learning technique **[2]**

Yoon-Joo Park et al [2018] Online customer studies are a prudent kind of casual (WOM) which accept an unyieldingly huge part in web business. Mediocre quality reviews will, in any case, consistently produce per users of trouble review. The inspiration driving this paper is to therefore foresee the convenience of studies . The inspiration driving this paper is to therefore foresee the convenience of studies. The results show that reviews for different thing types have particular mental and phonetic characteristics and the factors impacting the study steadiness of them are also exceptional. Our disclosures in like manner show that the assistance vector backslide methodology predicts review uphold most decisively among the four strategies for all of the five datasets. This assessment adds to improving powerful use of online surveys**.[3]**

Sunil Saumya1 et al [2018] In the hundreds and even in the massive numbers for some famous things, the item reviews are posted online. Dealing with an especially gigantic volume of continually delivered online substance is a troublesome task for buyers, sellers and even researchers. The purpose behind this assessment is to rank the amazing number of reviews using their foreseen convenience score. The system is specially made the reviews into low or high type by self-assertive forest area classifier. The convenience score of the brilliant reviews is simply foreseen using point boosting regression. The help score of the sub-par quality overviews isn't resolved considering the way that they are never going to be in the top k reviews. They are essentially added at the completion of the review summary to the overview posting site. The proposed system gives sensible study circumstance on review posting pages and making all superb studies recognizable to customers on the top. The exploratory results on data from two famous**[4].**

Liao, Shiyang [2017] gave an approach to comprehend real situations with the Sentiment Analysis of a Reviewer data centred on Deep learning techniques. With the suggested method, it was viable to forecast user satisfaction on a product. Lately, Deep Learning was competent to resolve problems in voice recognition or computerized vision. CNN worked fine for image analysis together with classification. An imperative reason to employ CNN for image analysis and image classification was that the CNN could extort an area of features as of global information precisely and also it was competent to regard the relations amongst those features. The above solution could attain the utmost accuracy in analysis together with classification. For NLP 'texts' data features could also be extorted piece by piece. Regarding the relations amongst those features without considering the context or complete sentence might incorrectly interpret the sentiment. And, it was the most effectual method to perform image classification. CNN comprised a convolutional layer to extort information by a large piece of text **[5]**

Xing Fang et al[2015] Sentiment analysis or evaluation mining is one of the main tasks of NLP (Natural Language Processing). Feeling research has given a lot of thinking of late. In this paper, we expect to deal with the issue of end furthest point order, which is one of the fundamental issues of idea examination. A general cycle for incline limit course of action is proposed with low down communication portrayals. Data used in this assessment are online thing overviews assembled from Amazon.com. Preliminaries for both sentence-level request and review level course of action are performed with promising outcomes. At last, we also give understanding into our future work on idea examination. **[6]**

Daichi Imamori , Keishi Tajima [2015] gave way to deal with idea Due to the dynamicity, new notable records reliably appear and disappear in scaled down scale writing for a blog organizations. Early distinguishing proof of new records that will wrap up standard in future is a fundamental issue that has a couple of utilizations, for instance, incline area, viral displaying, and customer proposal. Assessment of conspicuousness of a record is also significant for approximating the idea of information it posts. Assessment of the idea of information is imperative in various applications, yet it is generally difficult to measure it without human intervention. Similar idea has also been successfully associated with limited scope web diaries with interfacing limits.**[7]**

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. In this, technically advanced decade, the significance of decision making of market strategy depends highly on the analysis of marketing surveys and product reviews. Thus, in this literature we have tried to compare and learn some of the best models of sentiment analysis. Maximum researchers have tried to find out the overall analysis of the reviews but hardly anyone used that analysis for product marketing and enhancemen

Table 1: Different technique used for predicting reviews

YEAR	AUTHOR	PURPOSE	TECHNIQUE
2019	N. Sultana, P. kumar	Sentiment analysis for product review	NB,SVM & Linear model algorithm
2019	Alpna Patel & Arvind Kumar	Sentiment analysis by using RNN	Used RNN
2018	Yoon-Joo Park	Online review helpfulness across different product type	Linear regression, SVM, Random Forest,M5P
2018	S. Saumya, J. Prakash	Ranking Online customer reviews	Cosine similarity, SVM and Random Forest
2017	Liao Shiyang, Junbo Wang.	Sentiment Analysis of reviewer data	NLP and CNN for classification
2015	D. Imamori & K Tajima	Predicting popularity of reviewer account.	Cosine similarity and SVM
2015	Xing Fang & Justin Zhan	Sentiment analysis using product review data	NB, SVM, Random Forest

CHAPTER – 3

MATERIALS And METHODS

This work depicts about the prerequisites. It determines the equipment and programming prerequisite that are needed for software to keeping in mind the end goal, to run the application appropriately. The SoftwareRequirement Specification (SRS) is clarified in point of interest, which incorporates outline of this exposition and additionally the functional and non-practical necessity of this thesis.

3.1 GENERAL DESCRIPTION

Most past strategies have planned explicit indicators utilizing various highlights for every one of these three classes. The methodology we guarantee here contrasts from these current methodologies in that we propose a solitary learning based discovery system to recognize every one of the three significant classes of items. To additionally improve the speculation execution, we propose an item sub classification technique as a methods for catching the intra-class variety of articles.

3.1.1 USERS PERSPECTIVE

The Characteristic of this task work is to give information adaptability security while sharing information through cloud. It gives a proficient approach to share information through cloud.

3.2 FEASIBILITY STUDY

Believability is the determination of paying little respect to whether an undertaking justifies action. The framework followed in building their strength is called acceptability Study, these kind of study if a task could and ought to be taken.

Three key thoughts included in the likelihood examination are:

- Technical Feasibility

- Economic Feasibility
- Operational Feasibility

3.2.1 TECHNICAL FEASIBILITY

Here it is considered with determining hardware and programming, this will effective fulfill the client necessitythe specialized requires of the framework should shift significantly yet may incorporate

- The office to create yields in asecified time.
- Reaction time underparticular states.
- Capacity to deal with a particular segment of exchange at a specific pace.

3.2.2 ECONOMIC FEASIBILITY

Budgetary examination is the often used system for assessing the feasibility of a projected structure. This is more usually acknowledged as cost/favorable position examination. The method is to center the focal points and trusts are typical casing a projected structure and a difference them and charges. These points of interest surpass costs; a choice is engaged to diagram and realize the system will must be prepared if there is to have a probability of being embraced. There is a consistent attempt that upgrades in exactness at all time of the system life cycle.

3.2.3 OPERATIONAL FEASIBILITY

It is for the most part identified with human association and supporting angles. The focuses are considered:

What alterations will be carried through the framework?

- What authoritative shapes are dispersed?
- What new aptitudes will be needed?
- Do the current framework employee's individuals have these aptitudes?
- If not, would they be able to be prepared over the span of time?

3.3 TECHNOLOGY USED

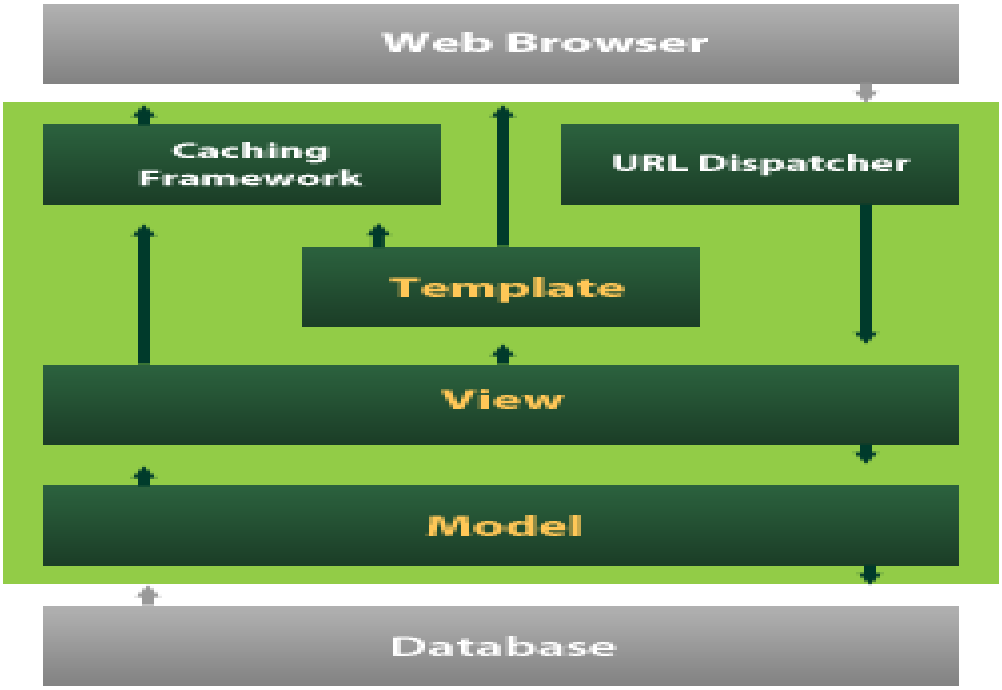
3.3.1 PYTHON

Python is a general-purpose interpreted, interactive, object oriented, and high-level programming language. An interpreted language Python has a design philosophy that emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer lines of code than might be used in languages such as C++or Java. It provides constructs that enable clear programming on both small and large scales.Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. C Python is managed by the non-profit Python Software Foundation. Python features a dynamic type system and automatic memory management. It supportsmultiple programming paradigms, including object-oriented, imperative functional and procedural, and has a large and comprehensive standard library

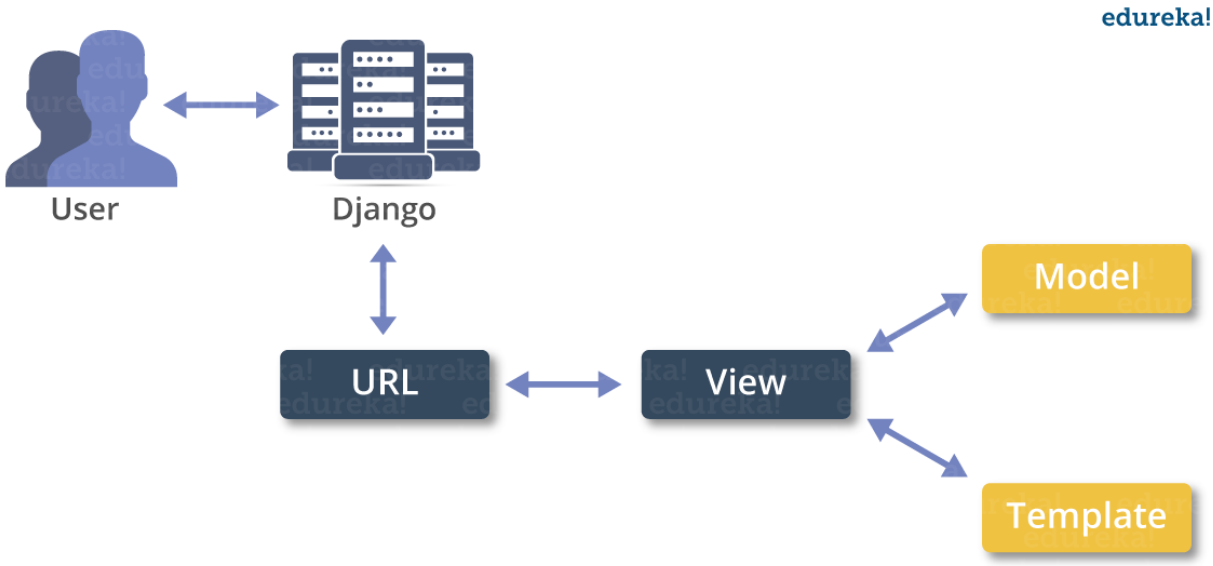
3.3.2 DJANGO

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Django's primary goal is to ease the creation of complex, database driven websites. Djangoemphasizes reusabilityand "pluggability" of components rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.



Django also provides an optional administrative create, read update and delete interface that is generated dynamically through introspection and configured via admin models



3.4 INPUT AND OUTPUT DESIGN

3.4.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

3.4.2 OBJECTIVES

Input Design is the process of converting a user-oriented description of the input into a computer based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

It is achieved by creating user friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

3.4.3 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision making.

- Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify

the specific output that is needed to meet the requirements.

- Select methods for presenting information.
- Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities problems, or warnings.
- Trigger an action.
- Confirm an action.

3.5 INTRODUCTION TO SYSTEM ANALYSIS

3.5.1 SYSTEM

A system is an orderly group of interdependent components linked together according to a plan to achieve a specific objective.Its main characteristics are organization, interaction, interdependence, integration and a central objective.

3.5.2 SYSTEM ANALYSIS

System analysis and design are the application of the system approach to problem solving generally using computers. To reconstruct a system the analyst must consider its elements output and inputs, processors, controls feedback and environment.

3.6 EXISTING SYSTEM

Previous research strongly emphasized the phenomenon that individuals are strongly influenced by the decisions of others, which can be explained by herd behavior. The

influence of early comments on later purchases can be understood as a special case of the herd effect. The first review contains important product reviews of previous users, which is a valuable reference resource for future purchase decisions. As shown in the figure, when consumers use other people's product ratings to estimate product quality online, herd behavior will appear in the online shopping process. Different from the existing herd behavior research, we focus on using large-scale real-world data sets to quantitatively analyze the general characteristics of the first batch of reviewers. In addition, we formalized the early reviewer prediction task as an ability problem, and proposed a new inlays-based scoring method for this task. As far as we know, the task of early reviewers' prediction itself has received little attention in the literature. Our contributions are summarized as follows: We present the first study that uses two large real-world data sets to characterize early reviewers on e-commerce websites. We quantitatively analyzed the characteristics of the first reviewers and their influence on the popularity of the product. Our empirical analysis supports many theoretical conclusions of sociology and economics. We treat the review publishing process as a multiplayer competition game, and developed an embedded ranking model to predict the first reviewers. Our model can solve the cold start problem by incorporating supplementary product information.

3.7 PROPOSED SYSTEM

To predict the initial reviewer, we propose a new approach when observing the review processing process as a multiplayer competitive game. Only the most competitive user can be W. R.t from the initial reviewer. To the products. The process of conflict can be degraded even more to several pairs comparisons between two players. In the two-tier competition, the winner defeats the loser early in the timestamp. Inspiring the recent advances in learning dispersion expression, first mapping users and products to

the same integrated space, and then determining the order of a pair of users who gave the product according to their respective distances, we propose using models Embedded margin base. Representation of the product.

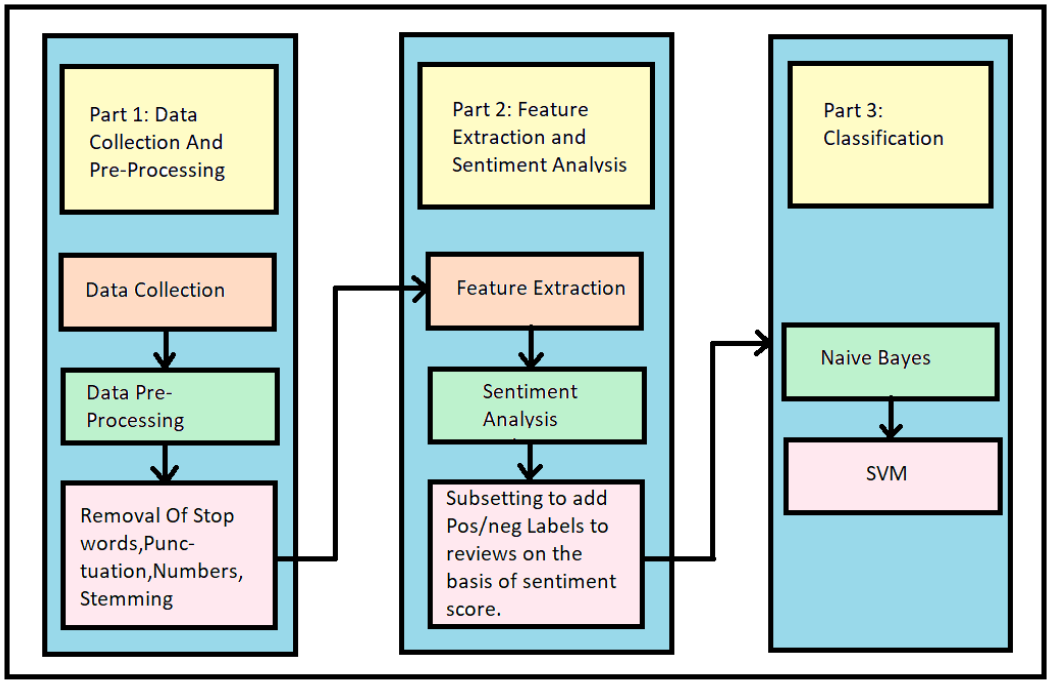


Figure 1: Proposed Framework

3.8 MODULES

- **UPLOAD PRODUCTS**

Uploading the products is done by admin. Authorized person is uploading the new arrivals to system that are listed to users. Product can be uploaded with its attributes such as brand, color, and all other details of warranty. The uploaded products are able to block or unblock by users.

- **PRODUCT REVIEW BASED ORDER**

The suggestion to user's view of products is listed based on the review by user and rating to particular item. Naïve bayes algorithm is used in this project to develop the whether the sentiment of given review is positive or negative. Based on the output of algorithm suggestion to users is given. The algorithm is applied and lists the products in user side based on the positive and negative.

- **RATINGS AND REVIEWS**

Ratings and reviews are main concept of the project in order to find effective product marketing. The main aim of the project is to get the user reviews based on how they purchased or whether they purchased or not. The major find out of the project is when they give the ratings and how effective it is. And this will helpful for the users who are willing to buy the same kind of product.

- **DATA ANALYSIS**

The main part of the project is to analysis the ratings and reviews that are given by the user. The products can be analysis based on the numbers which are given by user. The user data analysis of the data can be done by charts format. The graphs may vary like pie chart, bar chart or some other charts.

3.9 ALGORITHM

- **Preprocessing**

In this calculation, the reviews which are unfamiliar made to information base from the reviewer API, these reviews involve trivial words, whitespaces, hyperlinks and remarkable characters. First we need to do isolating interaction by emptying each and every pointless word, whitespaces, hyperlinks and remarkable characters.

The preprocessing steps mean to start the element extraction interaction and begin separating sacks of words from the examples. One of the primary center is to decrease the last measure of highlights removed. Undoubtedly, highlights decrease is significant to improve the precision of the forecast for both point displaying and conclusion investigation. Highlights are utilized to address the examples, and the more the calculation will be prepared for a particular element the more exact the outcomes will be. Consequently, if two highlights are comparable it is advantageous to consolidate them as one exceptional component. Also, if a component isn't important for the investigation, it tends to be taken out from the pack of words.

• **Lower capitalized letters**: The initial phase in the preprocessing is to go through all the information and change each capitalized letter to their relating lowercase letter. When handling a word, the examination will be case delicate and the program will consider "information" and "Information" as two entirely unexpected words. It is significant that, these two words are considered as similar highlights. Something else, the calculations will influence notions which may vary to these two words. For instance, on these three sentences: "information are acceptable", "Marvelous information", and "Awful Data". The first and second sentences both contain "information" and are positive, the third sentence contains "Information" and is negative. The calculation will figure that sentences containing "information" are bound to be positive and those containing "Information" negative. In the event that the uppercases had been eliminated the calculation would have had the option to figure that the way that the sentence contains "information" isn't extremely pertinent to identify whether or the sentence is positive. This preprocessing step is much more significant since the information are recovered from Reviewer. Online media clients are regularly writing in capitalized regardless of whether it isn't needed, hence this

preprocessing step will betterly affect web-based media information than other "traditional" information.

• **Remove URLs and client references**: Reviewer permits client to incorporate hashtags, client references and URLs in their messages. By and large, client references and URLs are not applicable for investigating the substance of a book. Hence, this preprocessing step depends on normal articulation to discover and supplant each url by "URL" and client reference by "AT_USER", this permits to decrease the aggregate sum of highlights extricated from the corpus [2]. The hashtags are not eliminated since they frequently contain a word which is important for the investigation, and the "#" characters will be taken out during the tokenization interaction. Remove digits: Digits are not significant for breaking down the information, so they can be taken out from the sentences. Besides, now and again digits will be blended in with words, eliminating them may permit to relate two highlights which may have been considered diverse by the calculation in any case. For instance, some information may contain "iphone", when other will contain "iphone7". The tokenization cycle, which will be presented later.

• **Remove stop words**: In normal language preparing, stop words are regularly taken out from the example. These stop words will be words which are normally utilized in a language, and are not applicable for a few regular language preparing techniques, for example, theme demonstrating and supposition investigation [10]. Eliminating these words permits to decrease the measure of highlights extricated from the examples.

•

- **Self-Learning and word standardization System**

In this calculation, first we need to instate the word reference (first accentuation dictionary).In the vocabulary generally we need to present the positive, negative fair and things. Each and every enormous datum and data mining adventures considering the pre-arranged data, without arranged data (presentation of words).So instatement of the pre-arranged data is indispensable. In oneself learning structure, we are doing word standardization, here we are not considering past, present and future status of the words, just we are pondering the word.

- **Sentiment Analysis**

In this calculation, pre-processed reviews are brought from the data set individually. In any case we require check individually watchword whether that expression is thing are not, if thing we will oust it from the particular audit. After that the remainder of the watchwords checked with evaluation create, whether or not those expressions are sure assessment or adverse end or unbiased inclination. The remainder of the watchwords in the review which doesn't has a spot with any of the assumption will be consigned fleeting end considering the more check of positive, negative and fair. In the subsequent cycle if the reaming word crosses the restriction of positive, negative or impartial, that watchword everlastingly included as improvement in the vocabulary.

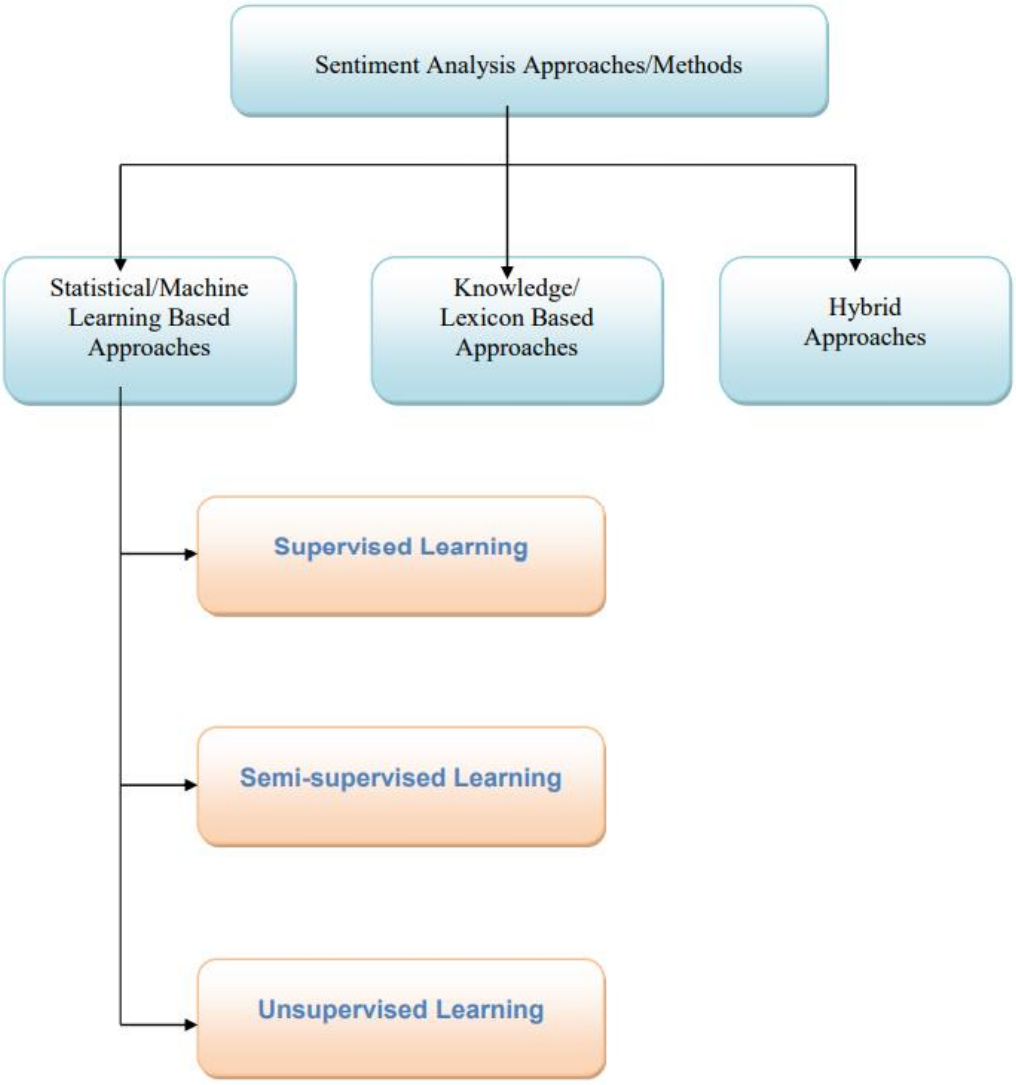


Figure 2: Sentiment Analysis Approaches/Method

ALGORITHM STEP IN SENTIMENT ANALYSIS

Step1:Get_some_sentiment_examples

As for every supervised learning problem, the algorithm needs to be trained from labeled examples in order to generalize to new data.

Step2:Extract_features_from_examples

Transform each example into a feature vector. The simplest way to do it is to have a vector where each dimension represents the frequency of a given word in the document.

Step3:Train_the_parameters

This is where your model will learn from the data. There are multiple ways of using features to generate an output, but one of the simplest algorithms is logistic regression. Other well-known algorithms are Naive Bayes. In the simplest form, each feature will be associated with a weight. Let's say the word "love" has a weight equal to +4, "hate" is -10, "the" is 0 ... For a given example, the weights corresponding to the features will be summed, and it will be considered "positive" if the total is > 0, "negative" otherwise. Our model will then try to find the optimal set of weights to maximize the number of examples in our data that are predicted correctly.
If you have more than 2 output classes, for example if you want to classify between "positive", "neutral" and "negative", each feature will have as many weights as there are classes, and the class with the highest weighted feature sum wins.

Step4:Test_the_model

After we have trained the parameters to fit the training data, we have to make sure our model generalizes to new data, because it's really easy to over fit. The general way of regularizing the model is to prevent parameters from having extreme values.

SVM

Support Vector Machine (SVM) is a supervised machine learning model that uses classification algorithms to solve two sets of classification problems. After providing the SVM model of the labeled training data set for each category, they were able to

classify the new text.

Compared to newer algorithms, such as neural networks, they have two main advantages: in the case of a limited number of samples (thousands), they are faster and have better performance. This makes the algorithm well suited for text classification problems, in which case it is common to access data sets containing up to thousands of labeled samples.

Naive Bayes

- Bayes classifiers are a group of straightforward probabilistic classifiers dependent on applying Bayes' hypothesis with solid (gullible) freedom presumptions between the highlights.

- Naive Bayes classifiers are exceptionally adaptable, requiring various boundaries direct in the quantity of factors (highlights/indicators) in a learning issue. Greatest probability preparing should be possible by assessing a shut structure articulation, which takes direct time, as opposed to by costly iterative guess as utilized for some different sorts of classifiers.

- In the insights and software engineering writing, credulous Bayes models are known under an assortment of names, including basic Bayes and autonomy Bayes. Every one of these names reference the utilization of Bayes' hypothesis in the classifier's choice guideline, yet innocent Bayes isn't (really) a Bayesian technique

- Naive Bayes is a straightforward method for building classifiers: models that appoint class marks to issue cases, addressed as vectors of highlight esteems, where the class names are drawn from some limited set. It's anything but a solitary calculation for preparing such classifiers, yet a group of calculations dependent on a typical guideline:

all gullible Bayes classifiers expect that the worth of a specific element is free of the worth of some other component, given the class variable. For instance, an organic product might be viewed as an apple on the off chance that it is red, round, and around 10 cm in breadth. An innocent Bayes classifier thinks about every one of these highlights to contribute autonomously to the likelihood that this organic product is an apple, paying little heed to any potential connections between's the shading, roundness, and breadth highlights.

- For a few sorts of likelihood models, credulous Bayes classifiers can be prepared productively in a managed getting the hang of setting. In numerous down to earth applications, boundary assessment for innocent Bayes models utilizes the strategy for most extreme probability; all in all, one can work with the credulous Bayes model without tolerating Bayesian likelihood or utilizing any Bayesian strategies.

- Despite their guileless plan and obviously distorted suspicions, credulous Bayes classifiers have functioned admirably in numerous intricate genuine circumstances. In 2004, an examination of the Bayesian order issue showed that there are sound hypothetical purposes behind the clearly doubtful viability of credulous Bayes classifiers. In any case, a complete examination with other order calculations in 2006 showed that Bayes arrangement is outflanked by different methodologies, like supported trees or irregular woods

3.10 METHODOLOGY

- **DATASET**

The system mainly deals with the problem of cold start. The significant role of early reviews has paying attention in depth attention from selling practitioners to induce client purchase intention. So this need a set of data related to the shopping products with respective user profile and product shopping review. The data collected from real time websites such as Amazon, flipkart and snapdeal shopping, doesn't have some attributes available in the website, so the experiment takes the synthetic dataset which is described below.

An e-commerce domain was created to collect the dataset of various product reviews.

siva	9685741230	gokul@gmail.com	shirt	Van Heusen	2	good
haris	9685744565	haris@gmail.com	T-shirt	Parx	4	It is worst
haris	9685744565	haris@gmail.com	T-shirt	Parx	1	poor product i have ever
haris	9685744565	haris@gmail.com	shirt	Van Heusen	5	worst product I have e purchased
haris	9685744565	haris@gmail.com	shirt	Van Heusen	3	not good
sabari	9685741230	gokul@gmail.com	cotten pant	LEVI'S	3	poor purchase
haris	9685744565	haris@gmail.com	shirt	Peter England	2	great purchase
haris	9685744565	haris@gmail.com	mobile accessories	Huwai	5	super
haris	9685744565	haris@gmail.com	mobile accessories	Huwai	2	It is very good produ
haris	9685744565	haris@gmail.com	shirt	Van Heusen	1	it is good
haris	9685744565	haris@gmail.com	shirt	Peter England	2	good
priyanka	9685741230	gokul@gmail.com	suit	Ives.	5	It is worst
priyanka	9685741230	gokul@gmail.com	shirt	Peter England	4	poor product i have ever
priyanka	9685741230	gokul@gmail.com	cotten pant	LEVI'S	3	worst product I have e purchased

Figure 3: Product Reviews Dataset

- **Preprocessing And Cleaning**

In this algorithm, the reviews which are foreign made to database from the reviewer API, these reviews comprise of pointless words, whitespaces, hyperlinks and unique characters. First we have to do separating process by evacuating every single superfluous word, whitespaces, hyperlinks and extraordinary characters.

The preprocessing steps aim to begin the feature extraction process and start extracting bags of words from the samples. One of the main focus is to reduce the final amount of features extracted. Indeed features reduction is important in order to improve the accuracy of the prediction for both topic modeling and sentiment analysis. Features are used to represent the samples, and the more the algorithm will be trained for a specific feature the more accurate the results will be. Hence, if two features are similar it is convenient to combine them as one unique feature. Moreover, if a feature is not relevant for the analysis, it can be removed from the bag of words.We have used NLTK package for data preprocessing.

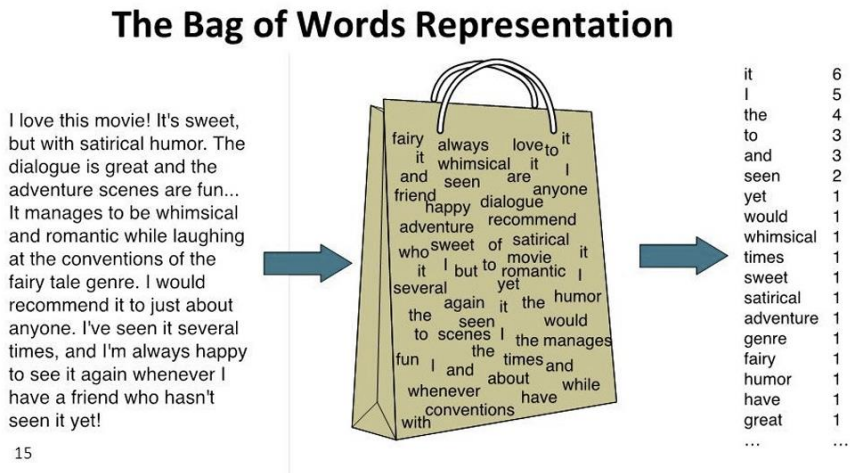


Figure 4: Simple illustration of Bag Of Words

• Lower uppercase letters: The initial phase in the preprocessing is to go through all the information and change each capitalized letter to their comparing lowercase letter. When preparing a word, the examination will be case touchy and the program will consider "information" and "Information" as two very surprising words. It is significant that, these two wordsare considered as similar highlights. Otherwise the algorithms will affect sentiments which may differ to these two words. For example, on these three sentences: “data are good”, “Awesome data”, and “Bad Data”. The first and second sentences both contain “data” and are positive, the third sentence contains “Data” and is negative. The algorithm will guess that sentences containing “data” are more likely to be positive and those containing “Data” negative. If the uppercases had been removed the algorithm would have been able to guess that the fact that the sentence contains “data” is not very relevant to detect whether or the sentence is positive. This preprocessing step is even more important since the data are retrieved from Reviewer. Social media users are often writing in uppercase even if it is not required, thus this preprocessing step will have a better impact on social media data than other “classical” data.

• Remove URLs and user references: Reviewer allows user to include hashtags, user references and URLs in their messages. In most cases, user references and URLs are not relevant for analyzing the content of a text. Therefore, this preprocessing step relies on regular expression to find and replace every URLs by “URL” and user reference by “AT_USER”, this allows to reduce the total amount of features extracted from the corpus [2]. The hashtags are not removed since they often contain a word which is relevant for the analysis, and the “#” characters will be removed during the tokenization process.

• Remove digits: Digits are not relevant for analyzing the data, so they can be removed from the sentences. Furthermore, in some cases digits will be mixed

with words, removing them may allow to associate two features which may have been considered different by the algorithm otherwise. For example, some data may contain “iphone8”, when other will contain “iphone 10”. The tokenization process, which will be introduced later.

- Remove stop words: In natural language processing, stop words are often removed from the sample. These stop words are words which are commonly used in a language, and are not relevant for several natural language processing methods such as topic modeling and sentiment analysis [10]. Removing these words allows to reduce the amount of features extracted from the samples.

- **Self-Learning and word standardization System**

In this algorithm, first we have to instate the word reference (first emphasis dictionary).In the lexicon for the most part we have to introduce the positive, negative nonpartisan and things. Every single huge datum and information mining ventures in view of the prepared information, without prepared information (introduction of words).

So instatement of the prepared information is vital. In the self-learning framework, we are doing word institutionalization, here we are not considering past, present and future status of the words, just we are thinking about the word.

3.11 SYSTEM DESIGN

3.11.1 Architecture Diagram

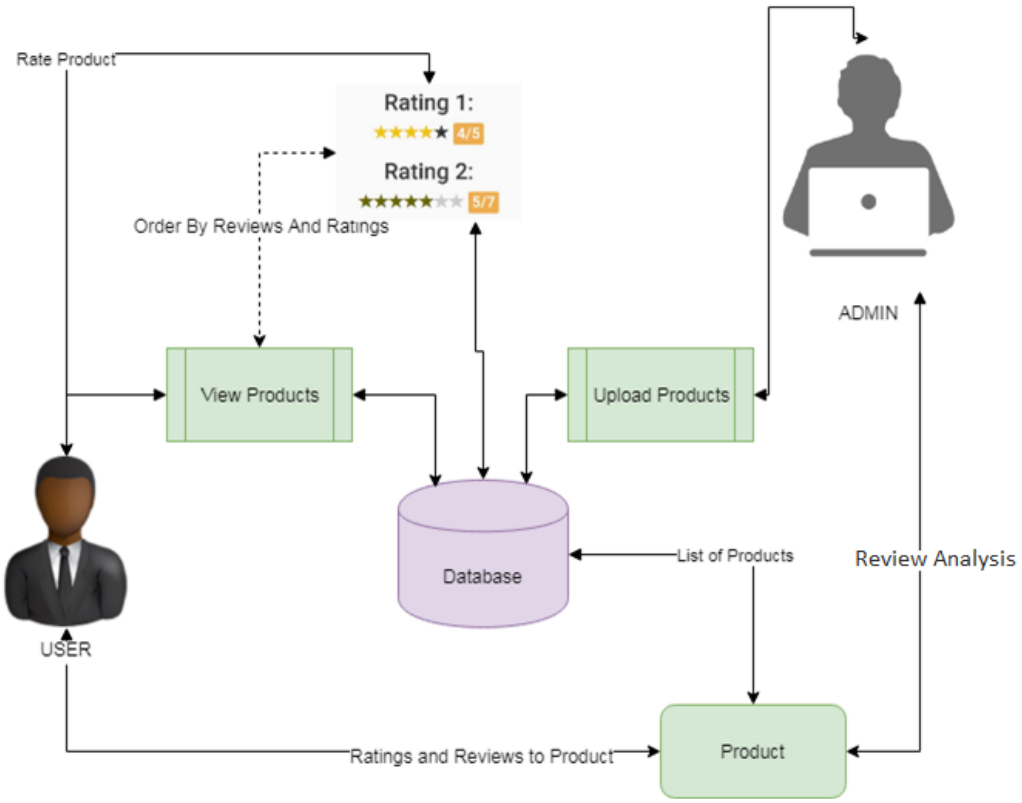


Figure 5: Architecture diagram

3.11.2 Component Diagram

- **User**

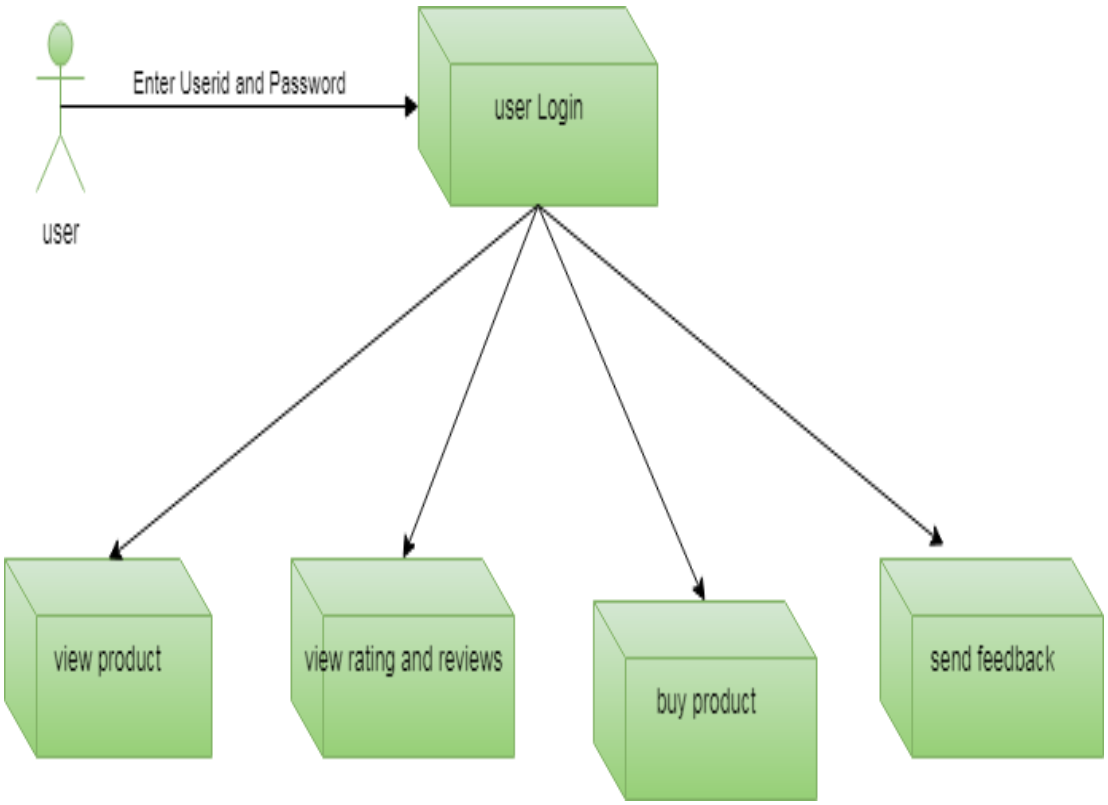


Figure 6: User Component diagram

- **Admin**

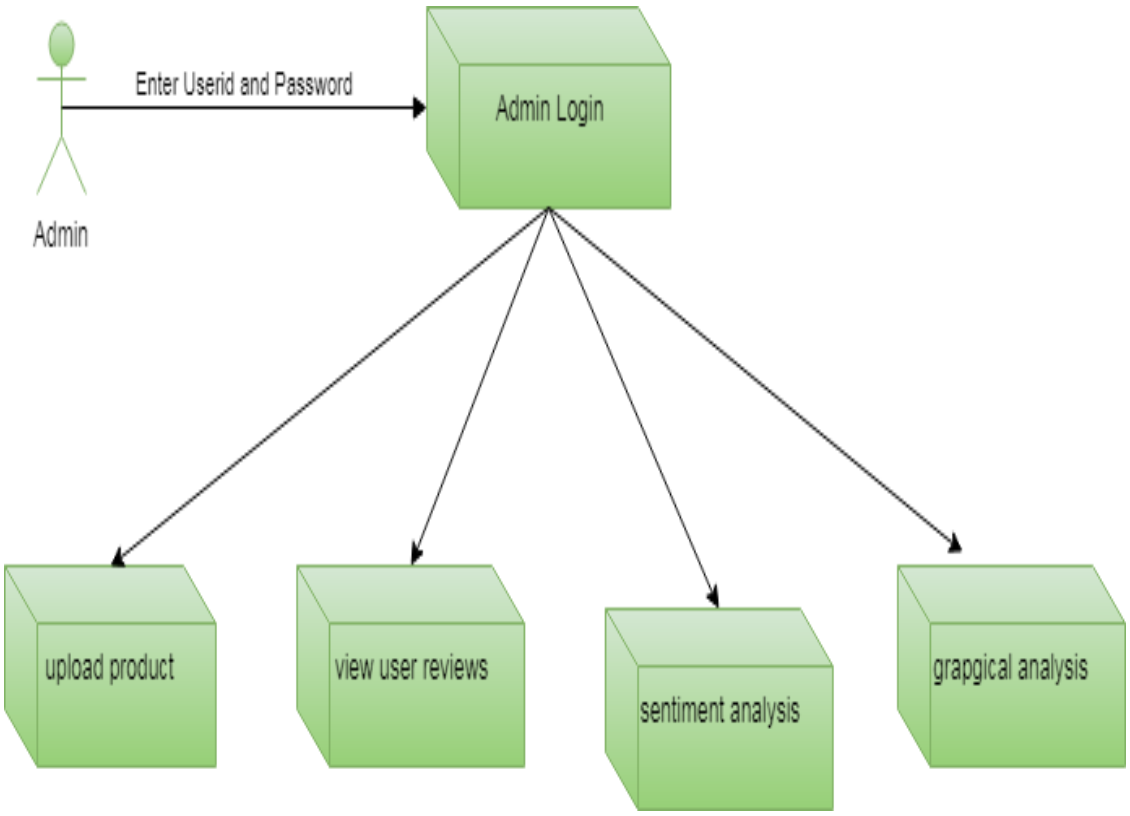


Figure 7: Admin Component diagram

3.11.3 ER Diagram

- **User**

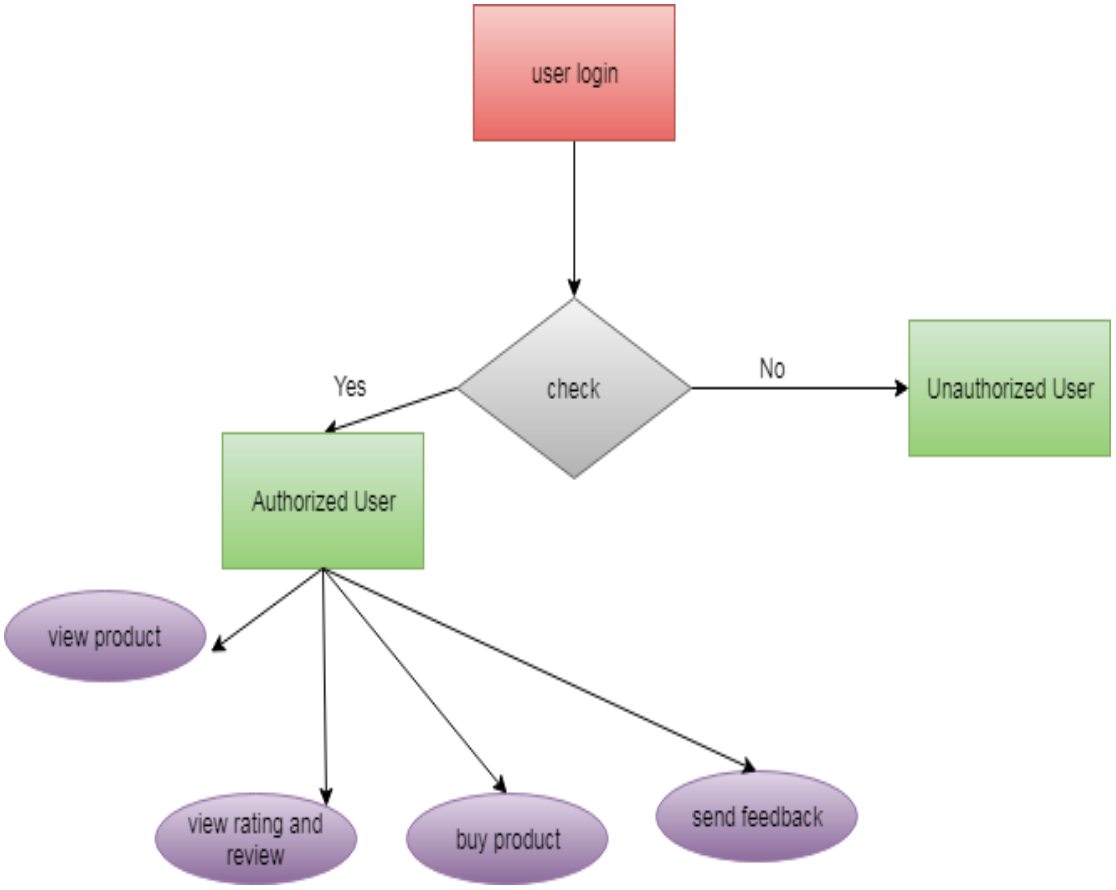


Figure 8: User ER diagram

Admin

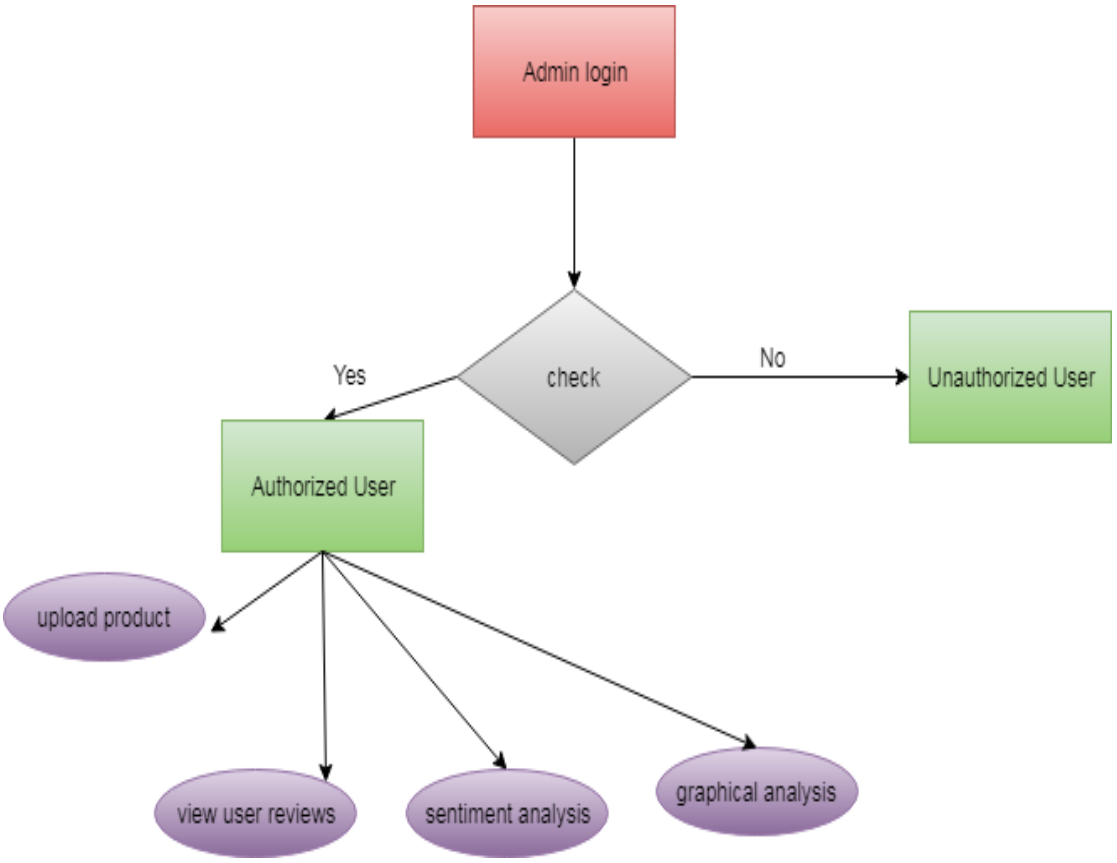


Figure 9: Admin ER diagram

3.11.4 Data Flow Diagram

- **User**

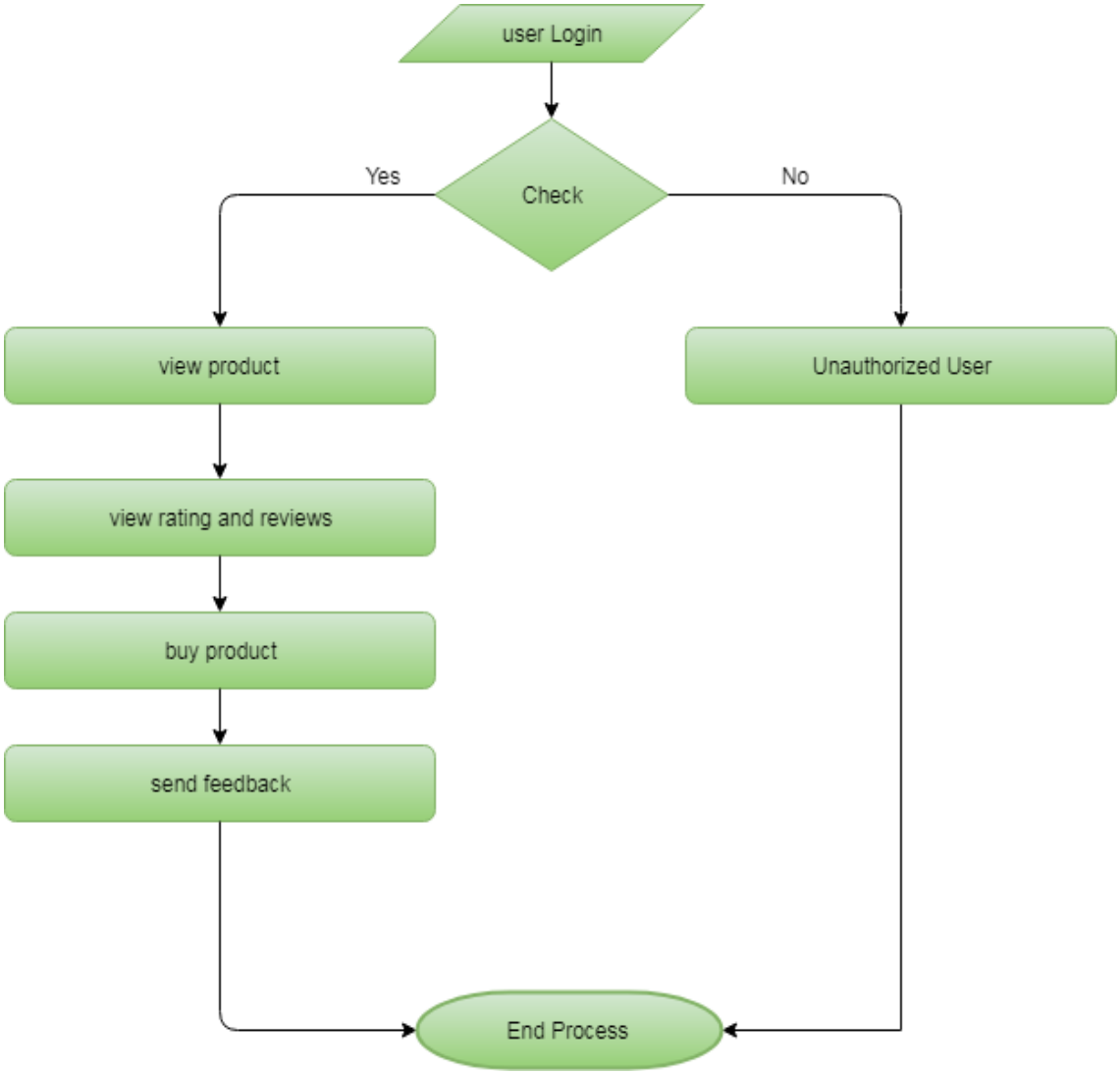


Figure 10: User Data flow diagram

- **Admin**

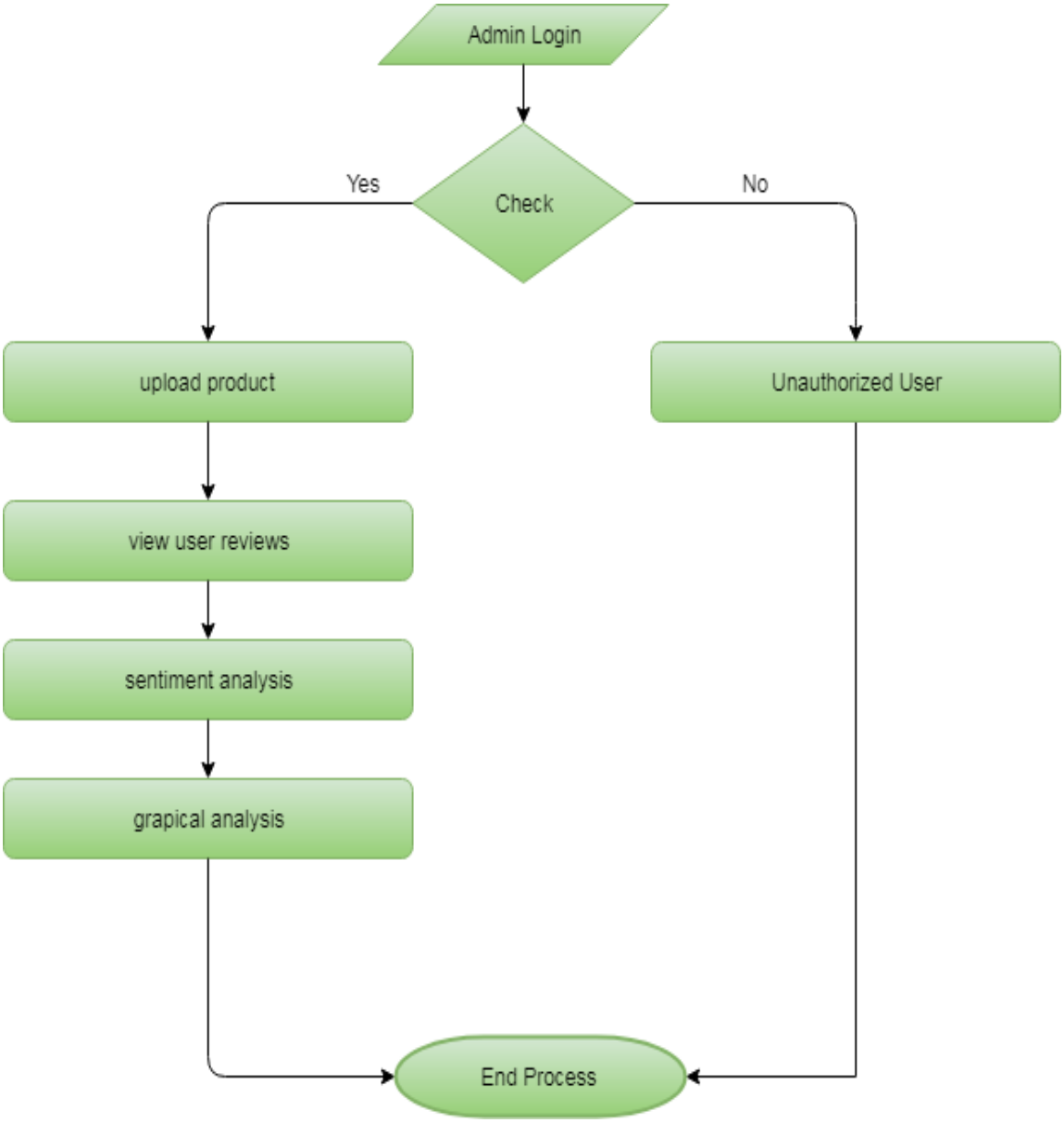


Figure 11: Admin Data flow diagram

CHAPTER – 4

RESULT ANALYSIS

AND

DISCUSSION

In this result chapter, we evaluate the outcome part, we step up and study the conduct qualities of posted audits on delegate online business stages. We plan to lead powerful investigation and make precise forecast towards item improvement. With the blasting of internet business, individuals are becoming accustomed to burning-through on the web and composing remarks about their buy encounters on vendor/audit Websites. These stubborn substance are important assets both to future clients for dynamic and to shippers for improving their items as well as administration.

These are the modules implemented in this research paper result part:

- **UPLOAD PRODUCTS**

Uploading the products is done by admin. Authorized person is uploading the new arrivals to system that are listed to users. Product can be uploaded with its attributes such as brand, color, and all other details of warranty. The uploaded products are able to block or unblock by users.

- **PRODUCT REVIEW BASED ORDER**

The suggestion to user's view of products is listed based on the review by user and rating to particular item. Naïve bayes algorithm is used in this project to develop the whether the sentiment of given review is positive or negative. Based on the output of algorithm suggestion to users is given. The algorithm is applied and lists the products in user side based on the positive and negative.

- **RATINGS AND REVIEWS**

Ratings and reviews are main concept of the project in order to find effective product marketing. The main aim of the project is to get the user reviews based on how they purchased or whether they purchased or not. The major find out of the project is when they give the ratings and how effective it is. And this will helpful for the users who are willing to buy the same kind of product.

- **DATA ANALYSIS**

The main part of the project is to analysis the ratings and reviews that are given by the user. The products can be analysis based on the numbers which are given by user. The user data analysis of the data can be done by charts format. The graphs may vary like pie chart, bar chart or some other charts.

RESULT

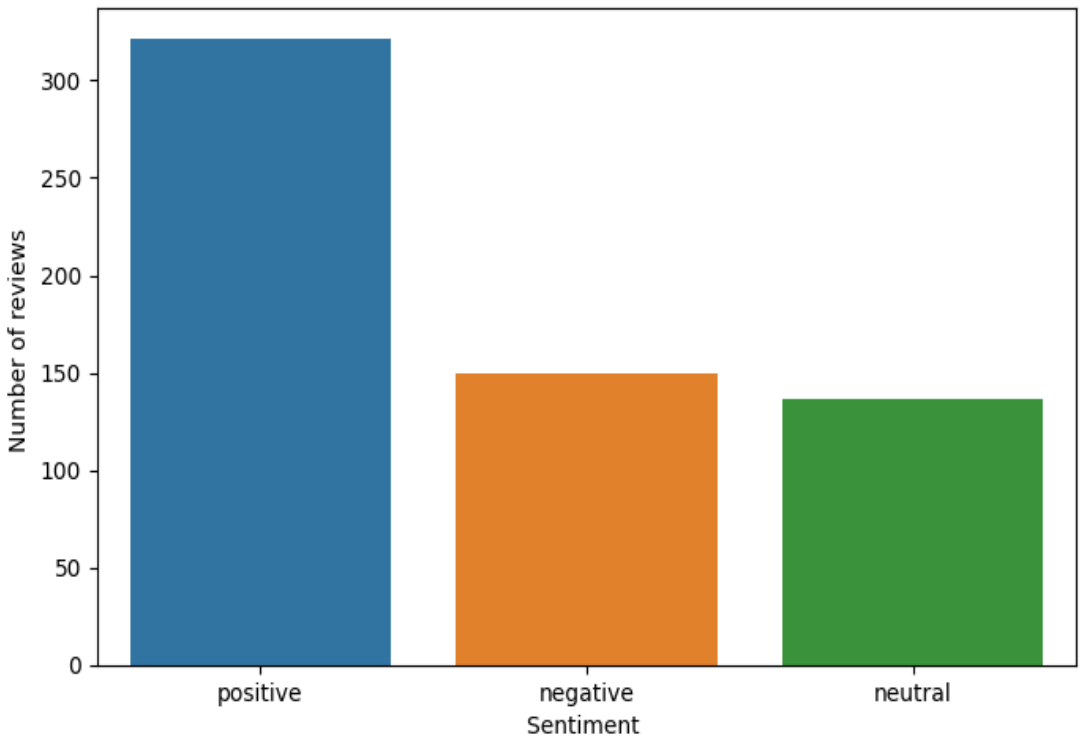


Figure 12: Polarity analysis

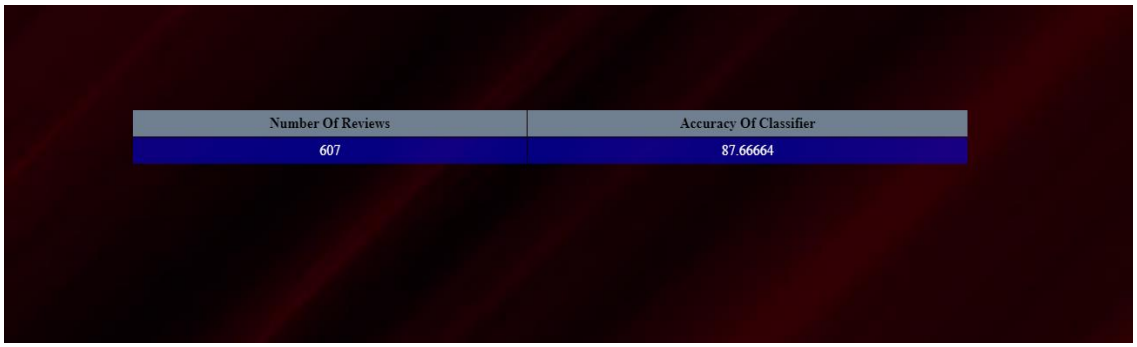


Figure 13: Accuracy of proposed Algorithm

Figure 11 shows the Polarity Analysis Of the Reviews given by users. Out of total 607 reviews, 315 are positive, 152 are negative and 140 are neutral reviews. Next figure is the table that shows the accuracy of our classifier. SVM has performed very good with an accuracy of 87.6664%.

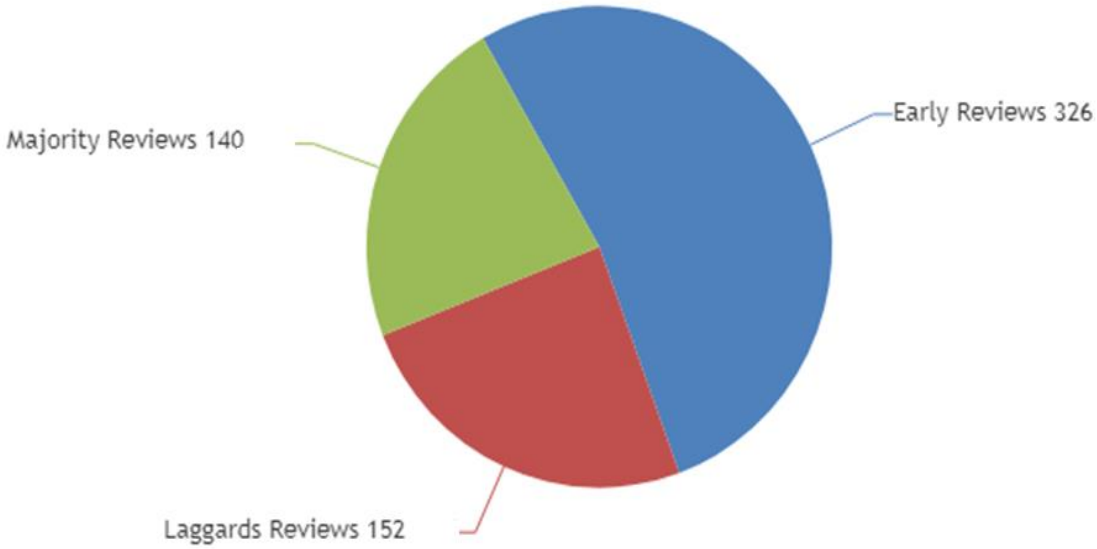


Figure 14: reviews classified into three types

Figure 13 represents three types of reviews which we have classified on the basis of product life time. Out of 607 reviews we can see that 326 are early reviews, 140 are majority reviews while the remaining 152 are laggard reviews.

The below table displays the average rating and average text letters per Review given by each category of reviews.

	Early	**Majority**	**Laggard**
Ave. Rating	3.97	3.56	3.42
Ave. Text per Review	69.2	51.9	32.4

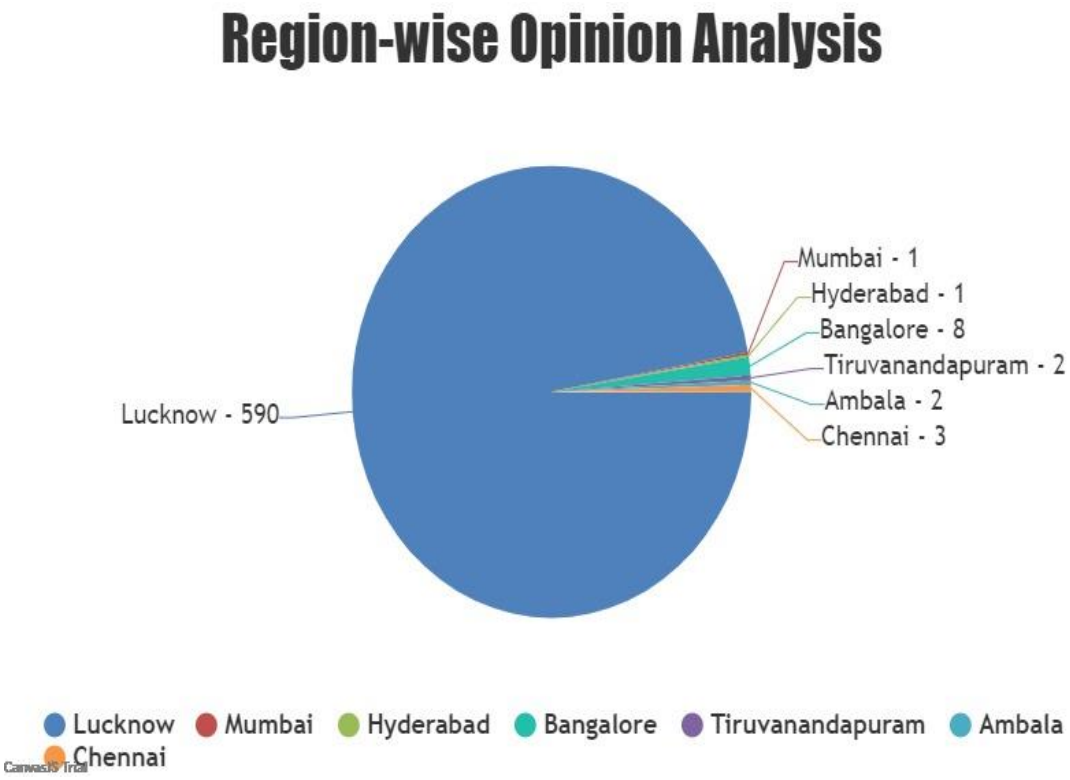


Figure 15: Region-wise opinion analysis

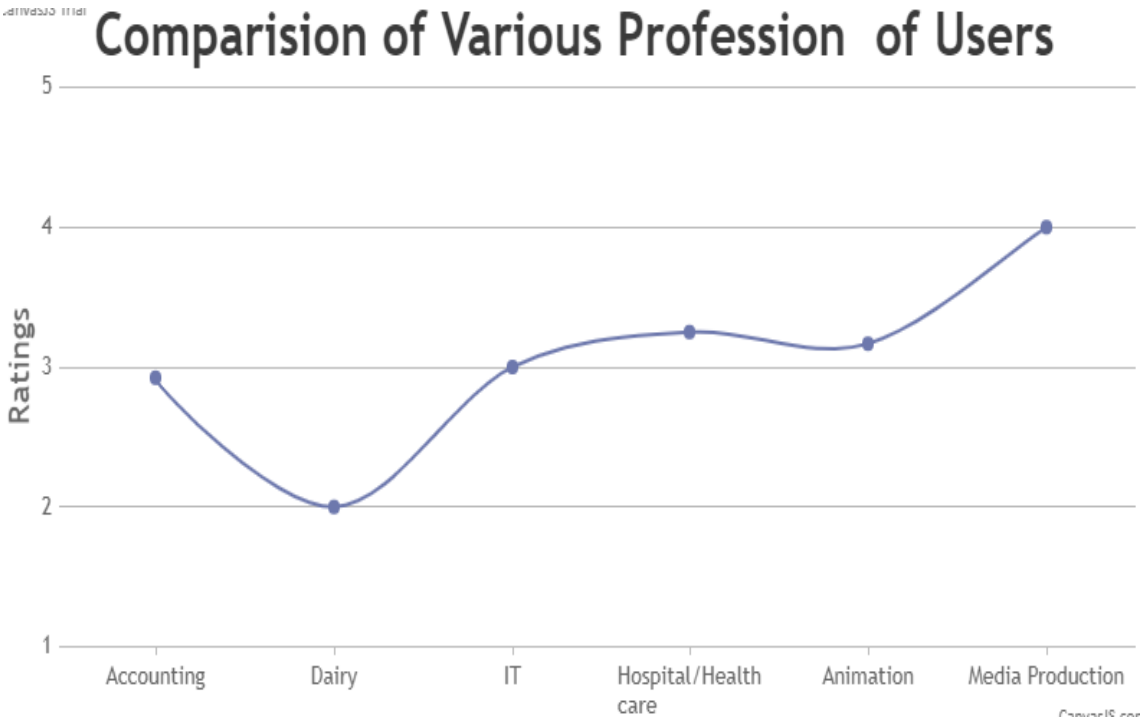


Figure 16: Comparison of various profession of users

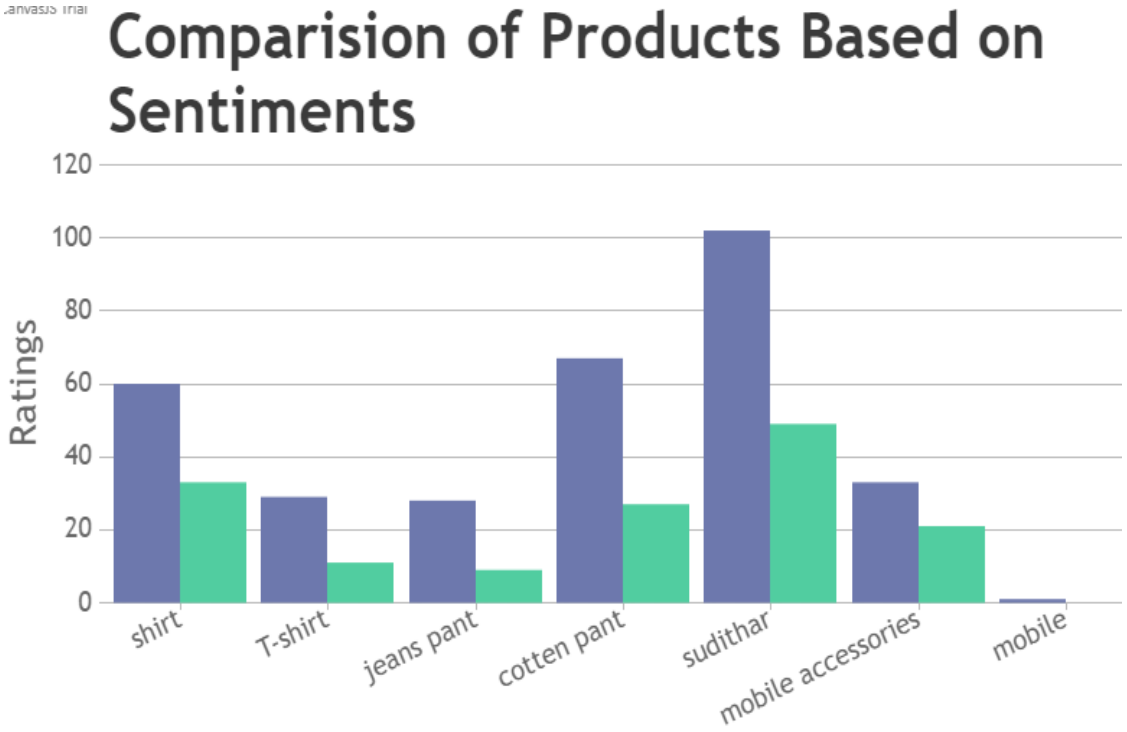


Figure 17: Comparison of product based on sentiments

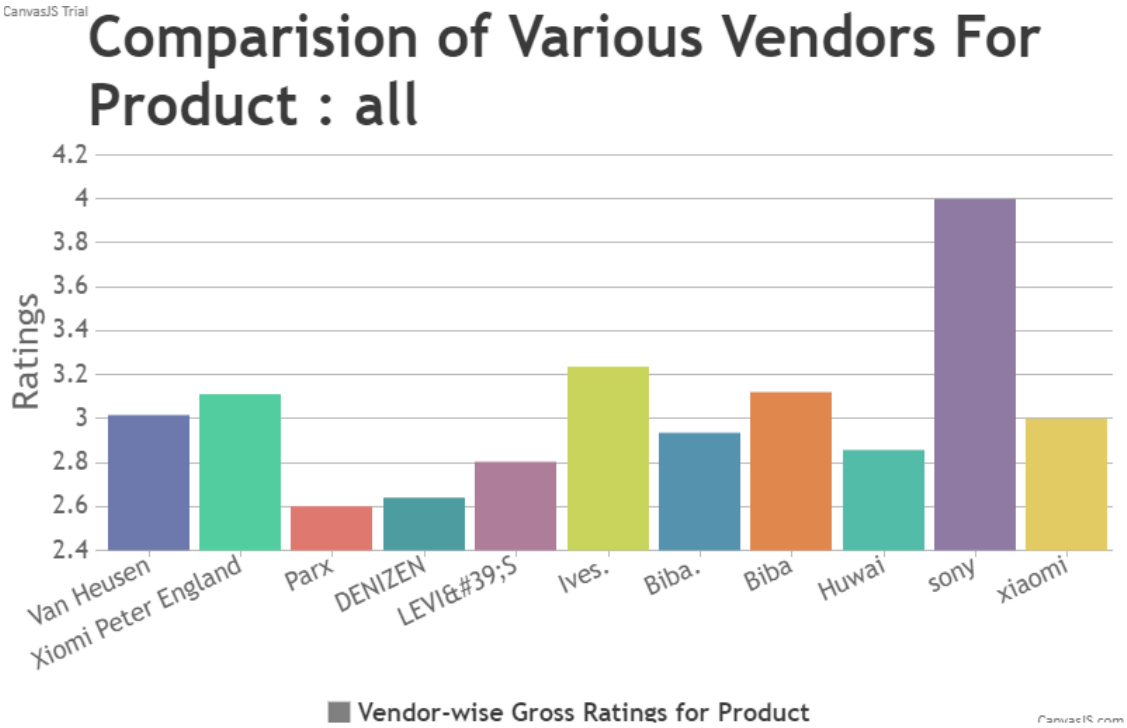


Figure 18: Comparison of various vendors for product

	Classifier	Number of Review	Accuracy (%)
1	NB (PROPOSED)	200	65%
2	SVM (PROPOSED)	200	87.6664
3	Random Forest	200	81%
4	RNN	200	85.42%
5	Linear Regression	200	78.60%
6	Cosine Similarity	200	73.23%

Table 2: Accuracy Comparison of Existing and Proposed System

The above table shows the accuracy comparision of existing and Proposed system. In this table we can see that Support Vector Machine has clearly outperformed other techniques.

CHAPTER – 5

CONCLUSION

AND

FUTURE SCOPE

6.1 CONCLUSION

In this document, we consider the new concepts described and expected in previous revisions. Our accurate inspection reinforces the advancement of hypothetical results from the social sciences and finance. We found that (1) early analysts generally assign higher normal ratings; (2) Previous reviewers generally issue more relaxed reviews. Sentiment analysis or opinion mining is a field of research that analyzes the emotions, attitudes or emotions of people towards certain entities. This project solves a basic problem in sentiment analysis, namely, the classification of the polarity of feelings. A detailed description of the emotion polarity classification process and each step has been proposed. Experiments have been conducted on sentence-level classification and comment-level classification.

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. Our research is aiming to achieve this by conducting sentiment analysis of product reviews and classifying the reviews into positive and negative sentiment.

6.2 FUTURE WORK

Currently, we focus on the analysis and prediction of reviews, early reviews, while there remains an important issue to address, i.e., how to improve product marketing with the identified early reviews. We will investigate this task with real e-commerce cases in collaboration with e-commerce companies in the future. In future, the work can be extended to perform multiclass classification of reviews which will provide delineated nature of review to the consumer, hence better judgement of the product. It can also be used to predict rating of a product from the review. This will provide users with reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. The proposed extension of work will be very beneficial for the e-commerce industry as it will augment user satisfaction and trust.

REFERENCES

[1] Najma Sultana , Pintu Kumar , Monika Rani Patra , Sourabh Chandra And S.K. Safikul Alam" Sentiment Analysis For Product Review" ICTACT Journal On Soft Computing, April 2019, Volume: 09, Issue: 03

[2] Alpna P. , Arvind K. T. , “Sentiment Analysis by using Recurrent Neural Network” proceedings of 2nd ICACSE, 2019

[3] Yoon-Joo Park "Envisioning the Helpfulness of Online Customer Reviews across Different Product Types" MDPI, Sustainability 2018, 10, 1735

[4] Sunil Saumya, Jyoti Prakash Singh, Nripendra P. Rana,Yogesh k. Dwivedi, Swansea University Bay Campus, Swansea "Situating Online Consumer Reviews" Article in Electronic Commerce Research and Applications, March 2018

[5] Liao, Shiyang, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, “CNN for situations understanding based on sentiment analysis of twitter data”, Procedia Computer Science, vol. 111, pp.376-381, 2017.

[6] Xing Fang* and Justin Zhan "Sentiment Analysis using product review data” Journal of Big Data , 2015

[7] D. Imamori and K. Tajima, "Foreseeing notoriety of twitter accounts through the revelation of link-propagating early adopters," in CoRR, 2015, p. 1512.

[8] Yeole V., P.V. Chavan, and M.C. Nikose, “Opinion mining for emotions determination”, ICIIECS 2015-2015 IEEE Int. Conf. Information, Embed. Commun. Syst., 2015.

[9] F. Luo, C. Li, and Z. Cao, Affective-feature-based sentiment analysis using SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support. Coop. Work Des., pp.276281, 2016.

[10] Kalaivani A., Thenmozhi D, Sentiment Analysis using Deep Learning Techniques. , IJRTE,2018

[11] Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. Scalable sentiment classification for big data analysis using naïve bayes classifier. In big data, 2013 IEEE International Conference on, pages 99-104. IEEE,2013

[12] D. N. Devi, C. K. Kumar, and S. Prasad, "A feature based approach for sentiment analysis by using support vector machine," in Advanced Computing (IACC), 2016 IEEE 6thInterna-tional Conference on. IEEE, 2016, pp. 3–8.

[13] McAuley, J.; Leskovec, J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In Proceedings of the 7th ACM Conference on Recommender Systems, RecSys', Hong Kong, China,12–16 October 2013; pp. 165–172.

[14] Ghose, A.; Ipeirotis, P.G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Trans. Knowl. Data Eng. 2011, 23, 1498–1512.

[15] V. G. D. W. Shih-Lun Tseng, Shuya Lu, "The effect of herding behavior on online review voting participation," in AMCIS, 2017.

[16] S. M. Mudambi and D. Schuff, "What makes a helpful online review? a study of customer reviews on amazon.com," in MIS Quarterly, 2010, pp. 185–200.

[17] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1–135, 2007.

[18] K. Dave, S. Lawrence, and D.M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web, pages 519_528. ACM, 2003. ISBN 1581136803.

[19] J.C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. AD- VANCES IN KNOWLEDGE ORGANIZATION, 9:49_54, 2004. ISSN 0938- 5495.

[20] Qiang Ye, Ziqiong Zhang, and Rob Law. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert systems with applications, 36(3):6527–6535, 2019.

[21] M. M. Altawaier & S. Tiun "Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis". International Journal on Advanced Science, Engineering and Information Technology, Vol. 6 (2016) No. 6, pages: 1067-1073.

[22]K. Dave, S. Lawrence and D. M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews". In Proceedings of the 12th international conference on World Wide Web, pp. 519-528.

[23]A. Balahur, R. Mihalcea and A. Montoyo "Computational approaches to subjectivity and sentiment analysis": Present and envisaged methods and applications. Computer Speech & Language, 28(1): 1-6, 2014.

[24]E. Boiy and M. F. Moens "A machine learning approach to sentiment analysis in multilingual Web texts". Information retrieval, 12(5): 526-558, 2009.

[25]M. Ghiassi, J. Skinner and D. Zimbra. "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network". Expert Systems with applications, 40(16): 6266-6282, 2013

[26] Tan LK-W, Na J-C, Theng Y-L, Chang K (2011) Sentence-level sentiment polarity classification using a linguistic
approach. In: Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation. Springer,
Heidelberg, Germany. pp 77–87

[27] Liu B (2012) Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan
& Claypool Publishers

[28] Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD
international conference on Knowledge discovery and data mining. ACM, New York, NY, USA. pp 168–177

[29] Gann W-JK, Day J, Zhou S (2014) Twitter analytics for insider trading fraud detection system. In: Proceedings of the
sencond ASE international conference on Big Data. ASE

[30] Roth D, Zelenko D (1998) Part of speech tagging using a network of linear separators. In: Coling-Acl, The 17th
International Conference on Computational Linguistics. pp 1136–1142

[31] Kristina T (2003) Stanford log-linear part-of-speech tagger. http://nlp.stanford.edu/software/tagger.shtml

[32] Marcus M (1996) Upenn part of speech tagger. http://www.cis.upenn.edu/~treebank/home.html

[33] Han J, Kamber M, Pei J (2006) Data Mining: Concepts and Techniques, Second Edition (The Morgan Kaufmann Series
in Data Management Systems), 2nd ed. Morgan Kaufmann, San Francisco, CA, USA

PLAGIARISM CHECK REPORT

Plagiarism Checker X Originality Report

Similarity Found: 6%

Date: Friday, August 06, 2021
Statistics: 482 words Plagiarized / 8399 Total words

PUBLICATION FROM THIS WORK

1) "**Characterizing and Predicting Reviews for Effective Product Marketing and Advancement** " has been accepted for oral presentation in the "International Conference on Artificial Intelligence (ICAI2021)" and publication in the Journal of Informatics, Electrical & Electronics Engineering (ISSN: 2582-7006).

2) **"Characterizing and Predicting Reviews for Effective Product Marketing and Advancement"** has been accepted and published by IJSRCSEIT Journal, Volume 7 issue 4 July-August
   UGC Journal No: 64718

PUBLICATIONS

Journal of Informatics Electrical and Electronics Engineering, 2021,
Vol. 02, Iss. 02, S. No. 006, pp. 1-4
ISSN (Online): 2582-7006

Characterizing and Predicting Reviews for Effective Product Marketing and Advancement

Aihsan Suhail[1], Halima Sadia[2]

[1]M.Tech, Scholar, Department of Computer Science & Engineering, Integral University, Lucknow, India
[2]Assistant Professor, Department of Computer Science & Engineering, Integral University, Lucknow, India

aihsansuhail@gmail.com[1], halima@iul.ac.in[2]

How to cite this paper: A. Suhail, H. Sadia (2021) Characterizing and Predicting Reviews for effective Product Marketing and Advancement. *Journal of Informatics Electrical and Electronics Engineering, Vol. 02, Iss. 02, S. No. 006, pp. 1-4, 2021*.

Received: 06/04/2021
Accepted: 26/05/2021
Published: 05/06/2021

Abstract

In the present made world, dependably, individuals around the planet grant through different stages on the Web. It has been addressed, about 71% of by and large online customers read online surveys going before buying a thing. Thing considers, particularly the early surveys (i.e., the investigations posted at the beginning time of a thing), astoundingly impact coming about thing deals. We call the clients who posted the early examinations as "early investigators". Be that as it may, early specialists contribute just a little level of surveys, their feelings can pick the achievement or disappointment of new things and associations. It is immense for relationship to perceive early spectators since their responses can assist relationship with changing publicizing frameworks and improve thing plans, which can at last incite the accomplishment of their new things. And in dependably, a mass extent of unstructured information is made. This information is as text, which is accumulated from get-togethers, online media regions, surveys. Such information is named as gigantic information. Client feelings are identified with a wide degree of spotlights like on express things also. These investigations can be mined utilizing different movements and are of everything considered significance to make checks since they unmistakably pass on the perspective of the bigger part. Online outlines moreover have become a basic wellspring of data for clients going before settling on an educated buy choice. Early examiner's appraisals and their got strength scores are apparently going to influence thing notoriety. The test is to assemble all the audits, in like way find and investigate the assessments, to locate something refined, that scores high evaluating.

Keywords

Online Review, Big data, Analysis, machine Learning

1. Introduction

The improvement of online business objections has connected with clients to disperse, or share buy encounters by posting thing outlines, which routinely contain huge closures, remarks and investigation towards a thing [8]. Thusly, a greater piece of clients will inspect online outlines prior to settling on an educated buy choice. It has been addressed, about 71% of by and large online customers read online audits prior to buying a thing. Thing contemplates, particularly the early audits (i.e., the surveys posted at the start time of a thing), astoundingly impact coming about thing deals [9], [10], [11]. We call the clients who posted the early surveys as "early specialists". However early specialists contribute just a little level of audits, their choices can pick the achievement or disappointment of new things and associations [12], [13-16]. It is colossal for relationship to see early inspectors since their responses can assist relationship with changing publicizing methods and improve thing plans, which can at long last incite the achievement of their new things. Thinking about the above conversations, we can see that early agents are essential for thing advancing. Hereafter, in this undertaking, we adventure up and study the immediate credits of early correspondents through their posted examinations on delegate electronic business stages. We desire to lead productive assessment and make exact supposition on early agents.

2. Literature Survey

Najma Sultana et al [2019] Sentiment research is defined as the way to extract data, opinions, surveys or sentences to predict the feeling of the sentence through common language handling (NLP). "Positive" "Negative" "Unbiased. It analyses the data and marks the 'better' and the 'more regrettable' supposition as sure and negative individually. Hence, in the previous years, the World Wide Web (WWW) has become an immense wellspring of crude information produced custom or client. Utilizing web-based media, online business site, films audits, for example, Facebook, twitter, Amazon, Flipkart and so on client share their perspectives, emotions in an advantageous way. Feeling examination is text-based investigation, however there are sure difficulties to locate the precise extremity of the sentence [1].

Alpna Patel [2019] E-business locations are being acclaimed and tremendous progress has been made in ads. In online business applications, customer devotion and persuasive fulfilment of customer needs are more relevant. To achieve effective results and contribution to online business applications, the necessity for suitably amassing customer reviews and analysis is required. The proposed work sufficiently crushes the issue of considering the customer lead from their online thing reviews. The experts are assembled into three classes explicitly unrefined, standard and present. This is sensible for both thing and expert portrayal. This portrayal is performed and associated among various investigators and their bit of leeway. The proposed work can in like manner choose the thing reputation and the thing use by different customers from web shopping destinations. In this paper to perform sentiment analysis, researcher has used IMDB movie review dataset and RNN as its machine learning technique [2].

Yoon-Joo Park et al [2018] Online customer studies are a prudent kind of casual (WOM) which accept an unyieldingly huge part in web business. Mediocre quality reviews will, in any case, consistently produce per users of trouble review. The inspiration driving this paper is to therefore foresee the convenience of

studies. The inspiration driving this paper is to therefore foresee the convenience of studies. The results show that reviews for different thing types have particular mental and phonetic characteristics and the factors impacting the study steadiness of them are also exceptional. Our disclosures in like manner show that the assistance vector backslide methodology predicts review uphold most decisively among the four strategies for all of the five datasets. This assessment adds to improving powerful use of online surveys [3].

Sunil Saumya1 et al [2018] In the hundreds and even in the massive numbers for some famous things, the item reviews are posted online. Dealing with an especially gigantic volume of continually delivered online substance is a troublesome task for buyers, sellers and even researchers. The purpose behind this assessment is to rank the amazing number of reviews using

their foreseen convenience score. The system is specially made the reviews into low or high type by self-assertive forest area classifier. The convenience score of the brilliant reviews is simply foreseen using point boosting regression. The help score of the sub-par quality overviews isn't resolved considering the way that they are never going to be in the top k reviews. They are essentially added at the completion of the review summary to the overview posting site. The proposed system gives sensible study circumstance on review posting pages and making all superb studies recognizable to customers on the top. The exploratory results on data from two famous [4].

Liao, Shiyang [2017] gave an approach to comprehend real situations with the Sentiment Analysis of a Twitter data centred on Deep learning techniques. With the suggested method, it was viable to forecast user satisfaction on a product. Lately, Deep Learning was competent to resolve problems in voice recognition or computerized vision. CNN worked fine for image analysis together with classification. An imperative reason to employ CNN for image analysis and image classification was that the CNN could extort an area of features as of global information precisely and also it was competent to regard the relations amongst those features. The above solution could attain the utmost accuracy in analysis together with classification. For NLP 'texts' data features could also be extorted piece by piece. Regarding the relations amongst those features without considering the context or complete sentence might incorrectly interpret the sentiment. It was the most effectual method to perform image classification. CNN comprised a convolutional layer to extort information by a large piece of text [5].

Xing Fang et al [2015] Sentiment analysis or evaluation mining is one of the main tasks of NLP (Natural Language Processing). Feeling research has given a lot of thinking of late. In this paper, we expect to deal with the issue of end furthest point order, which is one of the fundamental issues of idea examination. A general cycle for incline limit course of action is proposed with low down communication portrayals. Data used in this assessment are online thing overviews assembled from Amazon.com. Preliminaries for both sentence-level request and review level course of action are performed with promising outcomes. At last, we also give understanding into our future work on idea examination [6].

Daichi Imamori, Keishi Tajima [2015] gave way to deal with idea Due to the dynamicity, new notable records reliably appear and disappear in scaled down scale writing for a blog organization. Early distinguishing proof of new records that will wrap up standard in future is a fundamental issue that has a couple of utilizations, for instance, incline area, viral displaying, and customer proposal. Assessment of conspicuousness of a record is also significant for approximating the idea of information it posts. Assessment of the idea of information is imperative in various applications, yet it is generally difficult to measure it without human intervention. Similar idea has also been successfully associated with limited scope web diaries with interfacing limits [7].

3. Research Gap

One will certainly discover a lot of researchers relying on observation analysis and opinion examination, but unfortunately each of them leads to a simple graphical representation of knowledge that indicates either positive or negative of that object. While this undertaking is advertising-based, it helps to market the item

effectively by dissecting commentators and also improving the quality of the product and furthermore gives a client perspective on the item.

5. Conclusions

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. In this, technically advanced decade, the significance of decision making of market strategy depends highly on the analysis of

marketing surveys and product reviews. Thus, in this literature we have tried to compare and learn some of the best models of sentiment analysis. Maximum researchers have tried to find out the overall analysis of the reviews but hardly anyone used that analysis for product marketing and enhancement.

References

[1]. Najma Sultana, Pintu Kumar, Monika Rani Patra, Sourabh Chandra and S.K. Safikul Alam" Sentiment Analysis For Product Review"
ICTACT Journal On Soft Computing, April 2019, Volume: 09, Issue: 03

[2]. Alpna P., Arvind K. T., "Sentiment Analysis by using Recurrent Neural Network" proceedings of 2nd ICACSE, 2019

[3]. Yoon-Joo Park "Envisioning the Helpfulness of Online Customer Reviews across Different Product Types" MDPI, Sustainability 2018, 10, 1735

[4]. Sunil Saumya, Jyoti Prakash Singh, Nripendra P. Rana,Yogesh k. Dwivedi, Swansea University Bay Campus, Swansea "Situating Online Consumer Reviews" Article in Electronic Commerce Research and Applications • March 2018

[5]. Liao, Shiyang, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, "CNN for situations understanding based on sentiment analysis of twitter data", Procedia Computer Science, vol. 111, pp.376-381, 2017.

[6]. Xing Fang* and Justin Zhan "Sentiment Analysis using product review data", Journal of Big Data, 2015

[7]. D. Imamori and K. Tajima, "Foreseeing notoriety of twitter accounts through the revelation of link-propagating early adopters," in CoRR, 2015, p. 1512.

[8]. Yeole V., P.V. Chavan, and M.C. Nikose, "Opinion mining for emotions determination", ICIIECS 2015-2015 IEEE Int. Conf. Information, Embed. Commun. Syst., 2015.

[9]. 9. F. Luo, C. Li, and Z. Cao, Affective-feature-based sentiment analysis using SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support.
Coop. Work Des., pp.276281, 2016.

[10]. Kalaivani A., Thenmozhi D, Sentiment Analysis using Deep Learning Techniques. , IJRTE,2018

[11]. Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. Scalable sentiment classification for big data analysis using naïve bayes classifier. In big data, 2013 IEEE International Conference on, pages 99-104. IEEE,2013

[12]. D. N. Devi, C. K. Kumar, and S. Prasad, "A feature-based approach for sentiment analysis by using support vector machine," in
Advanced Computing (IACC), 2016 IEEE 6thInterna-tional Conference on. IEEE, 2016, pp. 3–8.

[13]. McAuley, J.; Leskovec, J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In Proceedings of the
7th ACM Conference on Recommender Systems, RecSys', Hong Kong, China,12–16 October 2013; pp. 165–172.

[14]. Ghose, A.; Ipeirotis, P.G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Trans. Knowl. Data Eng. 2011, 23, 1498–1512.

[15]. Singhal, P., Sharma, P., & Rizvi, S. (2019). Thwarting Sybil Attack by CAM Method in WSN using Cooja Simulator Framework. *International Journal of Engineering & Technology*, *8*(1.5), 116-125.

[16]. M. Vinny, P. Singh (2020) Review on the Artificial Brain Technology: BlueBrain. Journal of Informatics Electrical and Electronics Engineering,1(1), 3, 1-11


IARAI
Promoting Safe AI
International Conference on
Artificial Intelligence (ICAI2021)
May 22-23, 2021
AI Foundation
(www.aifoundation.in)
Participation Certificate
This certificate is presented to
AIHSAN SUHAIL
Integral University
for presenting the Paper with Title
Characterizing And Predicting Reviews For Effective Product Marketing And Advancement
in ICAI2021 on May 22-23, 2021.
Dr. Avimanyou Vatsa
Fairleigh Dickinson University, USA
General Chair, ICAI 2021
Certificate ID: ICAI22230521006
Dr. Agostini Alessandro
INHA University, South Korea
General Chair, ICAI 2021

Characterizing and Predicting Reviews for Effective Product Marketing and Advancement

Aihsan Suhail[1], Halima Sadia[2] ,

[1] MTech, Scholar, Department of Computer Science & Engineering, Integral University, Lucknow, India

[2] *Associate Professor*, Department of Computer Science & Engineering, Integral University, Lucknow, India

[1]aihsansuhail@gmail.com, [2] Halima@iul.ac.in

Abstract. Online surveys have become a significant wellspring of data for clients prior to settling on an educated buy choice. Early audits of an item will in general exceptionally affect the ensuing item deals. In this paper, we step up and study the conduct qualities of early reviewer through their posted audits on our shopping gateway. In explicit, we partition item lifetime into three back to back stages, in particular early, lion's share. A client who has posted a survey in the beginning phase is considered as an early analyst. We quantitatively describe early reviewer dependent on their rating practices, the supportiveness scores got from others and the relationship of their surveys with item prevalence. We have tracked down that (1) an early analyst will in general relegate a higher normal rating score; and (2) an early reviewer will in general post more supportive audits. Our examination of item surveys additionally demonstrates that early reviewers appraisals and their got support scores are probably going to impact item prominence. By survey audit posting measure as a multiplayer rivalry game, we propose a novel edge based implanting model for early analyst forecast. Broad investigations on two diverse web based business datasets have shown that our proposed approach beats various cutthroat baselines.

Keywords: Online surveys, early analyst, rating, audits, reviewer.

1 INTRODUCTION

The development of web based business sites has empowered clients to distribute or share buy encounters by posting item audits, which for the most part contain helpful conclusions, remarks and criticism towards an item. In that capacity, a lion's share of clients will peruse online surveys prior to settling on an educated buy choice. It has been accounted for about 71% of worldwide online customers read online surveys prior to buying an item. Item surveys, particularly the early audits (i.e., the surveys posted in the beginning phase of an item), profoundly affect ensuing item deals. We call the clients who posted the early surveys early commentators. Albeit early analysts contribute just a little extent of audits, their sentiments can decide the achievement or disappointment of new items and administrations. It is significant for organizations to distinguish early commentators since their criticisms can assist organizations with changing promoting techniques and improve item plans, which can in the long run lead to the achievement of their new items. Thus, early commentators become the accentuation to screen and draw in at the early advancement phase of an organization. The essential job of early audits has drawn in broad consideration from promoting experts to actuate shopper buy expectations. For instance, Amazon, one of the biggest online business organization on the planet, has supported the Early Reviewer Program1, which assists with getting early surveys on items that have not many or no audits. With this program, Amazon customers can study items and settle on more brilliant purchasing choices. As another connected program, Amazon Vine2 welcomes the most confided in analysts on Amazon to post suppositions about new and pre-release things to help their kindred clients settle on educated buy choices.

Past examinations have profoundly stressed the wonder that people are emphatically impacted by the choices of others, which can be clarified by group conduct. The impact of early audits on ensuing buy can be perceived as a unique instance of crowding impact. Early audits contain significant item assessments from past adopters, which are important reference assets for resulting buy choices. As demonstrated in, when buyers utilize the item assessments of others to gauge item quality on the Internet, crowd conduct happens in the web based shopping measure. Not the same as existing examinations on crowd conduct, we centre around quantitatively investigating the general qualities of early commentators utilizing huge scope genuine world datasets. Moreover, we formalize the early analyst expectation task as a contest issue and propose a novel installing based positioning way to deal with this undertaking. As far as anyone is concerned, the assignment of early commentator forecast itself has gotten almost no consideration in the writing. Our commitments are summed up as follows:

We present a first report to portray early analysts on a web based business site utilizing two certifiable huge datasets. We quantitatively dissect the attributes of early analysts and their effect on item prevalence. Our exact examination offers help to a progression of hypothetical ends from the social science and financial aspects. We see survey posting measure as a multiplayer contest game and foster an installing based positioning model

for the forecast of early analysts. Our model can manage the chilly beginning issue by consolidating side data of items. Broad examinations on two true enormous datasets, i.e., Amazon and Yelp have exhibited the viability of our methodology for the forecast of early commentators.

In this paper section I contains the introduction, section II contains the literature review details, section III contains the details about methodologies, section IV shows architecture details, V describe the result and section VII provide conclusion of this paper.

2 RELATED WORK

A developmental shift from disconnected business sectors to advanced business sectors has expanded the reliance of clients on online audits generally. Online surveys have become a stage for building trust and affecting purchaser purchasing behaviours. With such reliance there is a need to deal with such enormous volume of surveys and present believable audits before the customer. In this, actually progressed decade, the meaning of dynamic of market system relies exceptionally upon the investigation of advertising studies and item surveys. Hence, in this writing we have attempted to think about and get familiar with the absolute best models of assessment investigation. Greatest analysts have attempted to discover the general investigation of the surveys yet scarcely anybody utilized that examination for item advertising and improvement.

YEAR	AUTHOR	PURPOSE	TECHNIQUE
2019	N. Sultana, P. kumar	Sentiment analysis for product review	NB,SVM & Linear model algorithm
2019	Alpna Patel & Arvind Kumar	Sentiment analysis by using RNN	Used RNN
2018	Yoon-Joo Park	Online review helpfulness across different product type	Linear regression, SVM, Random Forest,M5P
2018	S. Saumya, J. Prakash	Ranking Online customer reviews	Cosine similarity, SVM and Random Forest
2017	Liao Shiyang, Junbo Wang.	SA of reviews data	NLP and CNN for classification
2015	D. Imamori & K Tajima	Predicting popularity of reviews account.	Cosine similarity and SVM
2015	Xing Fang & Justin Zhan	Sentiment analysis using product review data	NB, SVM, Random Forest

3 METHODOLOGY

- **Preprocessing**

  In this algorithm, the review which are foreign made to database from the reviews API, these reviews comprise of senseless words, voids, hyperlink and unique character. First we have to do separation by removing every single superfluous word, whitespace, hyperlink and extraordinary characters.

  The pre-processing step aim to start the “feature extraction” process and begin extracting “bags of words” from the samples. “One of the main focus is to reduce the final amount of features extracted”. Indeed, features reduction is vital so as to enhance the accuracy of the prediction for both topic modelling and sentiment analysis. Features are wont to represent the samples, and therefore the more the algorithm are going to be trained for a selected feature the more accurate the results are going to be . “Hence, if two features are similar it is convenient to combine them as one unique feature”. Moreover, if a feature isn't relevant for the analysis, it are often faraway from the bag of words.

- Lower uppercase letters: the last word step within the preprocessing is to travel through all the info and alter every uppercase letter to their corresponding lowercase letter. When processing a word, the analysis are going to be case”” sensitive and therefore the program will consider “data” and “Data” as two totally different words. it's important that, these two words are considered because the same features. Otherwise, the algorithms will affect sentiments which can differ to those two words. for instance , on these three sentences: “data are good”, “Awesome data”, and “Bad Data”. the primary and second sentences both contain “data” and are positive, the third sentence contains “Data” and is negative. The algorithm will guess that sentences containing “data” are more likely to be positive and people containing “Data” negative. If the uppercases had been removed the algorithm would are ready to guess that the very fact that the sentence contains “data” isn't very relevant to detect whether or the sentence is positive. This pre-processing step is even more important since the info are retrieved from Reviews. Social media users are often

writing in uppercase albeit it's not required, thus this pre-processing step will have a far better impact on social media data than other “classical” data.
• Remove URLs and user references: Reviews allows user to incorporate hashtags, user references and URLs in their messages. In most cases, user references and URLs aren't relevant for analyzing the content of a text. Therefore, this pre-processing step relies on regular expression to seek out and replace every URLs by “URL” and user reference by “AT_USER”, this enables to scale back the entire amount of features extracted from the corpus [2]. The hashtags aren't removed since they often contain a word which has relevancy for the analysis, and therefore the “#” characters are going to be removed during the tokenization process. • Remove digits: Digits aren't relevant for analysing the info , in order that they are often faraway from the sentences. Furthermore, in some cases digits are going to be mixed with words, removing them may allow to associate two features which can are considered different by the algorithm otherwise. for instance , some data may contain “iphone”, when other will contain “iphone7”. The tokenization process, which can be introduced later.
• Remove stop words: In tongue processing, stop words are often faraway from the sample. These stop words are words which are commonly utilized in a language, and aren't relevant for several tongue processing methods like topic modeling and sentiment analysis [10]. Removing these words allows to scale back the quantity of features extracted from the samples.
• Self-Learning and word standardization System
In this algorithm, first we've to instate the word reference (first emphasis dictionary).In the lexicon for the foremost part we've to introduce the positive, negative nonpartisan and things. Every single huge datum and knowledge mining ventures in sight of the prepared information, without prepared information (introduction of words).So instatement of the prepared information is significant . within the self-learning framework, we do word institutionalization, here we aren't considering past, present and future status of the words, just we are brooding about the word.

- **Sentiment Analysis**

In this calculation, pre-processed review are brought from the data set individually. In any case we require check individually watchword whether that expression is thing are not, if thing we will oust it from the particular audit. After that the remainder of the watchwords checked with evaluation create, whether or not those expressions are sure assessment or adverse end or unbiased inclination. The remainder of the watchwords in the tweet which doesn't has a spot with any of the assumption will be consigned fleeting end considering the more check of positive, negative and fair. In the subsequent cycle if the reaming word crosses the restriction of positive, negative or impartial, that watchword everlastingly included as improvement in the vocabulary.

Algorithm Step in Sentiment Analysis

Step1:Get-some-sentiment-examples
As for every supervised learning problem, the algorithm needs to be trained from labeled examples in order to generalize to new data.

Step2:Extract-features-from-examples
Transform each example into a feature vector. The simplest way to do it is to have a vector where each dimension represents the frequency of a given word in the document.

Step3:Train-the-parameters
This is where your model will learn from the data. There are multiple ways of using features to generate an output, but one of the simplest algorithms is logistic regression. Other well-known algorithms are Naive Bayes. In the simplest form, each feature will be associated with a weight. Let's say the word "love" has a weight equal to +4, "hate" is -10, "the" is 0 ... For a given example, the weights corresponding to the features will be summed, and it will be considered "positive" if the total is > 0, "negative" otherwise. Our model will then try to find the optimal set of weights to maximize the number of examples in our data that are predicted correctly.
If you have more than 2 output classes, for example if you want to classify between "positive", "neutral" and "negative", each feature will have as many weights as there are classes, and the class with the highest weighted feature sum wins.

Step4:Test-the-model
After we have trained the parameters to fit the training data, we have to make sure our model generalizes to new data, because it's really easy to over fit. The general way of regularizing the model is to prevent parameters from having extreme values.

- **Naive Bayes**
- Bayes classifiers are a group of straightforward probabilistic classifiers dependent on applying Bayes' hypothesis with solid (gullible) freedom presumptions between the highlights.
- Naive Bayes classifiers are exceptionally adaptable, requiring various boundaries direct in the quantity of factors (highlights/indicators) in a learning issue. Greatest probability preparing should be possible by assessing a shut structure articulation, which takes

direct time, as opposed to by costly iterative guess as utilized for some different sorts of classifiers.

- In the insights and software engineering writing, credulous Bayes models are known under an assortment of names, including basic Bayes and autonomy Bayes. Every one of these names reference the utilization of Bayes' hypothesis in the classifier's choice guideline, yet innocent Bayes isn't (really) a Bayesian technique

- Naive Bayes is a straightforward method for building classifiers: models that appoint class marks to issue cases, addressed as vectors of highlight esteems, where the class names are drawn from some limited set. It's anything but a solitary calculation for preparing such classifiers, yet a group of calculations dependent on a typical guideline: all gullible Bayes classifiers expect that the worth of a specific element is free of the worth of some other component, given the class variable. For instance, an organic product might be viewed as an apple on the off chance that it is red, round, and around 10 cm in breadth. An innocent Bayes classifier thinks about every one of these highlights to contribute autonomously to the likelihood that this organic product is an apple, paying little heed to any potential connections between the shading, roundness, and breadth highlights.

- For a few sorts of likelihood models, credulous Bayes classifiers can be prepared productively in a managed getting the hang of setting. In numerous down to earth applications, boundary assessment for innocent Bayes models utilizes the strategy for most extreme probability; all in all, one can work with the credulous Bayes model without tolerating Bayesian likelihood or utilizing any Bayesian strategies.

- Despite their guileless plan and obviously distorted suspicions, credulous Bayes classifiers have functioned admirably in numerous intricate genuine circumstances. In 2004, an examination of the Bayesian order issue showed that there are sound hypothetical purposes behind the clearly doubtful viability of credulous Bayes classifiers. In any case, a complete examination with other order calculations in 2006 showed that Bayes arrangement is outflanked by different methodologies, like supported trees or irregular woods.

SYSTEM ARCHITECTURE

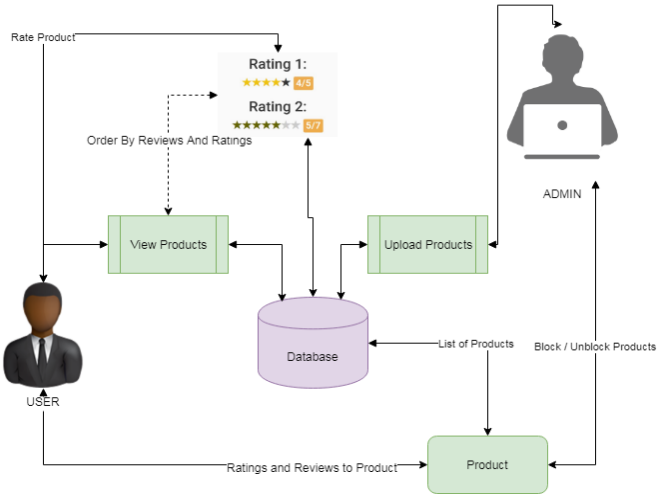


Figure 1 System Architecture

4 EXPERIMENTAL RESULTS AND ANALYSIS

In this outcome part, we step up and study the conduct qualities of posted audits on delegate online business stages. We plan to lead powerful investigation and make precise forecast towards item improvement. With the blasting of internet business, individuals are becoming accustomed to burning-through on the web and composing remarks about their buy encounters on vendor/audit Websites. These stubborn substances are important assets both to future clients for dynamic and to shippers for improving their items as well as administration.
These are the modules implemented in this research paper result part:

- **UPLOAD PRODUCTS**

  Uploading the products is done by admin. Authorized person is uploading the new arrivals to system that are listed to users. “Product” can be uploaded with its attributes such as brand, colour,

and all other details of warranty. The "uploaded products" are able to block or unblock by users.

- **PRODUCT REVIEW BASED ORDER**

The suggestion to user's view of products is listed based on the review by user and rating to particular item. 'Naïve bayes" algorithm is used in this project to develop the whether the sentiment of given review is positive or negative. Based on the output of algorithm suggestion to users is given. The algorithm is applied and lists the products in user side based on the positive and negative.

- **RATINGS AND REVIEWS**

Ratings and reviews are main concept of the project in order to find effective product marketing. The main aim of the project is to get the user reviews based on how they purchased or whether they purchased or not. The major find out of the project is when they give the ratings and how effective it is. And this will helpful for the users who are willing to buy the same kind of product.

- **DATA ANALYSIS**

The main part of the project is to analysis the ratings and reviews that are given by the user. The products can be analysis based on the numbers which are given by user. The user data analysis of the data can be done by charts format. The graphs may vary like pie chart, bar chart or some other charts.

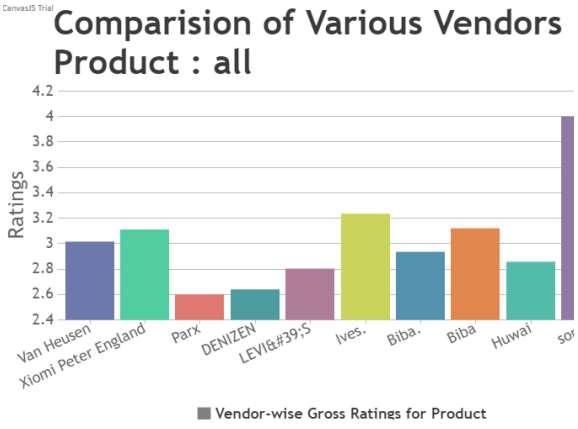


Figure 2: Comparison of various vendors for product

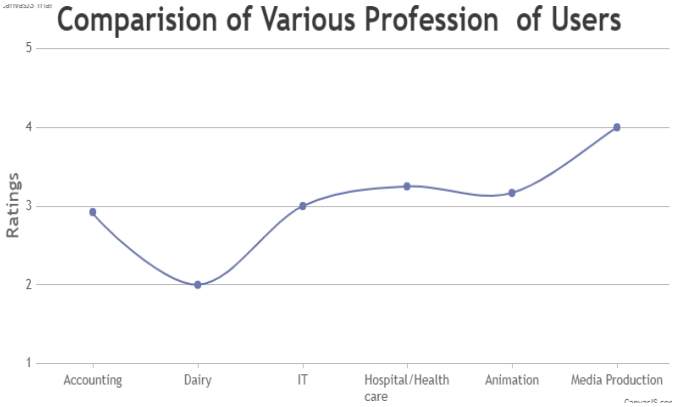


Figure 3: Comparison of various profession of users

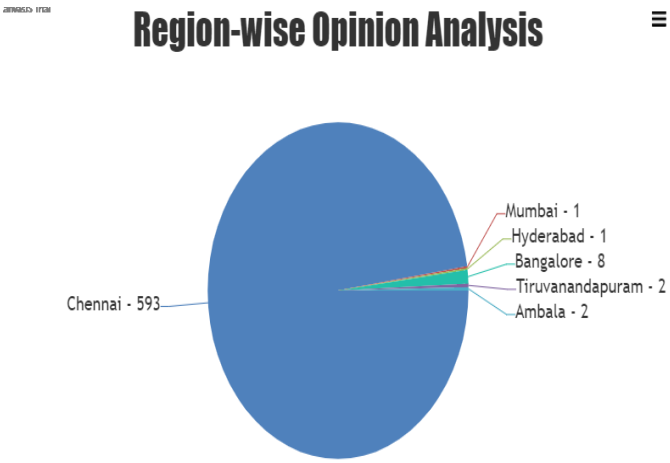


Figure 4: Region-wise opinion analysis

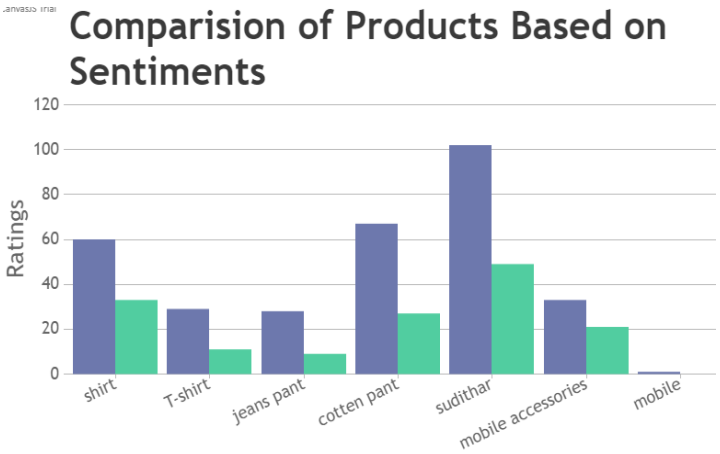


Figure 5: Comparison of product based on sentiments

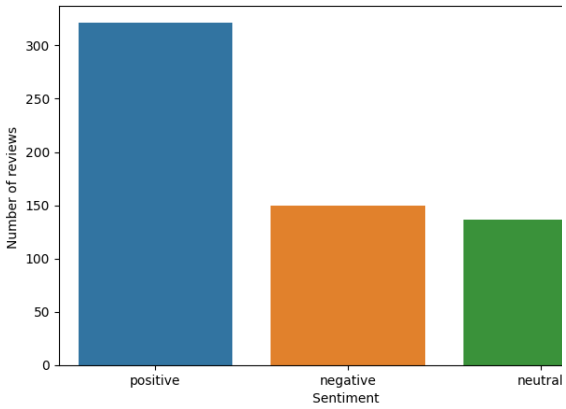


Figure 6: Merm analysis

Figure 8: shows the characteristics of early reviews

5 CONCLUSION

In this research paper, we have considered the novel concept of early reviewer portrayal and expectation on two certifiable online audit datasets. Our exact examination reinforces a progression of hypothetical ends from social science and financial aspects. We tracked down that (1) an early analyst will in general appoint a higher normal rating score; and (2) an early reviewer will in general post more accommodating audits. Our tests likewise show that early reviewers' evaluations and their got supportiveness scores are probably going to impact item ubiquity at a later stage. We have embraced a rivalry based perspective to display the survey posting measure, and fostered an edge based implanting positioning model (MERM) for foreseeing early reviewers in a beginning setting.

6 REFERENCES:

[1] Najma Sultana , Pintu Kumar , Monika Rani Patra , Sourabh Chandra And S.K. Safikul Alam" Sentiment Analysis For Product Review" ICTACT Journal On Soft Computing, April 2019, Volume: 09, Issue: 03

[2] Alpna P. , Arvind K. T. , “Sentiment Analysis by using Recurrent Neural Network” proceedings of 2nd ICACSE, 2019

[3] Yoon-Joo Park "Envisioning the Helpfulness of Online Customer Reviews across Different Product Types" MDPI, Sustainability 2018, 10, 1735

[4] Sunil Saumya, Jyoti Prakash Singh, Nripendra P. Rana,Yogesh k. Dwivedi, Swansea University Bay Campus, Swansea "Situating Online Consumer Reviews" Article in Electronic Commerce Research and Applications, March 2018

[5] Liao, Shiyang, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, “CNN for situations understanding based on sentiment analysis of reviews data”, Procedia Computer Science, vol. 111, pp.376-381, 2017.

[6] Xing Fang* and Justin Zhan "Sentiment Analysis using product review data” Journal of Big Data , 2015

[7] D. Imamori and K. Tajima, "Foreseeing notoriety of reviews accounts through the revelation of link-propagating early adopters," in CoRR, 2015, p. 1512.

[8] Yeole V., P.V. Chavan, and M.C. Nikose, “Opinion mining for emotions determination”, ICIIECS 2015-2015 IEEE Int. Conf. Information, Embed. Commun. Syst., 2015.

[9] F. Luo, C. Li, and Z. Cao, Affective-feature-based sentiment analysis using SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support. Coop. Work Des., pp.276281, 2016.

[10] Kalaivani A., Thenmozhi D, Sentiment Analysis using Deep Learning Techniques. , IJRTE,2018

[11] Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. Scalable sentiment classification for big data analysis using naïve bayes classifier. In big data, 2013 IEEE International Conference on, pages 99-104. IEEE,2013

[12] D. N. Devi, C. K. Kumar, and S. Prasad, “A feature based approach for sentiment analysis by using support vector machine,” in Advanced Computing (IACC), 2016 IEEE 6thInterna-tional Conference on. IEEE, 2016, pp. 3–8.

[13] McAuley, J.; Leskovec, J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In Proceedings of the 7th ACM Conference on Recommender Systems, RecSys’, Hong Kong, China,12–16 October 2013; pp. 165–172.

[14] Ghose, A.; Ipeirotis, P.G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Trans. Knowl. Data Eng. 2011, 23, 1498–1512.

Cite this article as :

Aihsan Suhail, Halima Sadia, "Characterizing and Predicting Reviews for Effective Product Marketing and Advancement", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. , July-August 2021. Available at doi : https://doi.org/10.32628/CSEIT2174107 Journal URL : https://ijsrcseit.com/CSEIT217407

Scientific Journal Impact Factor = 6.135
Online ISSN : 2456-3307
UGC Journal No : 64718

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 4/7534 28-Jul-2021

This is to certify that **Aihsan Suhail** has published a research paper entitled 'Characterizing and Predicting Reviews for Effective Product Marketing and Advancement' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 4, July-August 2021.

This Paper can be downloaded from the following IJSRCSEIT website link
https://ijsrcseit.com/CSEIT2174107

DOI : https://doi.org/10.32628/CSEIT2174107

IJSRCSEIT Team wishes all the best for bright future

Editor in Chief
IJSRCSEIT
website : http://ijsrcseit.com

Peer Reviewed and Refereed International Journal