International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 04 | Apr 2023 www.irjet.net p-ISSN: 2395-0072

E-Commerce Product Rating Based on Customer Review

Chikkela Sahithi[1], Bollu Swapna[1], Mulge Harsha[1], Sreedhar Bhukya[2], Bejjam Vaundhara devi[3]

[1]*Computer Science & Engineering, Sreenidhi Institute of Science & Technology, Hyderabad, Telangana, India*
[2]*Professor,Computer Science & Engineering, Sreenidhi Institute of Science & Technology, Hyderabad, Telangana, India*
[3]*Assistant Professor, Computer Science & Engineering, Sreenidhi Institute of Science & Technology, Hyderabad, Telangana, India*

Abstract — *Many users purchase products on EC websites. Due to online shopping, many e-commerce companies have been unable to know whether their customers are satisfied with their services. You can track and manage customer reviews by using algorithms to filter out theme and sentiment bias from online customer reviews. The system allows users to see different products and buy products online. Customers submit reviews of products and online shopping services. Specific keywords mentioned in customer reviews are mined and matched against keywords already in the database based on comparison, and the system evaluates the products and services offered by the company. This system uses a text mining algorithm to mine keywords. The system reviews various users and based on the reviews the system identifies whether the products and services offered by the e-commerce company are good, bad or worst. The system is a web application that allows users to browse various products online, purchase products, and rate products and online shopping services. This system helps many e-commerce companies improve or maintain their services based on customer ratings, and improve their products based on customer ratings.*

Key Words: **TF-IDF, Sentimental analysis, Tokenization, Lemmatization, Stemming.**

1.INTRODUCTION

In today's world, the web has become a great way to expressopinions about products. Your opinion matters a lot, especiallywhen it comes to making decisions about money and time. In these situations, people rely on opinions such as reviews. Likehis Facebook, Twitter, etc. on social media, people can discuss their opinions like a product of this research. That's also how we do it. Many people buy products online, but always check reviews and ratings before making a purchase online. This survey helps people save time and have a quick product discussion. Use these people to add sentiment keywords like good, bad, worst, best. Please judge whether this product is good or bad. Sentiment analysis, also known as opinion mining, is a branch of computer research that analyses how people express themselves in written keywords. People can express their opinions through sentiment keywords. It is virtually impossible to read all feedback to form conclusions and judgments. Or, in some cases, there may be conflicting reviews and ratings. A 5 star product may have verybad feedback. The reason for this may be that some people like to review products, others like to review them. This cannot mandate users to perform both tasks. To make unanimous decisions, I came up with a model that could set things right. It has a very simple but complex working method.

2. OBJECTIVES

The growing popularity of online reviews has also fuelled the fake review writing business. This refers to paid human writers creating misleading reviews to sway reader's opinions. Our project addresses this problem by building a classifier that takes as input the rating text and basic information from the giver of rating and returns whether the rating is trustworthy or not. This makes it difficult for potential customers to read them and decide whether to purchase the product. Product manufacturers also have problems maintaining overviews and managing customer opinions. Also, many other retailers on his website sell the same product with good reviews, and the manufacturer usually makes many kinds of products, so the manufacturer faces further difficulties.

3. REVIEW OF RELATED LITERATURE

Recently, many classification algorithms have been proposed, but SVM is still one of the most widely and most popular used classifiers. Applying the kernel equations arranges the data instances in such a way within the multi-dimensional space, that there is a hyper-plane that separates data instances of one kind from those of another. The kernel equations may be any function that transforms the linearly non-separable data in one domain into another domain where the instances become linearly separable. Kernel equations may be linear, quadratic, Gaussian, or anything else that achieves this particular purpose.Once we manage to divide the data into two distinct categories, our aim is to get the best hyper-plane to separate the two types of instances. The data instances that were not linearly separable in the original domain have become linearly separable in the new domain, due to the application of a function (kernel) that transforms the position of the data

points from one domain to another. This is the basic idea behind Support Vector Machines and their kernel techniques. Whenever a new instance is encountered in the original domain, the same kernel function isapplied to this instance too, and its position in the new domain is found out.

4. METHODS AND RESULTS

4.1 Proposed Method:

This system helps ensure that reviews can be done by authorized individuals. Also, don't show fake reviews. Based on the data, a word score is calculated and based on this score the user is presented with a quick analysis of product quality. Our motivation is to generate feature-related ratings for products. Product comments by users help create an overall rating. In our proposed work we develop a process of product aspect ranking consisting of three main Steps: (a) aspect identification; (b) sentiment classification on aspects (c) Product aspect ranking. Buyer ratings of products are used to first identify aspects of the ratings and then analyse those ratings to find buyers' opinions on the aspects via sentiment classifiers. Finally, the aspect frequencies and buyer ratings are used to rank the products based on the importance of the aspects and consider the ideas about every aspect of their overall opinion.

4.2 Methodology

Many people want a product review before paying to actually buy it. In some cases, products receive an overwhelming amount of customer feedback. It's very difficult to read all the reviews in order to reach consensus and make a decision. Our model analyzes user ratings and ranks products accordingly. This framework helps many e-commerce businesses develop or maintain their products based on consumer feedback, and improve their products based on customer reviews.

4.2.1 Data Gathering

We downloaded the dataset from Kaggle and are training our model to work upon all the reviews given on that particular dataset.

4.2.2 Data Cleaning

i)Lemmatization:

Lemmatization takes into account the morphological interpretation of the word. This requires a detailed dictionary that the algorithm can search to reconnect the form to the lemma. Reduce inflected words to ensure that the root word belongs to the language. The root of a lemma is called a lemma. A lemma is in the form of a collection of words in standard, dictionary, or citation form.

ii) Removing noisy data:

This process involves removing unnecessary and useless data.For example, if the input given is "The product works great". The output after the removal of unnecessary data will be "product, works, great".

iii)Removing Null Data :

For removing null data isnull() function is used, isnull() function detect missing values in the given series object and then reset_index() method sets a list of integer ranging from0 to length of data as index.

4.2.3 Data Shaping

i) Tokenization :

For tokenization, we used the punkt. Now with the help of tokenization, we divide the sentence into smaller parts i.e tokens. The role of it is to break the sentence into a lots of text which helps in building the model.

ii)Creating Corpus :

We then created a corpus list in which we added all the words broken down by the tokenizer. Now, apart from the alphabets in lower case (a-z) and the alphabets in upper case(A-Z), we removed everything from the text including numbers, symbols etc. Now we turned all the upper case texts into lowercase, tokenized them , lemmatized and sent it back into the corpus list.

iii)Indexing :

For indexing we used fit_on_text method ,this creates the vocabulary index based on word frequency. If you give it something like, "The rat was on the table." It will create a dictionary word_index["the"] = 1; word_index["table"] = 6 itis word -> index dictionary so every word gets a different integer value. 0 is reserved for padding. So lower integer means more frequent word.

The following are the stages of preprocessing: We perform four steps in Pre-processing :

1.Remove Punctuations ,special symbols, special characters.

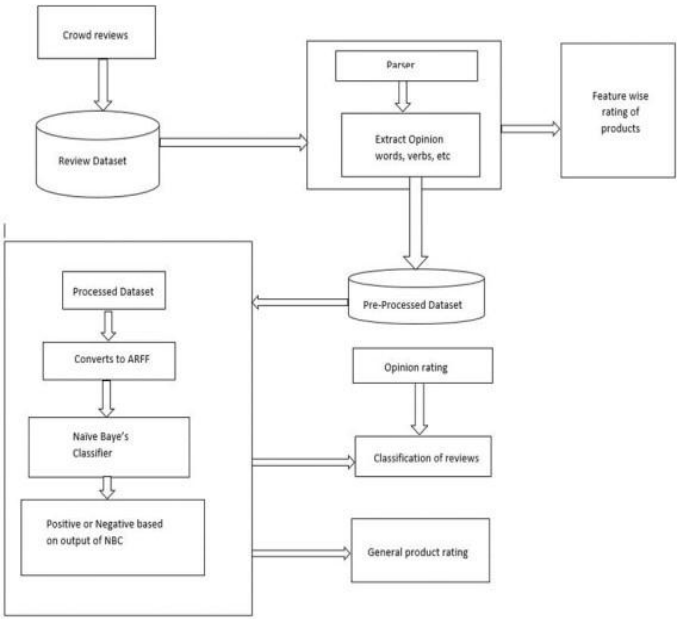
2.Stopword Removal

3.Tokenization

4.Stemming

5. SYSTEM ARCHITECTURE DIAGRAM

Sentiment analysis helps determine whether product reviews are positive or negative, or whether customers are satisfied based on online posts. Marketers can study people's reactions to advertising campaigns and new product releases. It also helps you determine how potential customers view your product and identify the pros and cons of that product from a sales perspective. Extracting the exact specifications of the products people have spoken about gives us concrete data on how price, usability, and viability are rated and compared to competitors' products. Rather than using data from surveys conducted on customers, it gathers valuable knowledge from consumers who are not customers of the company and helps identify reasons for purchasing a particular product. Applications of sentiment analysis can also be useful in many fields.



5.1 System Architecture

5.1 Data Flow Diagram:

DFD is also called bubble chart. It is a simple graphical format that can be used to represent a system in terms of the input data to the system, the various operations performed on that data, and the output data produced by the system. It maps the flow of information in a process or system, how data is processed in terms of inputs and outputs. Defined symbols such as rectangles, circles, and arrows are used to indicate data inputs, outputs, save points, and routes between each target.

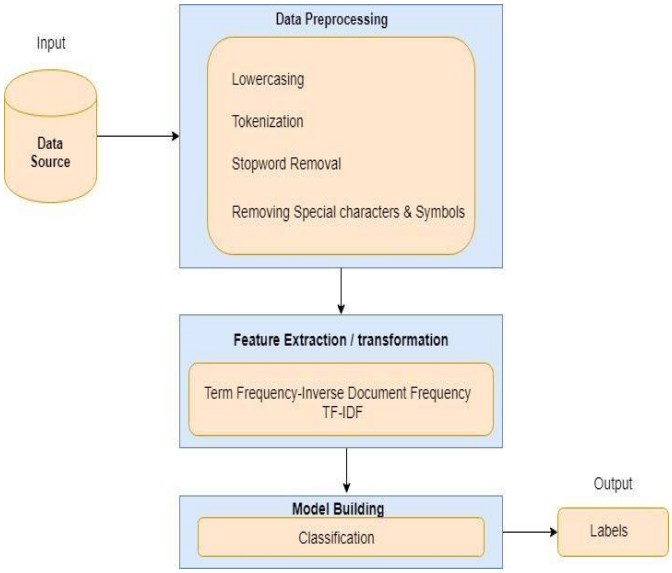


Fig. 5.1.1 Data flow Diagram

5.2 Usecase Diagram:

Use case diagrams are used to show the dynamic behaviour of the system. It encapsulates the functionality of the system by including use cases, actors and their relationships. It represents the high-level functionality of the system and also shows how users interact with it. The main purpose of use case diagrams is to show the dynamic aspects of the system. Accumulate system needs including both internal and external influences. People who call actors and elements responsible for implementing use case diagrams, use cases, and various other things. Represents how entities in the external environment can interact with parts of the system. Before starting to draw a use case diagram, it is important to analyse the system as a whole. Then, once individual features are identified, they are converted into use cases for use in use case diagrams. Then enter the actor that interacts with the system. An actor is a person or thing that invokes a function of the system. Identifies the number of times an actor communicates with the system Use case diagrams are used to show the dynamic behaviour of the system. It encapsulates the functionality of the system by including use cases, actors and their relationships. It represents the high-level functionality of the system and also shows how users interact with it. The main purpose of use case diagrams is to show the dynamic aspects of the system. Accumulate system needs including both internal and external influences.

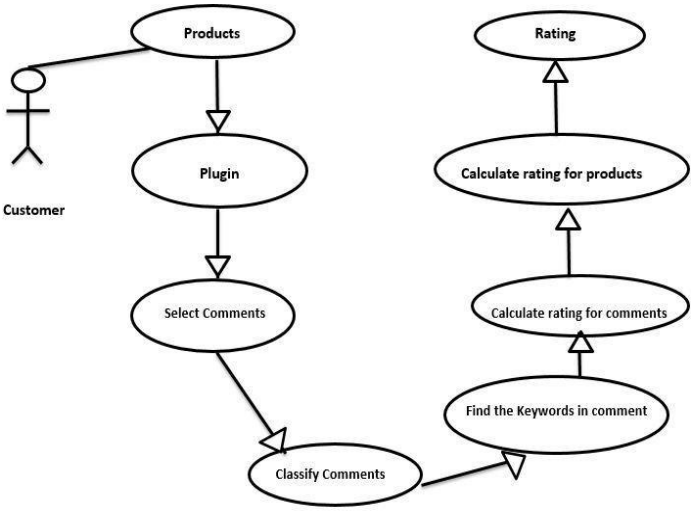
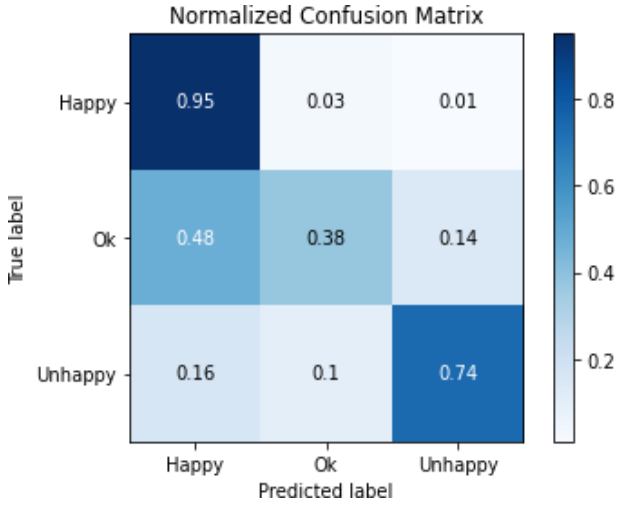


Fig-5.2.1 Usecase Diagram

6. RESULTS



7. CONCLUSIONS

Many users purchase products through e-commerce sites. Thanks to online shopping, many e-commerce businesses have been unable to find out if customers are satisfied with the services they provide. This motivated us to develop a system where multiple customers give reviews of products and services online shopping, thereby helping e-commerce companies and manufacturers get customer reviews to improve services and goods through extracting customer reviews. An algorithm can be used to track and manage customer reviews, uncovering themes and emotional direction from online customer reviews. In this system, users will see many different products and can buy products online. Customers give their opinions on goods and services purchased online. Some keywords mentioned in customer reviews will be extracted and matched with existing keywords in the database based on comparison, the system will evaluate products and services provided by the company. This system will use a text mining algorithm to extract keywords. The system evaluates many different users, based on the rating, the system will tell if the products and services provided by the e- commerce company are good, bad or worse. We use a database of keywords based on sentiment as well as the weight of positivity or negativity in the database, then based on those sentiment keywords extracted from the article reviews of the users in which they are classified. This system is a web application where users will see different products and purchase products online and can give their opinion on online shopping goods and services. This system will help many e-commerce businesses improve or maintain their services based on customer reviews as well as improve goods based on customer reviews.

8. FUTURE SCOPE

Finding biases in vast amounts of unstructured data has become an important research challenge. Industry groups, experts, and academics are now contributing efforts and ideas to find the best systems for verifying fake opinions. Some algorithms are used to analyse fake reviews. Although good results have been obtained, there is still no algorithm that can solve all the challenges and difficulties facing today's generation. Future work and knowledge are needed to further improve the performance of fake reviews. I need more. Every business wants to know what consumers really think about their products and services, as well as those of their competitors, by analysing real reviews rather than fake reviews. There is a great need for such applications in our daily life. A future research direction is to implement the system and validate its performance by applying the proposed approach to various benchmark datasets. Comparing the performance of different classification methods to find the best fit for the proposed erroneous classification method could be another future research direction. However, there are other types of reviews or reviewer functions that are likely to contribute to the predictive task. In the future, we will continue to explore different types of features to make more accurate predictions.

9. REFERENCES

[1]https://www.irjmets.com/uploadedfiles/paper/volume3/issue_6_june_2021/13418/1628083524.pdf

[2]https://ijesc.org/upload/7818a326e3e9a6c1c1f781949cb359ff.E-Commerce%20Product%20Rating%20Based%20on%20Customer%20Review%20Mining%20(1).pdf

[3] https://medium.com/@blogsupport/classification-of-product-reviews-in-e-commerce-website-and-proposing-a-balanced-rating-system-a1ebf6dd192c