IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

VISUAL REPRESENTATION OF PRODUCT REVIEWS USING MACHINE LEARNING AND SENTIMENT ANALYSIS

[1]Surabhi Agarwal , [2]Mohd Usman Khan

[1]P.G Student, CSE Department, Integral University, Lucknow, U.P, India
[2]Assistant Professor, CSE Department, Integral University, Lucknow, U.P, India

***Abstract*: Presently, very huge amount of data is available on internet. This data holds expressed opinions and sentiments. The volume, variety and velocity are the key properties of this data. Decision making on both individual and organizational level is always accompanied by the search of other's opinions on the same. With the tremendous establishment of opinion rich resources like product reviews, feedbacks are proved to be the most essential and valuable resources to market. Sentiment Analysis is an application of Natural Language Processing (NLP), also known as emotion extraction or opinion mining or text mining. It helps to understand the human decision making, categorizing, analyzing and extracting meaningful information in order to understand opinions of consumers. There are several tools and algorithms available to perform sentiment detection and analysis, which are better than unconventional, time consuming and error prone methods used earlier.**

Index Terms - Sentiment Analysis, Opinion Mining, Text Analysis, Natural Language Processing (NLP), Product Review, Data Classification, Polarity Detection.

I. INTRODUCTION

The advancement of electronic commerce with growth in internet and network technologies has led customers to move to online retail platforms such as Amazon, Walmart, etc. People usually rely on customer reviews of products before they buy online. These reviews are often rich in information describing the products and their quality. Customers choose to compare between various products and brands based on whether an item has a positive or negative review. These reviews act as a feedback mechanism for the seller. Through this medium, sellers strategize their future sales and the areas where the product or services needs improvement.

The enormous amount of competition to attract and maintain customers online is fascinating businesses to implement novel strategies to enhance the customer experiences. It is becoming compulsory for companies to examine customer reviews on online platforms such as Amazon to understand better how customers rate their products and services. The purpose of this study is to investigate how companies can conduct sentiment analysis based on Amazon reviews to gain more intuitions into customer experiences. The dataset selected for this research consists of customer reviews of Amazon products, which enables a business person to gain insights on customer reviews regarding specific product and services. The study will enable companies to pinpoint the reasons for positive and negative reviews, followed by implementing effective strategies to address them accordingly. The aim of this research is to help companies to use sentiment analysis to understand customer experiences and customers to understand whether a particular product is to be purchased or not.

II. LITERATURE SURVEY

Sentiment analysis has been present for some time, but many active researches had happened in the past few years to understand and exhibit customer reviews.

Levent Guner [1] from KTH Royal Institute of Technology, Stockholm selected 60,000 random product reviews from Amazon. He used the dataset available in Kaggle that contains 4 million reviews. The performance was compared with three different algorithms namely Naïve Bayes (NB), Support Vector Machine (SVM) and Long short-term memory network (LSTM). The authors used numerous performance metrics to determine the best performing classification algorithm on the test set. To determine the performance, the metrics used were Accuracy, Area Under Curve (AUC), Precision, Recall and F1-score. Based on the results of the evaluation, their study concluded that the LSTM model performed the best with precision > 0.90 and AUC = 0.96 for binary classification.

Xing Fang and **Justin Zhan** [2] collected over 5.1 million product reviews in 4 key categories: beauty, book, electronics, and home. They analyzed these reviews with 3 different classifiers, namely, Naïve Bayes, Support Vector Machine and Random Forest. Their paper addressed the basic question of judging sentiments, categorizing sentiment polarity and ends with random forest generating more accurate results. As per their findings, for larger data sets SVM worked better than Naïve Bayes.

Wan Liang Tan [3] performed both traditional machine learning algorithms including Naïve Bayes, SVM, K-Nearest Neighbor and Deep Learning Network Models such as Recurrent Network Models and LSTM on Amazon reviews dataset. They collected 34627 reviews and divided it into 21000 records of training datasets and 13627 test datasets respectively. In terms of test accuracy, LSTM showed the best performance among all of them with 71.5 percent accuracy. One of the main reasons for not high enough accuracy was the imbalance in their data, as they concluded in their work.

Callen Rain [4] used Naïve Bayes and Decision-List classifiers to sort product reviews from Amazon as positive and negative. He used a corpus that includes 50,000 reviews of 15 items that is used as the research dataset. The features such as bag-of-words and bigrams are compared with each other for labeling positive and negative reviews correctly. His analysis showed that Naive Bayes performed better than the decision-list and bag of words ended up being the best algorithm for feature extraction.

Nishit Shrestha and **Fatma Nasoz** [5] analyzed the reviews present on Amazon to get opinions. They had a model using Recurrent Neural Networks (RNN) with Gated Recurrent Unit (GRU) that learned low dimensional review vector representation using paragraph vectors and product embedding. The data used in this analysis is a collection of about 3.5 million product reviews gathered from Amazon.com. Paragraph Vectors are very much inspired by word vectors. PV system learns vectors by estimating the next term, given several sampled contexts from a paragraph. The concatenation of review embedding developed from paragraph vectors and GRU-derived product embedding is used to train a Support Vector Machine (SVM) to exhibit sentiments. With only review embedding, the anticipated classifier provided 81.29 percent accuracy. The product embedding techniques improved the accuracy to 81.82 percent. Authors assume that a similar technique can be used to acquire user information.

In a **research article** [6] different approach has been implemented for sentimental analysis. In this research, an algorithm called a BoW (Bag of words) is used in which the relation between the words is not taken in account. To measure the sentiment for the whole sentence, the sentiment of every single word of the sentences has been individually decided and values are collected using some aggregation of function. Along with this opinion summarization method based on features can also be used. For each product, a specific feature and their attributes are extracted, and the general feature for each product class is acquired. Then polarity is assigned to each function with the aid of Sequential Minimal Optimization and Support Vector Machines.

III. OBJECTIVE

- To provide visual representation of reviews in the form of word cloud, histogram, box plot.
- To provide visual representation of sentiments of reviews in the form of pie-chart
- To provide product recommendation to users.

IV. SYSTEM DISTRIPTION MODEL

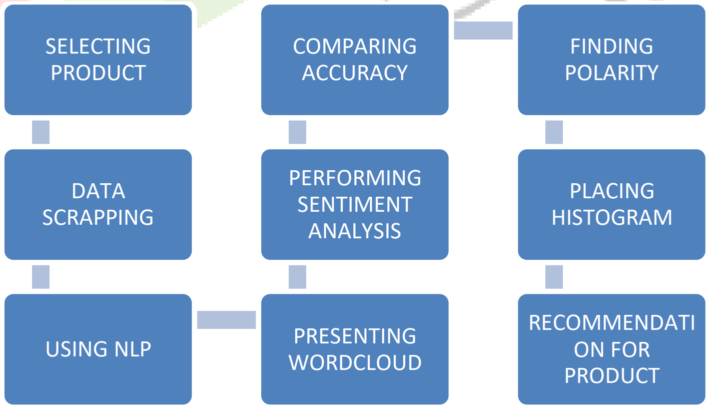

Fig. Model for Sentiment Analysis

4.1 NATURAL LANGUAGE PROCESSING

NLP is a branch of AI that helps computers to understand, interpret and manipulate human languages like English or Hindi to analyze and derive it's meaning. NLP helps developers to organize and structure knowledge to perform tasks like translation, summarization, named entity recognition, relationship extraction, speech recognition, topic segmentation, etc.

NLP aims at converting unstructured data into computer-readable language by following attributes of natural language. Machines employ complex algorithms to break down any text content to extract meaningful information from it. The collected data is then used to further teach machines the logics of natural language. Natural language processing uses syntactic and semantic analysis to guide machines by identifying and recognizing data patterns.

The natural Language Processing procedure is as follows –

4.1.1 <u>Data Collection</u>- The very first job in the process of sentiment analysis is data collection. Data can be collected from various sources like any website, from the several online opinion sets & ratings.

4.1.2 <u>Data Preprocessing</u>- It is the cleaning process of data. Unrequired words & symbols are omitted. This is required for further processing to be streamlined. Part of this move is eliminating hyperlinks, repeated sentences, emoticons, and special characters. It also performs lemmatization and stemming. Finally, it takes a reduced collection of features and passes them to the classifiers.

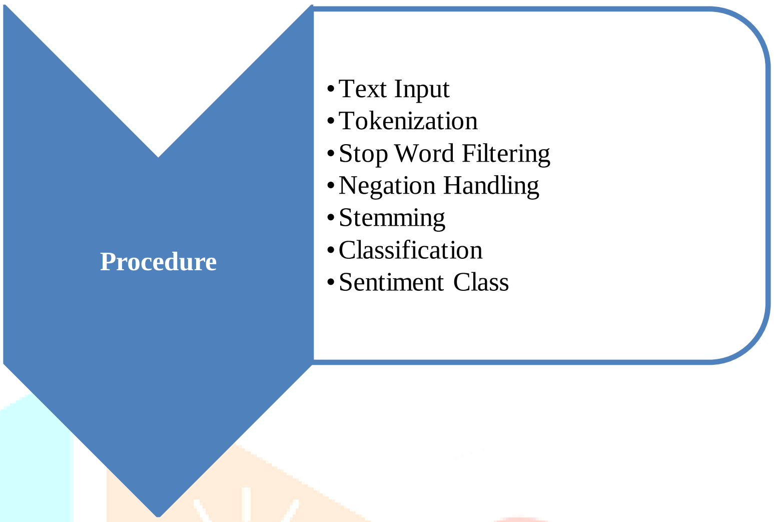

Fig. Natural Language Processing Procedure

4.1.3 <u>Classification-</u> The most critical aspect of a system for sentiment analysis is a classifier. Classification is achieved in negatives, positive, or neutrals categories. A third of the database is usually used as training sets to generate the classifiers. To a large degree, the precision of the classifier relies on the training collection. By using machine learning classifiers like SVM, Bayesian Classifiers and so on, the classification can be performed. However, before training and testing the classifier, machine learning classifiers do feature extraction, which can also use deep neural networks for classifying the data.

4.1.4 <u>Output</u>- After the data has gone through the classifier, the output data is shown. It shows the polarity of feelings of the whole data, and the degree of detail depends upon the type of classifiers which is used. The output can be represented in the form of a word cloud, histogram, box plot, pie chart, graph. They help user to understand reviews easily without spending too much time in scrolling down the list of reviews.

4.2 <u>WORD CLOUD</u>

Word clouds (also known as text clouds or tag clouds) work in a very simple way that the more a specific word appears in a source of text data (such as a speech, blog, post, or database), the bigger and bolder it appears in the word cloud. A word cloud is a collection, or cluster, of words represented in different sizes. The bigger and bolder the word appears, the more often it is raised up within a given text and the more important it is. Also known as tag clouds or text clouds, these are ideal ways to pull out the most relevant parts of textual data, from blog posts to databases. They can also help business users to compare and contrast two different pieces of text to find the word related similarities between the two.

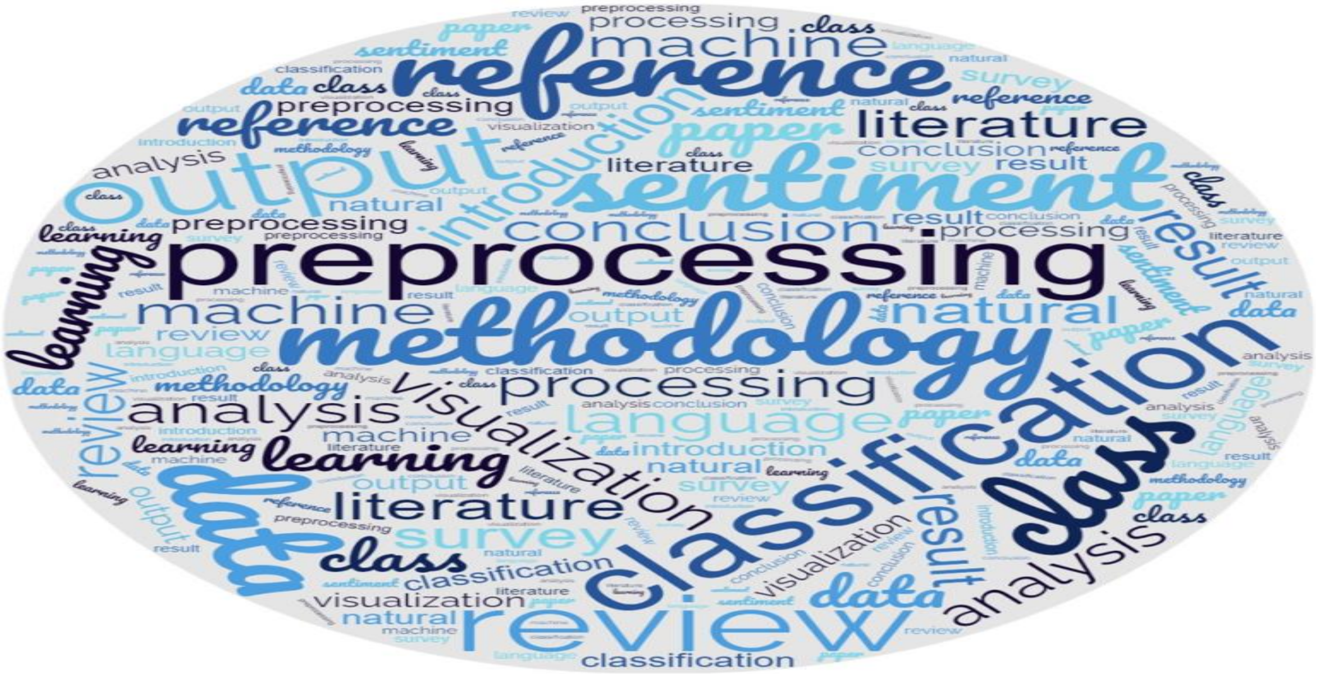

Fig. Word Cloud

4.3 SENTIMENT ANALYSIS

Sentiment Analysis or opinion mining is one of the important tasks of NLP (Natural Language Processing) that has acquired much attention in recent years. The sentiment is a feeling, expression, thought or judgement and using sentiment analysis one can study the target audience's sentiments towards a particular product. It is a form of text analysis that indicates polarity (e.g. a positive or negative opinion) within whole text, sentence or paragraph. Knowing people's emotions is important for companies and first time buyers because consumers can communicate their thoughts and feelings more freely. With the technological improvements in the field of machine learning and automation, companies can create systems that automatically analyzes customers feedback, survey responses and social media interactions. In this way, companies can listen to their customers closely and customize goods and services to suit the needs of their customers

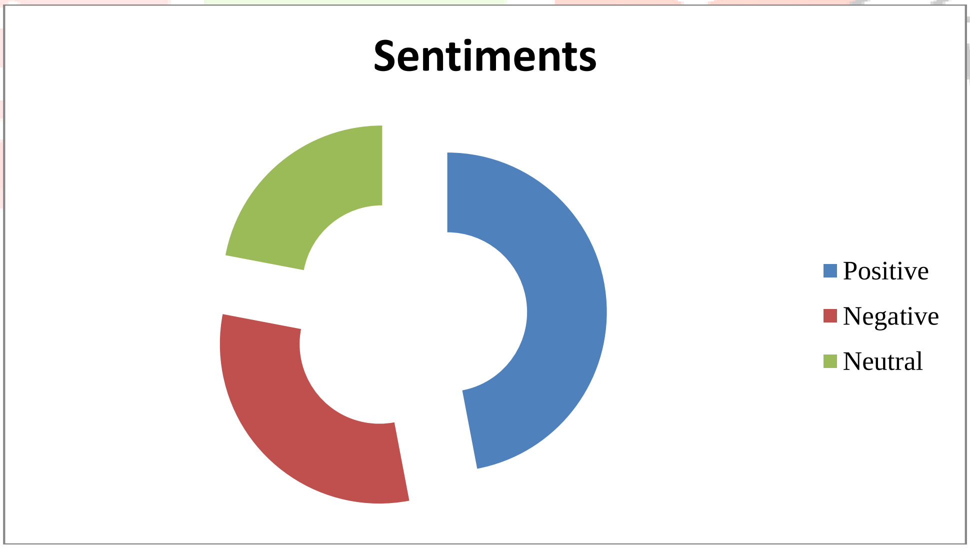

Fig. Categories of Sentiment

4.4 MACHINE LEARNING

ML is being used without specific programming to give computers, the ability to learn. It involves statistical and predictive analysis for allowing the machine to recognize several patterns and to use this knowledge to uncover secret insights into information supplied. Both algorithms for machine learning are split into the supervised machine learning & un-supervised machine learning algorithms. This is accomplished by observing the latest trend and studying how to apply it in a new pattern. On the other hand, unsupervised machine learning algorithms are mainly used to group various specific data types and are applied in various areas where data separation is necessary. In sentiment analysis, classification plays an important role. A pre-classified sample of database which is named a training dataset is used for training and creating a classifier in the classifications stage of sentiments analysis by using machine learning techniques. Upon learning the pattern, this classifier then can label the previous unlabeled data. The precision of a classifier, however, often depends heavily on the data used for training. Therefore, we see that for sentiment analysis, supervised machine learning methods are better suited. The key machine learning classifier used for sentiments analysis are the following:

4.4.1 <u>Naive Bayesian Classifiers</u>- It is believed that the Naive Bayesian Classifier is very simple and easy in terms of implementation. This is not any single algorithm but consists of the set based on Bayes theorem of various classification algorithms. A term used to define an event's probability. This probabilistic classifier utilizes and analyses all the characteristics present in the vector of the function differently, i.e., it considers them independently of each other. By analyzing a pre-categorized collection of documents, we can learn the pattern.

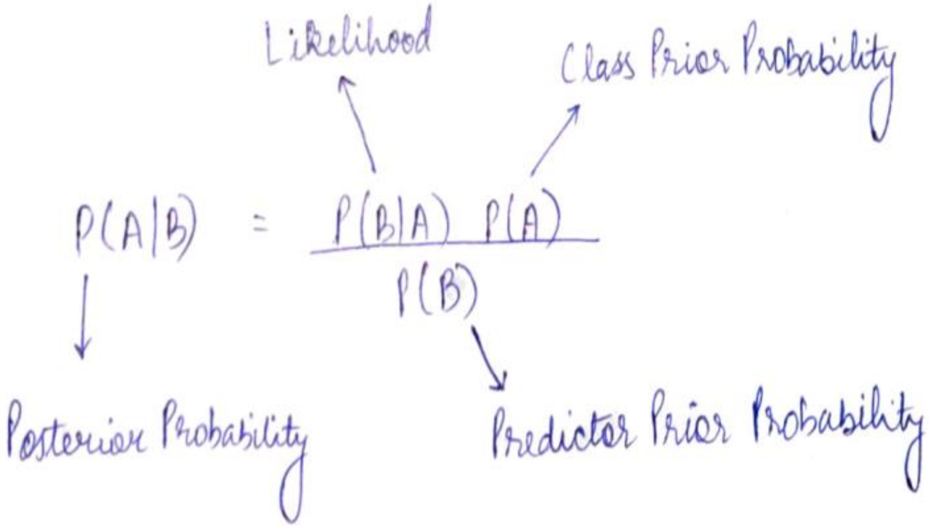

Fig. Formula for Naïve Bayes Classification

This model states that the conditional probabilities of the event P(A) occurring could be determined in presence of the two events, P(A) & P(B) if P(B) has already occurred.. The Naive Bayes classifier's input for training consists of preprocessed data along with its extracted features. The classifications process is conducted on the data set of test data after completing the training and then, depending on the outcomes, the new data. A polarity of the feelings of the data is given by this classification method. For instance, the "It was good" review statement would have resulted in positive polarity.

4.4.2 <u>SVM Classifier-</u>

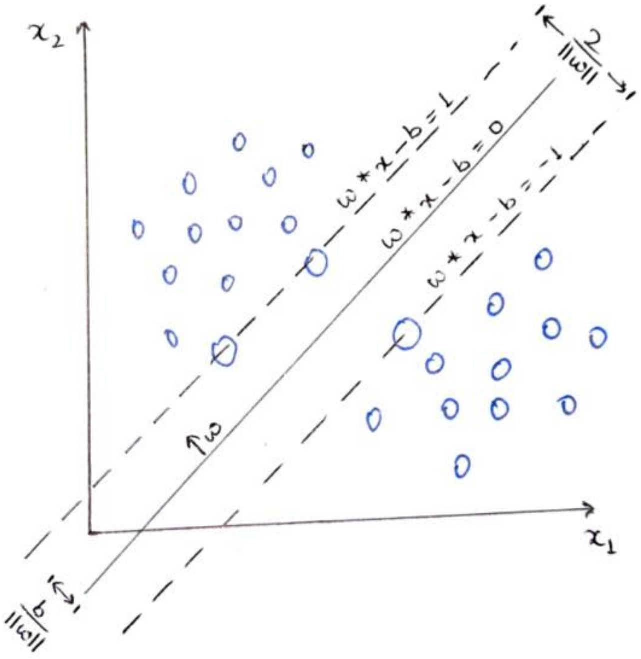

Fig. Graph for SVM

SVM is a popular machine learning technique that employs a statistical approach. It is extremely effective at text classification. There is an n-dimensional space in the SVM, in which n represents number or quantity of features presented in a vector of the function. In the n-dimensional space, each of the data elements presents in the training dataset is registered, the value of each character is the coordinate value.

In this particular n-dimensional space, the key concept of this approach is to find linear separators that best differentiate the various groups. SVM uses a differentiation function with the following parameters: "X" is the vector of the function; the weight vector is "w", and the bias vector is "b". On the training set, the weights & preload vector are automatically learned. Between these two classes, a margin that is far from a document is described. The classifier margins are defined by this distance and indecisive choices are reduced by maximizing this margin. While some features are important to this system, due to the sparse nature of the text, they are correlated and therefore well suited for SVM text classifications.

4.4.3 Decision Tree- For classification issues, the decision trees are mainly used. Depending on the important trait or attributes, also known as the independent variable, the tree is split. Based on these attributes, the space of training data is described in hierarchical form. There is a condition for each attribute value, which is the presence or the absence of one or more than one words. The inner nodes are labeled with characteristics, however, the edges that exit the nodes are called a trace of the weight of the dataset. The name of every leaf in the tree was a group or class.. In this way, in inferring what value is required of the element, a decision tree classifier associates data from an element.

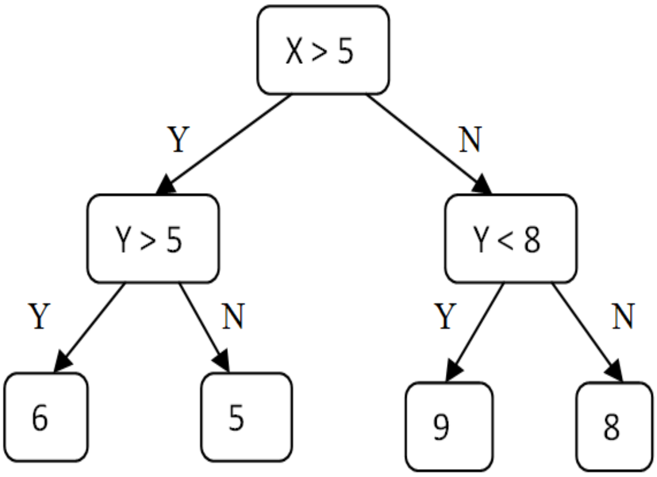

Fig. Decision Tree

4.5 PRODUCT RECOMMENDATION

A product recommendation system is a solution that provides relevant product suggestions to the customers in real time. It is a powerful data filtering platform that depends on algorithms, artificial intelligence, machine learning, and other data analyzing practices. It is a concatenation collecting, storing, analyzing, and filtering customers' data to provide highly personalized relevant products to each and every customer. Relevant products meet the customers' requirements, tastes, and preferences. The quality of data should be very high to achieve such refined targeting at an individual level. But most importantly, we need the right tool to understand the customer data and business needs.

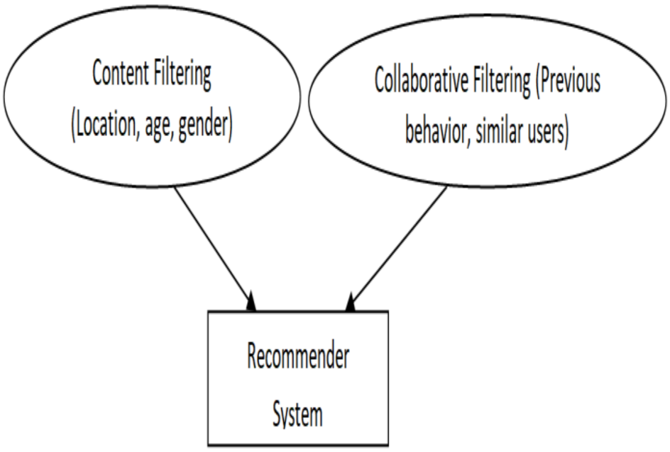

Fig. Model for Recommendation of the Product

V. CONCLUSION

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitude or emotions towards certain entities. This paper tackles with a fundamental problem of sentiment analysis, sentiment polarity categorization and hence, provides recommendation for the product. Users can see a visual representation of reviews instead of scrolling down and reading all the reviews of a product one by one. Hence, users can save their time and effort for finding what they want to purchase. A particular product is recommended if the polarity of the content of the review is positive or neutral, otherwise, the product will not be recommended to the user.

REFERENCES

[1] Raj Sinha, “Data analysis and sentiment analysis on Amazon reviews”, International Journal for Research in Applied Science and Engineering Technology [IJRASET], volume-9, pp. 2200-2206, 2021.

[2] Arwa S. M. AlQahtani, “Product sentiment analysis for Amazon reviews”, International Journal of Computer Science and Information Technology [IJCSIT], volume 13, pp.15-30, 2021.

[3] Somsurva Dutta and Santosh Bothe, “Analysis of Amazon reviews using machine learning approach”, International Journal for Research in Applied Science and Engineering Technology [IJRASET], volume-9, pp. 313-323, 2021.

[4] Xing Fang and Justin Zhan, “Sentiment analysis using product review data”, Journal of Big Data [JBD], 2015.

[5] Waqar Muhammad, Khurum Nazir Junejo, Maria Mushtaq and Muhammad Yaseen Khan, “Sentiment analysis of product reviews in the absence of labeled data using supervised learning approaches”, Research Gate, 2019.

[6] Najma Sultana, Sourabh Chandra, Pintu Kumar and Sk Safikul Alam, “Sentiment analysis for product review”, Research Gate, 2019.

[7] Pravesh Kumar Singh, “Analytical study of feature extraction techniques in opinion mining”, Research Gate, pp.85-94, 2013.

[8] Minu P Abraham and Udaya Kumar Reddy, “Feature based sentiment analysis of mobile product reviews using machine learning techniques”, International Journal of Advanced Trends in Computer Science and Engineering [IJATCSE], volume-9, pp. 2289-2296, 2020.

[9] Tanjim Ul Haque, Nudrat Nawal Saber and Faisal Muhammad Shah, “Sentiment analysis on large scale Amazon product reviews”, IEEE-International Conference of Innovative Research and Development [ICIRD], 2018.

[10] Raheesa Safrin, K.R.Sharmila and T.S.Shri subangi, “Sentiment analysis on online product review”, International Research Journal of Engineering and Technology [IRJET], volume -4, pp. 2381-2388, 2017.

[11] Rajkumar S Jagdale, Vishal S Shrisat and Sachin N Deshmukh, “Sentiment analysis on product reviews using machine learning techniques”, Springer, pp 639-647, 2018.

[12] T.K Shivaprasad and Jyothi Shetty, “Sentiment analysis of product reviews”, IEEE, 2017.

[13] Prashant Pandey, Muskan and Nitasha Soni, “Sentiment analysis on customer feedback data”, IEEE, 2019.

[14] Monir Yahya Ali Salmony and Arman Rasool Faridi, “Supervised sentiment analysis on Amazon product reviews”, IEEE, 2021.

[15] Anjana Madhav C and Lavanya M, “Sentiment analysis of product reviews for overall product rating”, IEEE, 2020.

[16] Duvvuru Mahammad Dawood Khan, “Sentiment analysis of product based reviews”, [IJIRT], volume-8, pp. 467-473, 2021.

[17] Panthati Jagadeesh, Ranga Tarun Kumar, Challa Manish Reddy and Jasmine T. Bhaskar, “Sentiment analysis of product reviews”, Research Gate, 2018.

[18] P Rakesh, M Sandeep and G Jagadeesh, “Amazon product review sentiment analysis using machine learning”, International Research Journal of Computer Science [IRJCS], volume-8, pp.136-141, 2021.

[19] Arpita Lasod and Rahul Pawar, “Sentiment analysis using machine learning techniques”, International Journal of Innovative Research in Technology [IJIRT], volume-6, pp.153-157, 2019.

[20] K Ashok Kumar, “Sentiment analysis of Amazon product reviews using machine learning”, Research Gate, volume-82, pp.5245-5254, 2020.

[21] Kiran Shehzadi and Usman Ahmed Raza, ”Sentiment analysis by using deep learning and machine learning techniques”, International journal of Advanced Trends in Computer Science and Engineering [IJATCSE], volume-10, pp.754-761, 2021.

[22] Sobia Wassan, Xi Chen, Tian Chen, Muhammad Waqr and N Z Jhanjhi, ”Amazon product sentiment analysis using machine learning techniques”, Research Gate, volume 30, pp.695-703, 2021.

[23] Vineet Jain and Mayuri Kambli, ”Amazon product reviews: Sentiment analysis”, Research gate, 2020.

[24] Jyoti Budhwar,“Sentiment analysis based method for Amazon product reviews”, International Journal of Engineering Research and Technology [IJERT], volume-9, pp.54-57, 2021.

[25] Sayyed Johar and Samara Mubeen, “Sentiment analysis on large scale Amazon product reviews”, International Journal of Science Research in Computer Science and Engineering [IJSRCSE], volume-8, pp.07-15, 2020.

[26] P Rakesh, M Sandeep and G Jagadesh, “Amazon product review sentiment analysis using machine learning”, International Research Journal of Computer Science [IRJCS], volume-08, pp.136-141, 2021.