International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.7, July 2021 E-ISSN: 2347-2693

Spammer Detection and Fake User Identification in E-Commerce Site

P.S. Gayke[1*], Snehal Kardile[2], Nutan Dongare[3], Shweta Pathare[4], Pallavi Sakat[5]

[1,2,3,4,5]Department of Information Technology, DVVP COE Ahmednagar, Maharashtra, India

[*]*Corresponding Author: pratibha.gayke@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v9i7.2225 | Available online at: www.ijcseonline.org

Received: 18/Jul/2021, Accepted: 20/Jul/2021, Published: 31/Jul/2021

***Abstract*—** Sentiment analysis is a technique which is used for Natural language Processing, text analysis, text pre-processing etc. are the trending research field in current time. Sentiment analysis using different techniques and tools for analyze and arrange the unstructured data in a manner that objective results can be generated from them. By using these techniques, allow a computer to understand what is being said by humans. Sentiment analysis uses different techniques to determine the sentiment from a text or sentence or expression. The Internet is a huge source of natural language. People share their thoughts and experiences which are subjective in nature. Many a time, it is difficult for customer to identify whether the product shown by seller is good or bad. Companies may also unaware of customer requirements. Based on product reviews it is necessary to understand the perspective of customer towards a particular product. However, these are in huge amount; therefore a summary of positive and negative reviews needs to be generated. In this project, the main focus is on the review of products and techniques used for extract feature wise summary of the product and analyzed them to form an authentic review. Future work will include more product reviews websites and will focus on higher level natural language processing tasks. Using best and new techniques or tool for more accurate result in which the system except only those keywords which are in dataset rest of the words are eliminated by the system.

Keywords—Sentiment Analysis, Polarity, Natural Language Processing, Product reviews

I. INTRODUCTION

Many Customers buys selecting product according to reviews on Internet. The internet is a large source of natural language. People share their thoughts and experiences which are natural. Many a time, it is difficult for customer to differentiate whether the product is good or bad. Companies may not be fully aware of customer requirements. Product reviews can be analyzed to understand the sentiment of the people towards a particular product[1]. The sentences below are extracted from a product review on the Internet product feature database: In Sony mobile, camera is very good but battery life is bad. In this example, reviewer stated that in Sony mobile, camera is very good but battery life is bad. Several tasks are involved to understanding such sentiments. Firstly, data is collection from the e-commerce store amazon.com and Flipkart.com. It consists of customer reviews about various products such as mobile phones, television, camera etc. However, for our project, the focus is on electronic goods. These include but are not limited to phones, televisions, laptops and cameras. The data is collected dynamically through web scraping. Then, the polarity of the opinions must be analyzed. For instance, good and bad respectively hold a negative and a positive opinion. Secondly, for more accurate result can be obtained, the opinion strength of the opinion should determine. If the strength of words used to present the users opinion is taken into account. For example, good and excellent indicate different levels of positive sentiment. The addition of very before any word can also be used to determine the strength. Finally, the classification of review is classified with respect to sentiment classes, i.e. the polarity of reviews are deciding which are positive and negative reviews. The main objective of sentiment analysis on product reviews is to review different algorithm and techniques to extract feature wise summary of a product and analyze it to form an authentic review[2]. Sentiment analysis for a particular product can help companies know about these expectations before product launch. Thus, product analysis can prove beneficial to companies and allow them to develop efficient marketing strategies. The customers will be able to judge a product wisely before purchasing it. Comparative analysis will allow them to pick the right product and the right brand that suits their requirements. The determination of features and their rating will help customers take an informed decision. The aim of sentiment analysis on product review are large number of products are being talked about on the internet. The problem is to extract the information for a particular product or brand and classify the opinion based on various product features. This involves extracting features and training the classifier based on the same. The features should be such that the training speed is improved. A good accuracy on test data has to be achieved. The features of a product that are being talked about have to be determined. That is, it should be known to the users the basic on which the product is given a certain rating. The features have to be classified into positive and negative[3].

Sentiment analysis is a technique which is used for Natural language Processing, text analysis, text pre-processing etc. are the trending research field in current time. Sentiment analysis using different techniques and tools for analyze and arrange the unstructured data in a manner that objective results can be generated from them. By using these techniques, allow a computer to understand what is being said by humans. Sentiment analysis uses different techniques to determine the sentiment from a text or sentence or expression[4]. Many Customers buys selecting product according to re- views on Internet. The internet is a large source of natural language. People share their thoughts and experiences which are natural. Many a time, it is difficult for customer to differentiate whether the product is good or bad. Companies may not be fully aware of customer requirements. Product reviews can be analyzed to understand the sentiment of the people towards a particular product[5].

Sentiment analysis (or opinion mining) is a natural language processing technique used to determine whether data is positive, negative or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs. Sentiment analysis is the process of detecting positive or negative sentiment in text. Its often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers. Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs. For example, using sentiment analysis to automatically analyze 4,000+ reviews about your product could help you discover if customers are happy about your pricing plans and customer service. Maybe you want to gauge brand sentiment on social media, in real time and over time, so you can detect disgruntled customers immediately and respond as soon as possible. The applications of sentiment analysis are endless[6].

II. RELATED WORK

ST Benchmark Towards a Benchmark for Mapping Systems A fundamental problem in information integration is to precisely specify the relationships, called mappings, between schemas. Designing mappings is a time consuming process. To alleviate this problem, many mapping systems have been developed to assist the design of mappings. However, a benchmark designs of mappings[7]. For comparing and evaluating these systems has not yet been developed. We present ST Benchmark, a solution towards a much needed benchmark for mapping systems.

LUBM A Benchmark for OWL Knowledge Base Systems In this paper authors propose a new data model which is an extension of RDF data model to support the type information and on this extended data model we propose a new query language, PredQL (Predicate Query Language). Authors have used the Lehigh University benchmark (LUBM), to evaluate the performance, reasoning capabilities and storage mechanisms of our Semantic Repository. Authors have demonstrated our results against one billion triples[8]. Results show that our method of storing RDF data supports all types of queries with higher scalability and faster query performance.

Semantics and Query Answering Data exchange is the problem of taking data structured under a source schema and creating an instance of a target schema that reflects the source data as accurately as possible. In this paper, authors address foundational and algorithmic issues related to the semantics of data Exchange and to the query answering problem in the context of data exchange. These issues arise because, given a source instance, there may be many target instances that satisfy the constraints of the data exchange problem[9].

On Benchmarking Data Translation Systems for Semantic Web Ontology In this paper, authors present a benchmark that provides a catalogue of seven data translation patterns that can be instantiated by means of seven parameters. This allows us to create a variety of synthetic, domain-independent scenarios one can use to test existing data translation systems. Authors also illustrate how to analyze three such systems using our benchmark. Find the different user classes that you anticipate will use this product. User classes can be differentiated based on use frequency or product functions subset used or technical expertise or privilege levels or educational level and experience. It also describe the pertinent behavior or characteristics of each user class. Few requirements may limited only to specific user classes. Differentiate the very most important or useful user classes for this item or product from those who are less significant to satisfy[10].

III. METHODOLOGY

To conduct the research, training datasets are consider here which retrieve from Stanford dictionary. All the product re- views have been scanned. The data set has been prepared by taking 1554 positive and 4864 negative words,175 stop words from each of the mentioned dictionary. The main goal of the sentimental analysis is to analyze the reviews from the sets and to decide whether quality of product. Data mining is a process of mined valuable data from a large set of data[11]. Review to be analyzed with the use of the discussed data mining methods. A graphical description of the processes involve in sentiment analysis is detailed in Figure 1.

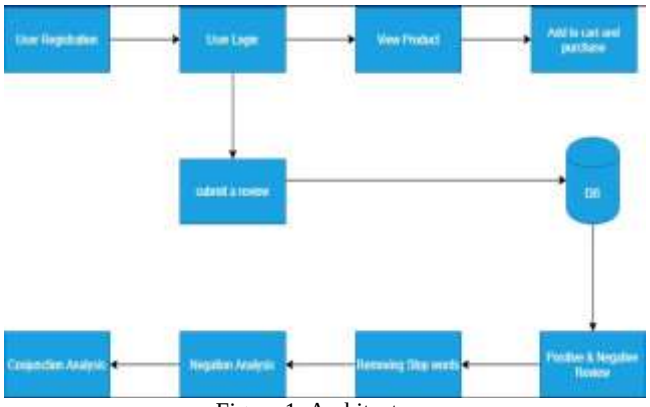

Figure 1 Architecture

Algorithm Naive Bayes :

Bayesian network classifiers are a popular supervised classification paradigm. A well-known Bayesian network classifier is the Nave Bayes classifier is a probabilistic classifier based on the Bayes theorem, considering Nave (Strong) independence assumption. It was introduced under a differ- ent name into the text retrieval community and remains a popular(baseline) method for text categorizing, the problem of judging documents as belonging to one category or the other with word frequencies as the feature. In Nave Bayes technique, the basic idea to find the probabilities of categories given a text document by using the joint probabilities of words and categories. It is based on the assumption of word independence. The starting point is the Bayes theorem for conditional probability, stating that, for a given data point x and class C:

$$P(C / x) = P(x/C)/P(x)$$

Furthermore, by making the assumption that for a data point x = x1,x2,...xj, the probability of each of its attributes occurring in a given class is independent, we can estimate the probability of x as follows:

$$P(C/x)=P(C).P(xi/C)$$

Review	Review Statement	Class
1	I like this procuct	Positive
2	It's a gud booc. Nice story.	Positive
3	look of mobile is bad but features is very good .	Positive
4	Not good.	negative

Figure 2 Example Chart

IV. RESULTS AND DISCUSSION

Detection of skin disease with symptoms and diagnosis and working of the proposed systems are given the below snapshots.

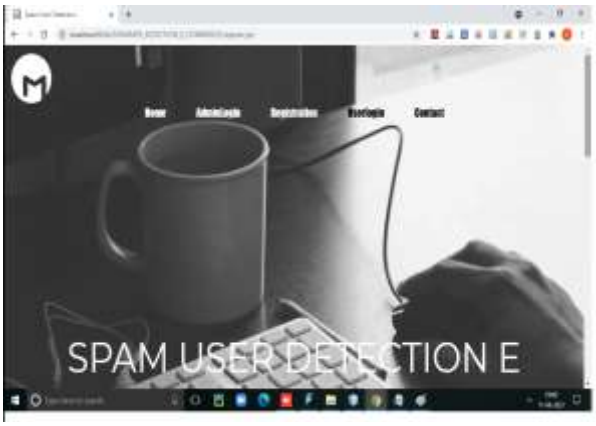

Figure 3 Home Page

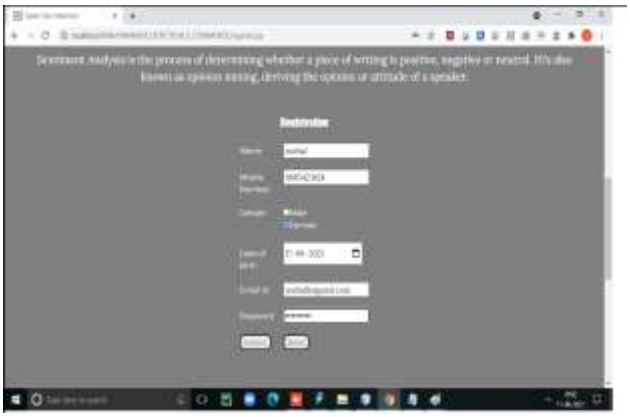

Figure 4 Registration

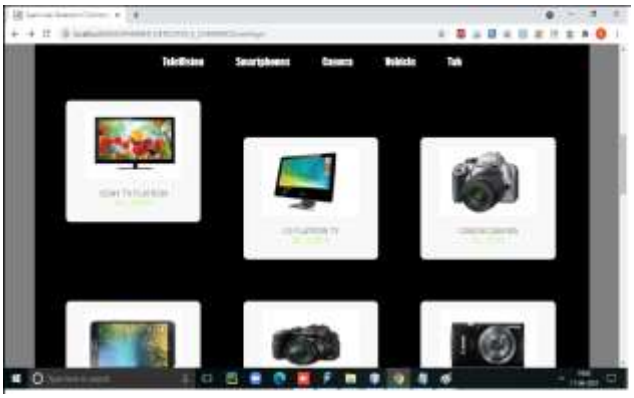

Figure 5 User Interface

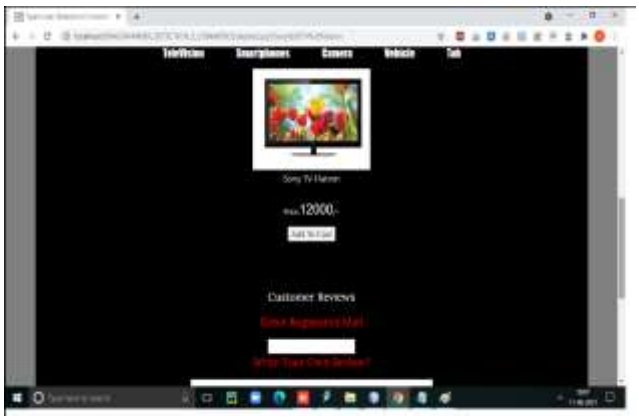

Figure 6 Add to Cart

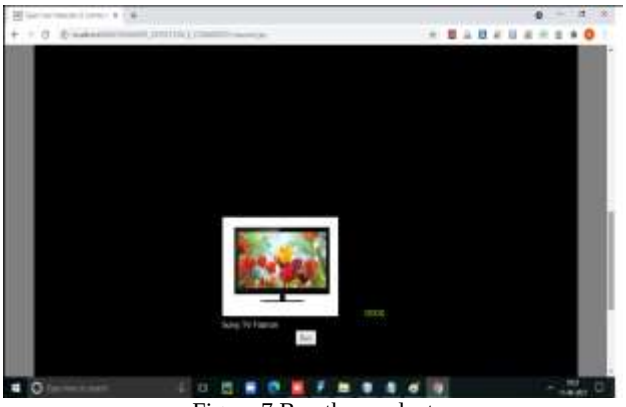

Figure 7 Buy the product

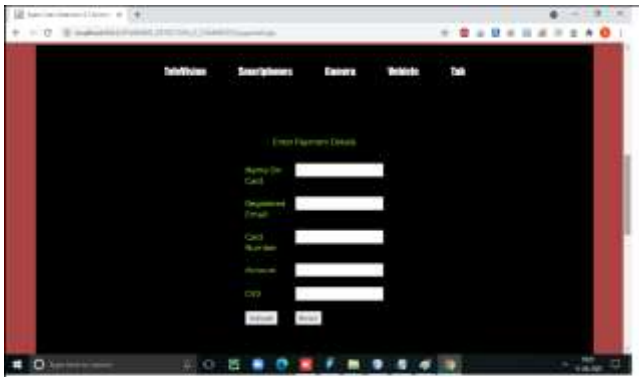

Figure 8 Make Payment

V. CONCLUSION

We have developed a system for sentimental analysis of user reviews. This system discusses various approaches to Opinion Mining and Sentiment Analysis. It provides a detailed view of different applications and potential challenges of Opinion Mining that makes it a difficult task. Some of the machine learning techniques like Nave Bayes, Maximum Entropy has been discussed. Many of the applications of Opinion Mining are based on bag-of-words, which do not capture context which is essential for Sentiment Analysis. The recent developments in Sentiment Analysis and its related sub-tasks are also presented. The state of the art of existing approaches has been described with the focus on the following tasks: Subjectivity detection, Word Sense Disambiguation, Feature Extraction and Sentiment Classification using various data mining techniques.

REFERENCES

[1] D. Vidhate, P. Kulkarni,"*Performance comparison of multiagent cooperative reinforcement learning algorithms for dynamic decision making in retail shop application*", International Journal of Computational Systems Engineering, Inderscience Publishers (IEL), Vol **5**, Issue **3**, pp **169-178**, **2019**.

[2] Tin Kam Ho, "*Random decision forests*", in Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. **1**, pp. 278 – 282, 2018.

[3] D. Vidhate, P. Kulkarni, "*A Framework for Dynamic Decision Making by Multi-agent Cooperative Fault Pair Algorithm (MCFPA) in Retal Shop Application*",Information and Communication Technology for Intelligent Systems, Springer, Singapore, pp **693-703**, **2018**.

[4] R. Moraes, J. F. Valiati, andW. P. Gavio Neto, "*Document-level sentiment classification: An empirical comparison between SVM and ANN*", Expert Syst. Appl.,vol. **40**,no. **2**,pp. 621-633,Feb. 2016.

[5] D. Vidhate, P. Kulkarni,"*A Novel Approach by Cooperative Multiagent Fault Pair Learning (CMFPL)*", Communications in Computer and Information Science, Springer, Singapore, Volume **905**, pp **352-361**, **2018**.

[6] H. Tang, S. Tan, and X. Cheng, "*A survey on sentiment detection of reviews*", Expert Systems Applications, vol. **36**, no. **7**, pp. 10760-10773, Sep. 2018.

[7] D. Vidhate, P. Kulkarni,*"Exploring Cooperative Multi-agent Reinforcement Learning Algorithm (CMRLA) for Intelligent Traffic Signal Control"*, Smart Trends in Information Technology and Computer Communications, Volume **876**, pp **71-81**,**2018**.

[8] The Random Forest Algorithm Towards Data Science.: https://towardsdatascience.com/the-random-forest-algorithm

[9] D. Vidhate, P. Kulkarni, "*A Framework for Improved Cooperative Learning Algorithms with Expertness (ICLAE)",* International Conference on Advanced Computing and Communication Technologies Advances in Intelligent Systems and Computing, Springer Singapore, Volume **562**, pp. **149-160**, **2017**.

[10] Random Forest Simple Explanation William Koehrsen Medium. https://medium.com/@williamkoehrsen/random-forest-simple

[11] D.Vidhate, P. Kulkarni,"*Expertise Based Cooperative Reinforcement Learning Methods (ECRLM)",* International Conference on Information & Communication Technology for Intelligent System, Springer book series Smart Innovation, Systems and Technologies (SIST, volume **84**), Springer Cham, pp **350-360**, **2017** .

