Detection of Fake Online Reviews Using Semi-Supervised and Supervised Learning

Muhammed Hashim H

II MCA Department of Computer Applications,
Nehru College of Management, Coimbatore,
Tamilnadu, India
muhammedhashimh68@gmail.com

Ms Surabhi K S

Assistant professor, Department of Computer Applications,
Nehru College of Management, Coimbatore,
Tamilnadu, India
ksurabhi454@gmail.com

ABSTRACT

Social media's growth has made online data research crucial to comprehending consumer behaviour. In order to assess consumers' thoughts on movies, this study focuses on sentiment analysis of Twitter tweets. To find hidden trends, a lexicon that includes internet reviews and social media terms is created. As e-commerce has grown, internet reviews have a big impact on what people decide to buy. On the other hand, dishonest material has increased as a result of the financial incentives linked to false reviews. False reviews have the potential to increase or decrease sales. This study uses support vector machines, logistic regression, and the Naïve Bayes classifier to determine if reviews are authentic or fraudulent.

Keywords: neural networks, microblogs, e-commerce, and product recommenders

I. INTRODUCTION

Social media's pervasiveness has changed how individuals communicate with businesses and exchange ideas. Consumer decisions are greatly influenced by online reviews, especially in e-commerce, where prospective customers consult reviews before making purchases. But the growing number of phony reviews—which are purposefully written to deceive customers—has sparked questions about their veracity. Fraudulent reviews can skew customer perceptions and have a negative financial impact on businesses by either enhancing a product's reputation or undermining that of a rival.

Because of the complexity of human language, different writing styles, and deliberate deceit, identifying bogus reviews is a difficult process. The large amount of online material makes traditional manual verification ineffective, thus machine learning and natural language processing (NLP) approaches are required. To distinguish between authentic and fraudulent reviews, these techniques examine user behaviour, language characteristics, and review

trends.

To improve false review identification, this work uses supervised and semi-supervised learning methods, such as Naïve Bayes, Logistic Regression, and Support Vector Machines. This study intends to increase classification model accuracy by utilizing sentiment analysis, which would guarantee more trustworthy online reviews and promote confidence in e-commerce platforms. The results aid in the creation of more potent fraud detection systems for online marketplaces.

Because high-star ratings may be strongly associated with positive evaluations, people assume that reviews and reviewers will aid in rating prediction. Therefore, web mining, machine learning, and natural language processing have made understanding how to mine reviews and the relationships between reviewers in social networks a key concern. The rating prediction task is its main focus.

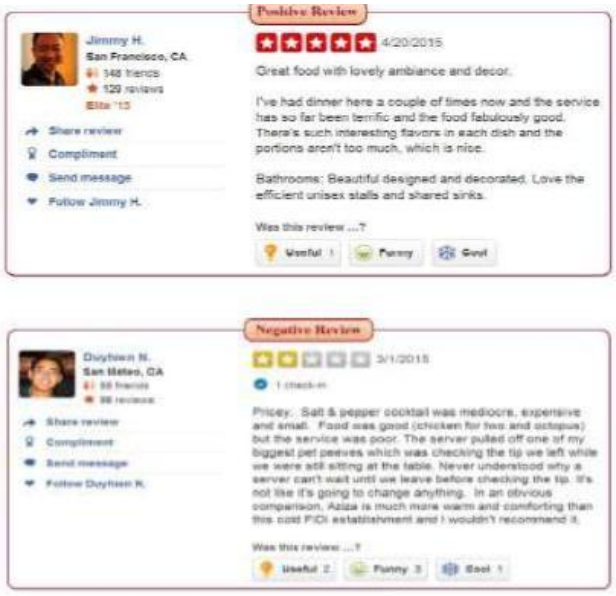


Fig.1: Examples of Positive Review and Negative review on Websites

We naturally illustrate a sample of both good and negative internet evaluations in Fig. 1. A 5-star rating contains a lot of positive words, such "excellent" and "beautiful," as shown in Fig. 1. However, we discover derogatory terms like "expensive" and "bad" in a two-star review. Accordingly, a positive review indicates a high star rating, while a negative review indicates a low one. We can make a decision with ease after we are aware of the benefits and drawbacks of the two types of reviews.

Sentimental Analysis

Sentiment analysis, or simply the process of determining the polarity of the text, is the application of natural language processing to find and extract biased information from source materials. Because it extracts a user's opinion or attitude, it is also known as opinion mining. This is frequently used to explain how individuals think about a certain subject. Sentiment analysis aids in ascertaining the opinions of a writer or speaker regarding a

particular topic or the general contextual polarity of a document. The attitude may be the user's judgment or estimation of their emotional condition at the time of writing.

Sentiment Analysis is hard

Today, Sentiment analysis which uses a variety of machine learning techniques to ascertain the sentiment of vast volumes of text or speech, is becoming increasingly relevant. Examples of application tasks include figuring out how excited someone is about a forthcoming film, relating various opinions about a political party to people's favorable thoughts about voting for that party or turning written hotel reviews into 5-star ratings based on scaling across categories like "quality of food," "services," "living room," and "facilities" offered. It is simple to understand why sentiment analysis is necessary for the vast volume of information published on social media, forums, blogs, newspapers, and other platforms. Processing this data by hand is not feasible in the modern era.

Text Analysis Process

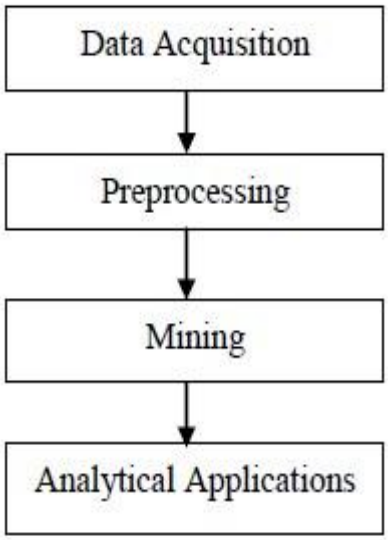


Fig.2 : Process of Analysing Text

The steps involved in this process are as follows :

Data Acquisition: This process collects information from a variety of pertinent sources, including web crawling, document scanning, online reviews, Twitter tweets, and newsfeeds.

Preprocessing is the process of eliminating inconsistent, noisy, and incomplete data. Text preprocessing and feature extraction are the first steps in the classification process.

Three steps are involved in preprocessing:

1. **Tokenization or segmentation**: The process of separating a written language string into its individual words is known as tokenization or segmentation. Tokens are blocks of characters that make up text data. As a result, the documents are being used for additional processing after being separated as tokens.
2. **Elimination of stop words**: These are words that must be filtered, either prior to or following natural language processing. Words that do not convey much information are known as stop words. To facilitate phrase search, a number of tools specifically refrain from eliminating these stop words. For any

purpose, a variety of word collections can be selected as stop words. To improve performance, some search engines eliminate the majority of common words from a text, including lexical words like "want." Many stop words can be found in natural language processing or search engines. It contains English stop words that are regarded as "functional words" because they lack meaning, such as "and," "the," "a," "it," "you," "may," "that," "I," "an," and "of."

3. **Data mining**: The process of using various mining techniques to extract information from stored data. Classification, clustering, statistical analysis, natural language processing, and other techniques are examples of different mining approaches. Text analytics primarily uses classification techniques. A supervised learning technique called classification aids in giving an unclassified tuple a class label based on an instance set that has already been classified. The goal of data classification and identification is to tag the data in order to facilitate its rapid and effective creation. However, many organizations can benefit from re-transforming their data, which speeds up data searches and reduces storage and backup expenses. The main reason for using different data classification technologies is that classification can assist an organization in meeting legal and regulatory requirements to retrieve particular information within a given time frame. Analytical Application: Text mining yields useful information that can be used to enhance decision-making and procedures. Sentiment analysis, document imaging, fraud analysis, and other methods are among them.

II. EXISTING METHODOLOGIES

It can be unfair and deceptive to fabricate reviews and ratings to promote products on your website in an attempt to boost sales and reputation without receiving real feedback. Nowadays, this is a typical practice, and fake review detectors are becoming more and more necessary. The review's content is the main focus of the content-based approach. This is what is said in the review, or its text. Heydari aimed to identify fraudulent reviews by evaluating the review's language. Ott, etc. to categorize data in three different ways. Three strategies include text classification, genre identification, and psycholinguistic deception detection. Behavioural trait-based research focuses on reviewers and takes into account their characteristics. Limetal. resolved particular problems with users who are the origin of spam reviews from various sources or review spammer detection. The behaviour of those who purposefully post phony reviews differs greatly from that of regular users. They observed the following deceptive assessments and evaluation practices. Recognizing, deceptive online reviews is frequently regarded as a classification problem .The primary remedy is supervised text classification algorithms. While training on a sizable dataset of instances with labels from both

classes, false beliefs (positive examples), and genuine beliefs (negative examples), these tactics are robust (negative).

for instance).

Researchers' observations and test findings indicate that current systems classify spam using naive Bayes classifiers and non-spam, which is highly erroneous and might not give users accurate results. The method of semi-supervised classification has been utilized by numerous researchers as well. An established technique for determining scores with parameters is logistic regression analysis. establishes the element's significance.

The Naive Bayes algorithm makes use of non-aligned assumptions and conditional probability of function. Additionally, decisions are used. A tree that chooses which new instances to use based on the value of the attribute. The writers of fake reviews are constantly searching for more effective techniques for producing numerous phony reviews with little assistance from humans. In order to counter these strategies, researchers must test computer-generated phony reviews and develop classifiers using synthetic output to replicate the actions of actual fraudlent reviewers

III. PROPOSED SYSTEM

Every rating takes into account how the word is still used or how it affects the product. As a result, the tokenization procedure is used for all validation of the suggested system. Ratings will be verified and reviewed in light of the data for prompt response.

Due to the necessity and process of cooking words. eliminated following deletion Potential feature words are formed by words. Most of the time

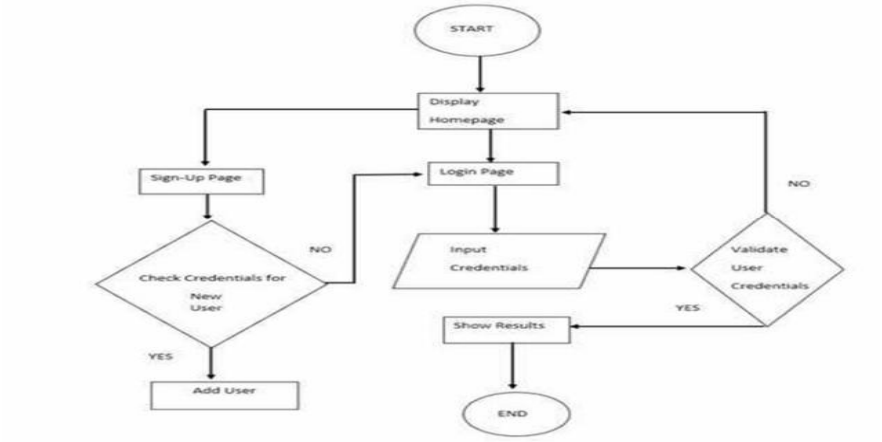
Users frequently use terms that were created instantly and do not accurately represent the original features of the product. Or introduce them to the application of similar-sounding words, copy, write under duress, or do a favour for a loved one. This will establish and facilitate an expandable a potential word functionality version. Recognize the text and contrast it with the suggested system to determine which one it is system lexicon. To determine whether they are displayed, all potential feature words are compiled against this dictionary. Its frequency there is tallied and appended to the feature vector's matching column. for a word map with numbers. The view's length is measured and included with the count frequency in the feature vector. Lastly, the emotion score comes from the dataset that is part of the vector feature. Negative emotions were given a zero in the feature vector. Positive values include value and a pleasant atmosphere.

1.How the system works

The home page, where users can choose to sign up or log in, is displayed when the system first boots up. The user can use their login credentials to log in if they are already a member. Before continuing, new users must create an account. After logging in, users can verify a review's legitimacy by manually entering the review text or by entering the review ID. To ascertain whether the review is authentic or fraudulent, the system employs machine learning algorithms to process the input. For ease of identification, the outcome is then shown in bold.

By assisting users in confirming online reviews, this system guarantees e-commerce platforms' transparency. An automated detection system is essential because fake reviews have the potential to mislead customers and influence their purchasing decisions. The system improves accuracy in detecting fraudulent reviews by applying machine learning techniques. From login to review analysis, it is easy thanks to the smooth user interface. This guarantees that the features of the system are easily accessible to both new and returning users. Users can quickly ascertain the authenticity of reviews thanks to the results' bold display, which enhances readability. This system offers a dependable solution for users looking for reliable feedback before making purchase decisions, as the number of fraudulent reviews on the internet is on the rise.

Fig.3 shows the exact working of the system from the primary section:



IV. SYSTEM DESIGN

1. Input Design

The interface between the user and the information system is the input design. It includes creating specifications and data preparation procedures that are required to transform transaction data into a form that can be processed. This can be done by having people key the data directly or by checking the computer to read data from a written or printed document into the system.

Controlling the quantity of input needed, minimizing errors, preventing delays, and avoiding extra steps are the main goals of input design

and maintaining a straightforward procedure.

The input is made in a way that maintains privacy while offering security and usability. Taking into account the input design

the following items:

1) What information ought to be entered?

2) How should the information be coded or organized?

3) The dialogue to direct the input from the operating staff.

4) Techniques for getting input validations ready and what to do in the event of an error.

2. Objectives

1) The process of translating a user-focused input description into a computer-based system is known as input design. This design is crucial to preventing data entry errors and guiding management in the right direction so they can obtain accurate information from the computerized system.

2) It is accomplished by designing user-friendly data entry screens that can manage high data volumes. The objective of input design is to eliminate errors and facilitate data entry. The layout of the data entry screen ensures that all the data manipulation is possible. It also offers facilities for viewing records.

3) It will verify the accuracy of the data after it has been entered. With the aid of screens, data can be entered. Messages that are appropriate are supplied as needed to prevent the user from becoming stuck in a bind. Consequently, the goal of input design is to produce an input arrangement that is simple to understand.

3. Output Design

A high-quality output is one that clearly conveys the information and satisfies the end user's needs. Any system's outputs convey the processing results to its users and to other systems.

How the information is to be displaced for immediate need and the hard copy output are decided in output design. For the user, it is the most crucial and straightforward source of information. An intelligent and efficient output design enhances the system's ability to support user decision-making.

1) The process of designing computer output should be systematic and well-considered; the appropriate output should be created while making sure that every output component is made in a way that makes the system simple

and efficient for users. They should determine the precise output required to satisfy the requirements when analysing and designing computer output.

2) Choose how to present information.

3) Produce reports, documents, or other formats that include data generated by the system.

An information system's output form should achieve one or more of the following goals.

a) Provide details about previous actions, present circumstances, or anticipated future events.

c) Indicate significant occurrences, chances, issues, or cautions.

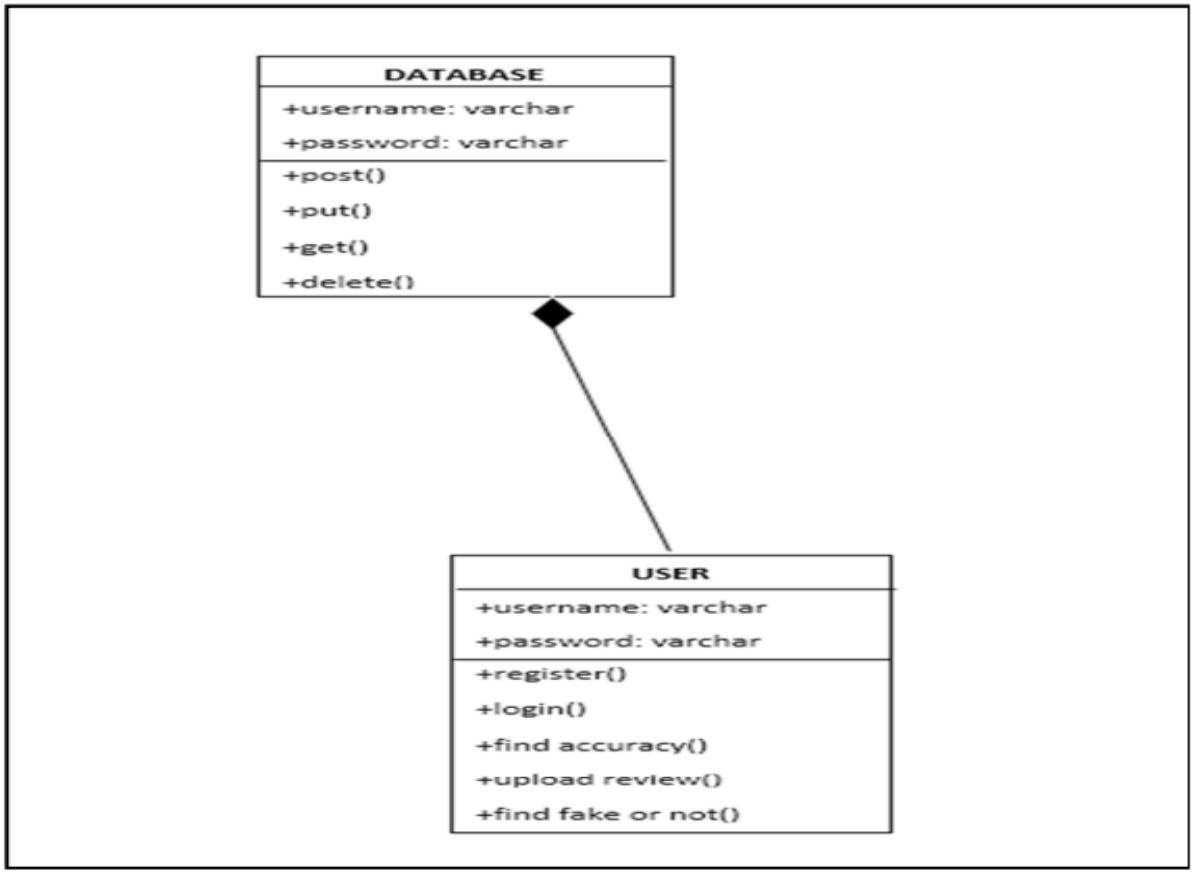
d) Set off an event.

e) Verify a move.

V. UML DIAGRAMS

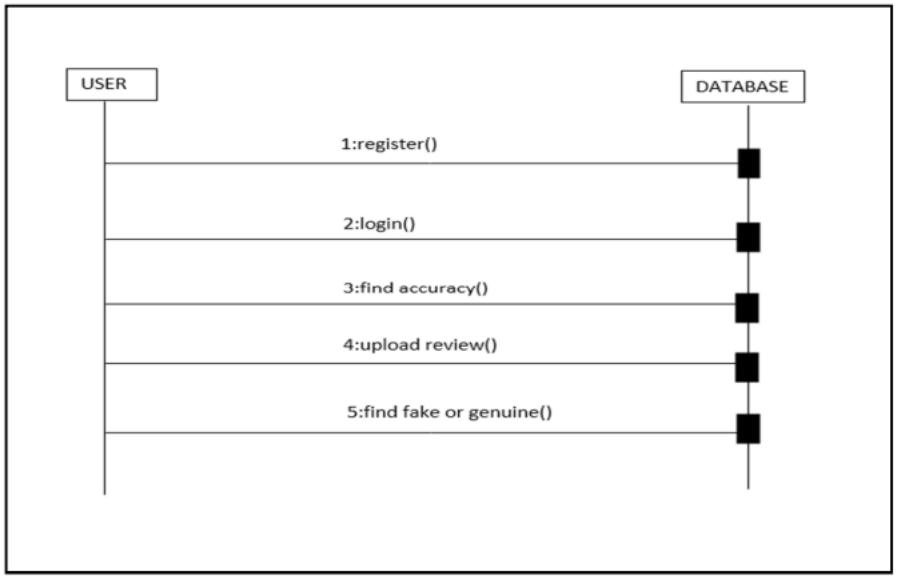
1. Class Diagram

It serves as the primary component of object-oriented modelling. It is employed for both general conceptual modelling of the application's structure and detailed modelling, which involves converting the models into computer code.



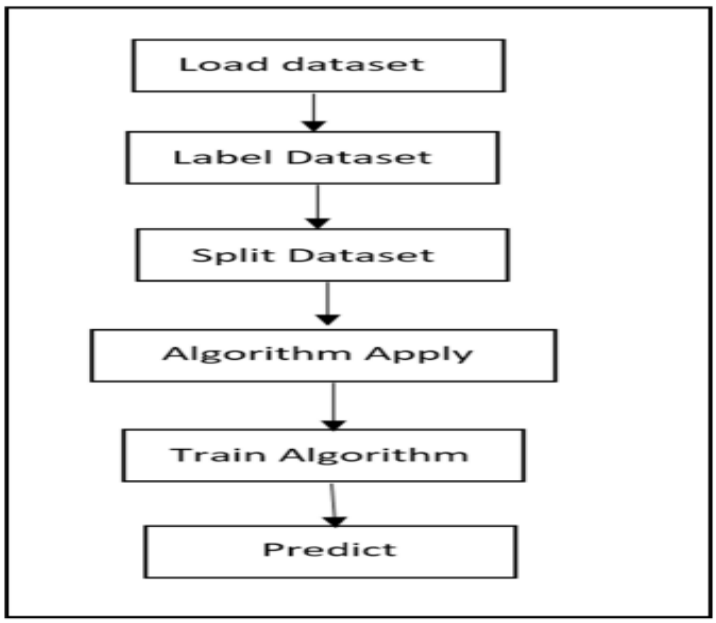
2. Sequence Diagram

It explains how and what in order, it is a type of interaction diagram. It is employed to document an existing procedure or to comprehend the requirements for a new system.



3. System Architecture

It is the conceptual model that outlines a system's behaviour, structure, and other aspects. A system's formal description and representation, structured to facilitate inference about the system's behaviours and structures, is called an architectural description.



VI. RESULT AND SCREENSHOTS

Online fake review detection aids in identifying and eliminating fraudulent reviews from e-commerce platforms and websites. The Python programming language, the DJANGO framework, and a SQL database in the backend were used to implement this project.

On any online platform, including websites and e-commerce sites, this system separates genuine reviews from fraudulent ones. Anyone can sign up for the system and use the reviews and review ID to determine whether or not the review is genuine. The results produced by this system will be determined by the computation of the machine learning algorithm that it uses.

Using machine learning models like Naïve Bayes, Logistic Regression, and Support Vector Machines, the system successfully identifies fraudulent reviews. According to performance evaluation, Twitter data has an accuracy of 82.70% and online reviews have an accuracy of 94.97%; structured reviews have a higher accuracy rate than informal tweets. The accuracy of classification is improved by combining behavioural and linguistic characteristics. By effectively detecting fraudulent reviews, the system assists users in making well-informed choices on e-commerce platforms.

Fig 1:Home Page

This is the web application's home page, from which users can access the sign-up and login pages.

Fig 2:Sign-up Page

Anyone can register for the web application on this sign-up page to verify the legitimacy of a review.

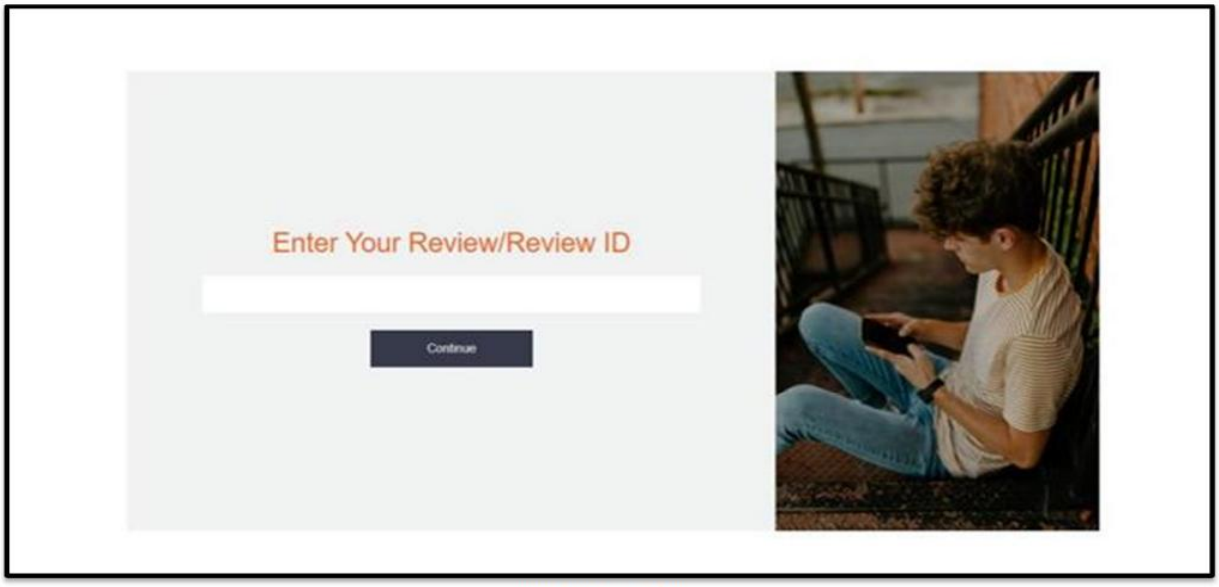


Fig 3:Review Page

This page is displayed to the user once they have successfully logged in. To determine if a review is authentic or fraudulent, a user can input a review or review ID.

VII. CONCLUSION

In this work, we illustrated various supervised and semi-supervised text mining techniques for identifying fraudulent online reviews.

We combined skills from one-of-a-kind investigations to create an advanced characteristic set. Additionally, we employed a unique classifier that was not employed in the previous study. Consequently, we have been able to improve Jiten et al.'s accuracy in advance semi-supervised techniques. We also found that the supervised Naive Bayes classifier is the most accurate classifier. Because we understand that semi-supervised designs function well in situations when honest labelling is not possible, this ensures that our dataset is effectively tagged. Our study was entirely focused on the opinions of consumers.

To develop a more accurate categorization algorithm in the future, consumer movements and text can be incorporated. Better tokenization training techniques could be used to increase the dataset's accuracy. The effectiveness of the suggested technique may be assessed using a sizable statistics collection. Prediction with an enlarged dataset library and an accurate decimal display system will be introduced in the future.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, “Exploiting Product Related Review Features for Fake Review Detection,” Mathematical Problems in Engineering, 2016. [2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection of review spam: a survey”, Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015. [3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the

49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011. [4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry and Word Count: Liwc," vol. 71, 2001. [5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012. [6] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014. [7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010. [8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5, pp. 1319–1327, 2017 [9] Beutel A, Murray K, Faloutsos C, Smola AJ (2014) CoBaFi - Collaborative Bayesian filtering. In: Proceedings of 23rd international conference on world wide web, pp 97–108 [10] Cao Q, Sirivianos M, Yang X, Pregueiro T (2012) Aiding the detection of fake accounts in large scale social online services. In: Proceedings of 9th USENIX symposium on networked systems design and implementation, pp 197–210 [11] Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H (2015) Survey of review spam detection using machine learning techniques. J Big Data 2(1):23 [12] Harris CG (2012) Detecting deceptive opinion spam using human computation. In: Proceedings of workshops at the 26th AAAI conference on artificial intelligence, vol WS-12-08, pp 87–93 [13] Aghakhani H, Machiry A, Nilizadeh S, Kruegel C, Vigna G (2018) Detecting deceptive reviews using generative adversarial networks. [14] Badresiya A, Vohra S, Teraiya J (2014) Performance analysis of supervised techniques for review spam detection. [15] Banerjee S, Chua A, Kim, J (2015) Using supervised learning to classify authentic and fake online reviews. [16] Bhattarai A, Dasgupta D (2012) A self-supervised approach to comment spam detection based on content analysis.