SPAM REVIEW DETECTION USING MACHINE LEARNING

Sinchana M N[1], Kavana S K[2], HemaKrishna[3], Kousalya H[4]

[1] Assistant Professor, Information science and engineering, RajaRajeswari college of Engineering, Bengaluru, Karnataka, India
[2] Student, Information science and engineering, RajaRajeswari college of Engineering, Bengaluru, Karnataka, India
[3] Student, Information science and engineering, RajaRajeswari college of Engineering, Bengaluru, Karnataka, India
[4] Student, Information science and engineering, RajaRajeswari college of Engineering, Bengaluru, Karnataka, India

ABSTRACT

With the Covid-19 pandemic, massive lockdowns and social distancing tactics have contributed to an increase in the prevalence of the WWW and online service platforms. User-generated material, including customer reviews, has proliferated due to the proliferation of products and services offered by this digital world. These evaluations are extremely powerful, influencing both consumer choices and company product improvements. Nevertheless, inside this ecosystem, a harmful practice has surfaced, whereby certain entities utilize phony reviews—fake recommendations for their own products or false criticism of rivals'—in order to profit financially. The credibility of internet business is compromised by this dishonesty, which also misleads prospective customers. This research uses ML techniques to give a comprehensive framework for detecting bogus reviews. The objective of the suggested framework is to determine which categorization method is best suited to assign. Through a thorough examination of multiple machine learning algorithms, the framework aims to build a dependable system to differentiate between authentic and fraudulent reviews, protecting users' interests and promoting trust in online platforms.

Keyword : - Machine Learning, Sentiment Analysis, Fake Review

1. INTRODUCTION

Everyone can freely express his/her views and opinions anonymously and without the fear of consequences. Social media and online posting have made it even easier to post confidently and openly. These opinions have both pros and cons while providing the right feedback to reach the right person which can help fix the issue and sometimes a con when these get manipulated These opinions are regarded as valuable. This allows people with malicious intentions to easily make the system to give people the impression of genuineness and post opinions to promote their own product or to discredit the competitor products and services, without revealing identity of themselves or the organization they work for. Such people are called opinion spammers and these activities can be termed as opinion spamming.There are myriad types of opinion spamming. One type is giving positive opinions to some products with intention to promote giving untrue or negative reviews to products to damage their reputation. Second type consists of advertisements with no opinions on product.

There is lot of research work done in field of sentiment analysis and created models while using different sentiment analysis on various sources, but the primary focus is on the algorithms and not on actual fake review detection. One of many other research works by E. I. Elmurngi and A. Gherbi used machine learning algorithms to classify the product reviews on Amazon.com dataset including customer usage and buying experiences. Utilizing Opinion Mining, a form of language processing, to monitor people's feelings and ideas regarding a product can aid research efforts. Building a system to gather and analyze social media postings, comments, online product and service reviews, or even tweets is known as opinion mining, or sentiment analysis. Machine learning, an aspect of artificial intelligence, is used in automated opinion mining. Software to extract information from datasets and incorporating additional data to enhance performance to build an opinion mining system.

Online and e-commerce reviews of consumer goods, opinions, and services are among the most common uses of opinion mining. E-commerce websites encourage their users to provide feedback and reviews regarding the goods or services they have purchased, as these reviews are very beneficial to both the merchant and the user. Before deciding to buy a product from that merchant, prospective customers check these reviews to learn what past or present users have to say. In a similar vein, vendors and service providers utilize it to pinpoint any flaws or issues customers have with their offerings and to comprehend market data to distinguish their goods of their comparable rivals.Opinion mining has a wide range of applications and uses.
particular customers Before making a choice, a customer can also evaluate the summaries against those of rival products, ensuring they don't pass up any superior options.
Enterprises/Vendors: Opinion mining assists sellers in connecting with their target market and learning how they see their offerings and those of their rivals. These assessments also assist the vendors in identifying problems or shortcomings so that they can enhance subsequent iterations of their product. In the current era, encouraging customers to submit product reviews has shown to be an effective marketing tactic that leverages the voices of actual customers. Such priceless data has been twisted and spammed.

2. LITERATURE SURVEY

Manual review analysis has traditionally been the mainstay of spotting fake reviews, relying on human judgment to spot deception. Researchers have developed rules based on factors like review length, tone, and perceived usefulness to tell apart real feedback from fake ones. Filieri's work highlights various aspects of review assessment, such as content depth, writing style, presence of visuals, and extreme sentiment, as crucial in determining reliability. However, manual detection faces challenges. While heuristic rules provide insights, they lack precision and may miss some fake reviews or be outsmarted by savvy fraudsters. Studies show that traditional methods are less accurate at detection compared to Machine Learning models, especially when dealing with a large number of reviews. Therefore, automated solutions are becoming necessary.

3. Methodology

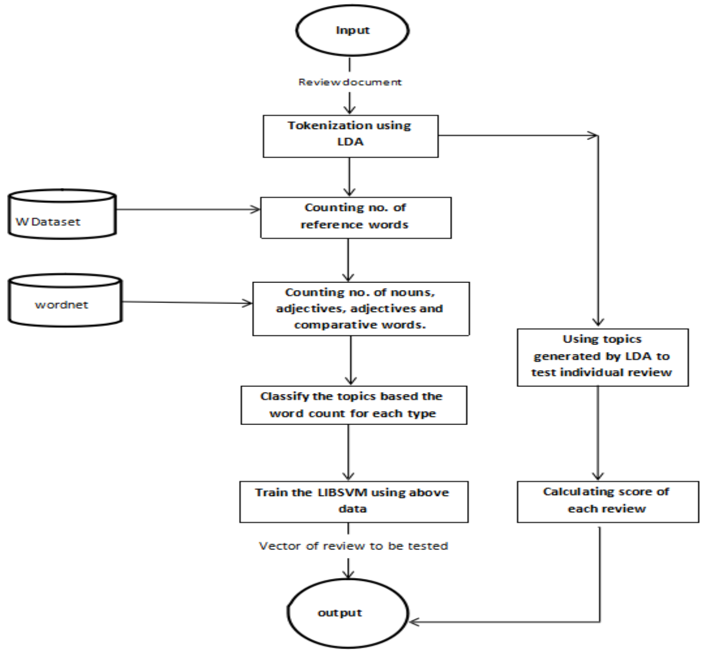
Input
Reviewdocument
Tokenization using LDA
W Dataset
Counting no. of reference words
wordnet
Counting no. of nouns, adjectives, adjectives and comparative words.
Classify the topics based the word count for each type
Train the LIBSVM using above data
Vector of review to be tested
output
Using topics generated by LDA to test individual review
Calculating score of each review

Figure 1 : System Model

1. **Data Collection:** Gathering customer reviews from various sources like Amazon, airline booking sites, hotel and restaurant booking platforms, and CarGurus to create a diverse dataset of 21,000 reviews.

2. **Data Preprocessing:** Cleaning up the collected data by removing irrelevant, redundant, and noisy information. This involves breaking down reviews into sentences, removing punctuation marks, splitting reviews into individual words, and getting rid of common words that don't carry much meaning.

3**. Feature Extraction:** Converting the preprocessed data into a group of features can help identify fake reviews. Includes considering factors like the length of the review, reviewer ID (to spot multiple reviews from the same person), rating given, and whether the review is marked as a verified purchase.

4. **Sentiment Analysis**

Assessing the sentiment of each review to determine if it's positive, negative, or neutral. Research suggests that fake reviews often exhibit stronger emotions than genuine ones, and they may focus more on conveying opinions rather than describing facts. Factors like the ratio of subjective to objective information, positive vs. negative sentiment, and the language used are considered.

5. **Fake Review Detection**: Using classification techniques to categorize reviews as either fake or genuine on the extracted features. All reviews are assigned a weight, and this determines its classification.

6. **Performance Evaluation and Results:** Comparing the accuracy of different models and classifiers to identify the most effective approach. Any enhancements made to improve accuracy are also discussed.

7. **Experimental Configuration:** Implementing supervised learning techniques on the dataset, where the labels of fake and genuine reviews help validate the classification results. Datasets are collected from many sources like hotel and product reviews, and they are combined into a single file for analysis.

4. PROPOSED WORK

The proposed system consists of four main phases to develop an effective fake review detection model, as illustrated in Figure 2.

1. **Data Preprocessing:** This initial phase involves preparing the raw data for analysis. Since raw data is often messy and unfit for machine learning, several preprocessing techniques are applied to clean and organize the data for further analysis.

2. **Splitting Data:** Evaluates the effectiveness of different machine learning algorithms, we divide the dataset into two subsets: the training dataset and the test dataset. The training set used in the model, while the test set is used to assess its performance by comparing the predicted values with the actual ones.

3. **Feature Extraction:** Feature extraction aims to enhance the performance of the model by selecting the most relevant features from the input data. This involves reducing the data to its essential components while eliminating irrelevant or redundant features that may decrease the model's accuracy.

4. **Classification:** In this phase, we employ various classification algorithms to categorize reviews as either fake or genuine. The following algorithms are considered:

Naive Bayes: This algorithm calculates probabilities based on the frequency of occurrences in the dataset, making it suitable for applications like text classification and spam filtering.

K-Nearest Neighbors (KNN): KNN categorizes instances based on the votes of similar cases, using a distance function to measure similarity.

Decision Tree: Decision trees represent a series of decisions done on training data, splitting it on the most informative features at each step.

Support Vector Machines (SVM): SVM finds the optimal hyperplane to separate the data into classes, effectively discriminating between them.

Random Forest: RF constructs multiple decision trees using different dataset samples and features, mitigating overfitting issues commonly associated with decision trees.

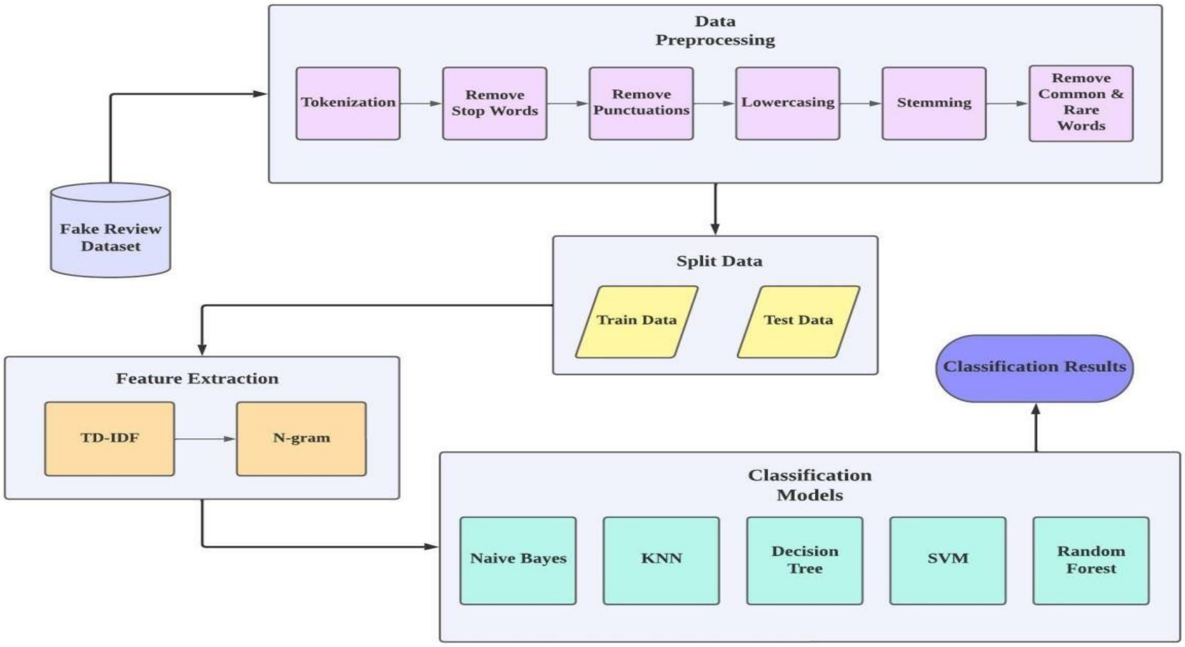

Figure 2 : Proposed Framework

Dataflow Diagram

A Data Flow Diagram (DFD) is a visual tool used to represent how data moves through a system. It helps to illustrate the flow of data from external into the system, how it's processed within the system, and how it's stored and accessed. There are four main symbols used in a DFD:

1. **Squares:** These represent external entities, such as users or other systems, which are sources or destinations of data entering or leaving the system.

2. **Rounded rectangles:** These represent processes or activities within the system that take data as input, perform some processing on it, and produce output data.

3**. Arrows:** These represent the flow of data between different components of the system. Data flows can be electronic or physical items.

4**. Flat three-sided rectangles:** These represent data stores, which both receive information for storage and provide it for further processing within the system.

Overall, a DFD provides a clear and structured visualization of how data moves through an information system, facilitating communication and understanding among stakeholders**.**

LEVEL 0 DATA FLOW DIAGRAM

The Level 0 Data Flow Diagram (DFD) illustrates how the system is broken down into processes, each handling different data flows to or from external entities. It encompasses all the essential functionalities of the system. The diagram identifies internal data stores necessary for system operation and depicts the flow of data between different components. Essentially, it provides a high-level overview of how data moves through the system and interacts with its various parts.

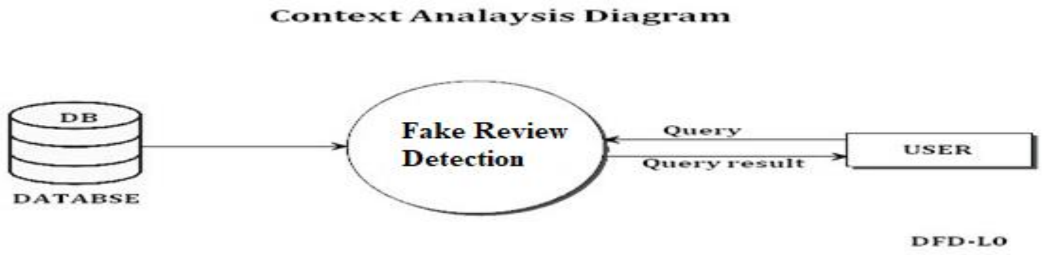

Figure 3: Level 0 Data Flow diagram

In simple terms, intermediate nodes carry out the process, and the requested data is delivered to the requesting node. The process involves two main nodes: the source, which sends the data packet to the respective node, and the destination, which receives the packet and retrieves the required data.

LEVEL 1 DATA FLOW DIAGRAM

In the Level 1 Data Flow Diagram, a file is selected and transferred to a server. The server receives the file and generates a Message Digest (MD). Once the MD file is generated, it retrieves all the public keys belonging to the user group. Then, it combines the MD with the public keys to generate a secure MD. This secure MD is encrypted with the users' private keys and a Ring-Signature generated. Finally, a mail containing the Ring-Signature is sent to all users.

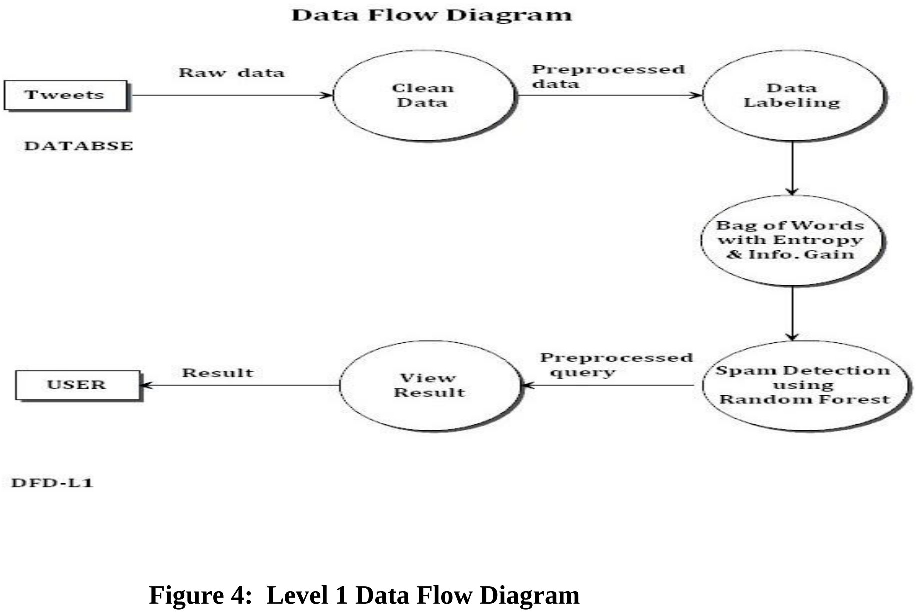

Figure 4: Level 1 Data Flow Diagram

LEVEL 2 DATA FLOW DIAGRAM

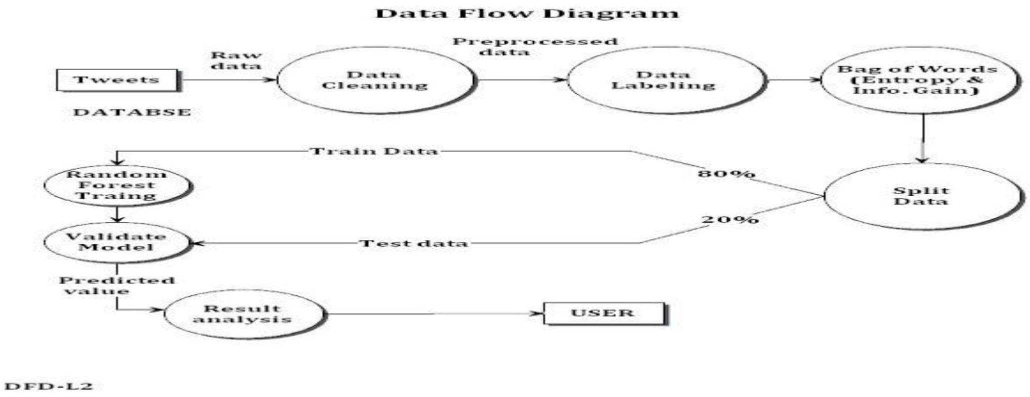

Figure 5 : Level 2 Data Flow Diagram

5. RESULTS AND DISCUSSION

Detecting spam reviews in e-commerce is vital for upholding the credibility and reliability of online marketplaces. Spam reviews can skew product ratings, mislead shoppers, and unfairly sway purchasing choices. Effective spam detection employs various methods, including NLP to spot unusual patterns in review content, ML algorithms to recognize irregularities in reviewer behavior, and network analysis to uncover coordinated actions across multiple accounts. Sophisticated systems also utilize sentiment analysis to flag excessively positive or negative reviews that diverge significantly from typical feedback. By implementing robust spam detection measures, e-commerce platforms can offer a more trustworthy and user-friendly shopping environment, safeguarding the interests of both consumers and legitimate sellers.

6. CONCLUSION

Our spam review detection system filters out fake reviews. We found that SVM classification was more accurate in classifying test data, while Naïve Bayes performed better on training data. This indicates Naïve Bayes's ability to generalize and predict fake reviews efficiently. Our method can be applied to similar datasets successfully.Data visualization was crucial for exploring the dataset, and identified features boosted classification accuracy. Different algorithms were compared on their accuracy. We also provide a feature to recommend the most truthful reviews to assist buyers in making informed decisions. Incorporating new factors like ratings, emojis, and verified purchase status has enhanced classification accuracy. Overall, our approach combines various techniques to effectively detect and filter out spam reviews, improving online marketplace integrity.

7. REFERENCES

1. E. I. Elmurngi and A.Gherbi, “Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques,” *Journal of Computer Science*, vol. 14, no. 5, pp. 714–726, June 2019.
2. J. Leskovec, “WebData Amazon reviews,” [Online]. Available: http://snap.stanford.edu/data/web-Amazon-links.html [Accessed: October 2021].
3. 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*, Baltimore, MD, USA, vol. 1, no. 11, pp. 1566-1576, November 2019.
4. N. O’Brien, “Machine Learning for Detection of Fake News,” [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/119727/1078649610-MIT.pdf [Accessed: November 2018].
5. J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, “Supervised Learning for Fake News Detection,” *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 76-81, May 2022.

6. B. Wagh, J. V. Shinde and P. A. Kale, “A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques,” *International Journal of Emerging Research in Management and Technology*, vol. 6, no. 12, pp. 37-44, December 2017.
7. A. McCallum and K. Nigam, “A Comparison of Event Models for Naive Bayes Text Classification,” in *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*, Pittsburgh, PA, USA, vol. 752, no. 1, pp. 41-48, July 1998.
8. B. Liu and M. Hu, “Opinion Mining, Sentiment Analysis and Opinion Spam Detection,” [Online]. Available: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon [Accessed: January 2022].