FAKE PRODUCT REVIEW DETECTION USING MACHINE LEARNING

1st Mehak Parashar
Computer Science and Engineering
Sharda University, Greater Noida
mehak1512@gmail.com

2nd Shivank Mehrotra
Computer Science and Engineering
Sharda University, Greater Noida
shivank.mehrotra.2004@gmail.com

3rd Suraj Belwal
Computer Science and Engineering
Sharda University, Greater Noida
surajbelwal@gmail.com

Abstract - In today's digital world, the role that genuine reviews play towards the audience is massive, and some companies employing deceptive or misleading reviews on their products to manipulate public opinion is a matter of concern. This research addresses the unavoidable task of fake product review detection by using advanced natural language processing techniques. Inspired by previous work utilizing TF-IDF and BOW features, our approach introduces Word2Vec embeddings for feature extraction, resulting in better performance and accuracy across various machine-learning algorithms. The research deeply examines the detailed approach of pre-processing steps, highlighting the importance of tokenization, stemming, and Word2Vec feature extraction to collect and accumulate meaningful representations of text data. After running a detailed experiment and also after comparing the results of both methods, we establish that Word2Vec-based features yield higher accuracy compared to the TF-IDF and BOW methods. Furthermore, the study evaluates the effectiveness of several classification algorithms, such as Random Forest, Support Vector Machines, K-Nearest Neighbours, Decision Tree, Naive Bayes, and Logistic Regression. The results prove that our Word2Vec-enhanced models every time outperform their counterparts, exhibiting the potential for robust fake review detection. The implications of this research extend to multiple diverse online platforms, particularly e-commerce websites, where the proposed approach can be integrated to automatically identify as well as flag suspicious reviews. By enhancing the trustworthiness of user-generated content, this system empowers customers with more reliable and honest information, developing a transparent and genuine online shopping experience.

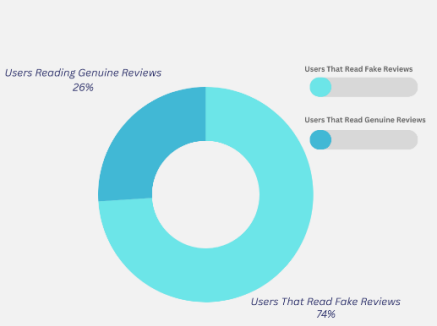
I. INTRODUCTION

In today's world of digitization and globalization, everything is on the internet today. From watching movies to meeting people everything we do today is through the internet. Amongst these wide ranges of activities one major activity, we do through the Internet is online shopping for products or services. E-commerce has emerged as a giant venture in recent years. People tend to buy products/services online nowadays than going to a physical store, and unlike the physical store when you can't judge the stuff by your eyes or feel, the only way to trust or rely on that product/service is through the reviews of that product/service.

As reviews are the only validation which is provided to the online user hence reviews make up the major decision of online buyers.

When you go to buy a product online you can't stop yourself from checking its reviews to see whether that product is exactly what it seems like, whether other people like it or not, what they have to say about it, what's their experience with it, is it of good quality etc. But can you be sure that the review you are valuing so much is written by an authentic customer?

Here is where the concept of fake reviews comes in, the reviews you see might be a review that is not been written by a genuine buyer but by a bot system or spammer just to affect the sales of that product/service. It is startling that nearly 74% of consumers read fake reviews on different sites last year, as represented in the chart below.



Reviews are an important factor for both customers and the producers for the purchase and sales of the product respectively. By planting fake reviews, the scammers try to manipulate the sales and the market value of the product. It can be in both ways as fake reviews can be both the directions negative and positive. These reviews can surely manipulate the consumer's mindset in decision-making as the customer doesn't know whether the review is fake or not.

Users on Amazon can mark whether the reviews are valuable or not. Unfortunately, this still necessitates human effort and is prone to spammers. According to a report on Amazon, only 17% of users truly trust reviews. Only 3% to 10% of individuals post reviews, and 61% of electronic reviews are fake. As of March 2019, Amazon had over 2 million unverified reviews. Detecting fake reviews is complex since no one knows the exact volume of spam.

II. RELATED WORKS

Jacob et al in their research paper for detecting fake product review and their removal using opening mining through machine learning used CNN and RNN to classify and group similarities/dissimilarities in the review dataset. CNN had an input layer, convolution layer, pooling layer, and also output layer. The problem of a large number of layers was solved by introducing the LSTM model. CNN model had the highest accuracy out of all, with 80.33% a recall of 78.02%, and an F1 measure of 82.79%. Both models implemented convolutional neural networks and beat baseline linear SVM in accuracy, therefore the usage of CNNs for spam detection seems promising.

Jiang et al aimed a practical feature-based detection of large-scale fake reviews mainly based on three categories: Linguistic, Behavioural and product factors and information (as a disagreement between the product information and contents of the review allows the revelation of truth.) Analysed it using various feature selections like Internal Text Similarity, Comment Latency, Comment Time Interval, Brand Concentration, etc. AdaBoost, Neural network, Gradient boosting, Logistic regression, Naive Bayes, Random Forest, and Ensemble were tested. All of them were tested on the JD dataset and also on the Amazon dataset. watching both the results and output random forest is the one that produced an accuracy of 0.877, an f1 score of 0.881, and a precision of 0.850. The limitation of this paper is that it is supervised learning based hence the dataset is carefully labelled hence it's labour intensive.

Sadafale et al stated proposed a comprehensive Natural Language Processing (NLP) framework for detecting fake product reviews in online shopping platforms. Their methodology involves rigorous analysis of reviews, emphasizing verified purchases, and employs various algorithms such as Logistic Regression, Support Vector Classifier (SVC), and Random Forest Classifier. Their model aims to strike a balance between accuracy and interpretability, showing promise in fake review detection. Future work suggestions include continuous monitoring and extending similar systems to other online marketplaces for identifying fake reviews.

Deshai et al used Deep Learning Hybrid Approaches to Detect Fake Reviews. This paper introduces two novel deep-learning hybrid techniques, namely CNN-LSTM for detecting fake online reviews and LSTM-RNN for detecting fake ratings in the e-commerce domain, particularly on Amazon datasets. The models leverage advanced word embedding

techniques, including Glove and One-Hot Encoding. Experimental results indicate that the CNN-LSTM technique achieves the highest prediction accuracy in detecting fake online reviews, while the LSTM-RNN model outperforms existing models in detecting fake online ratings with an impressive precision of 93.8%. The study demonstrates the efficiency and practicality of the proposed hybrid models for optimal results in fake online review detection.

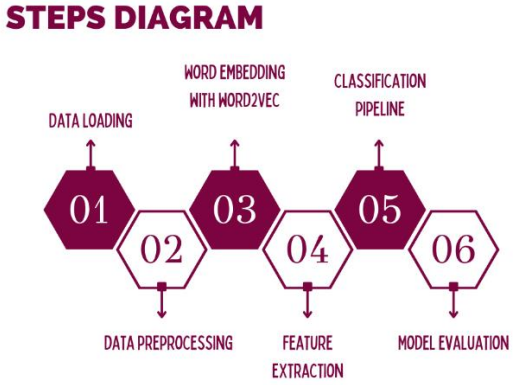
Tabany and Gueffal aimed Sentiment Analysis and Fake Amazon Reviews Classification Using the SVM Model. This project attempts to conduct sentiment analysis of short and long Amazon reviews and report their effects on the supervised learning Support Vector Machines (SVM) model, to bridge for fake reviews classification. Hyperparameter tuning improved the SVM model for sentiment analysis (accuracy of 93%), then altering the review length affected the model's performance, which validated that review length affects the classifier (first assumption). Secondly, conducted fake reviews classification on the fake reviews' dataset yielded 88% accuracy, while the merged subsets of the two datasets yielded 84% accuracy.

Mir et al in their paper addressed the growing issue of fake reviews in online shopping environments and proposed the use of the BERT (Bidirectional Encoder Representation from Transformers) model for extracting word embeddings from reviews. Various classification methods including SVM, Random Forests, and Naive Bayes are employed to improve accuracy. Results demonstrate that SVM outperforms other classifiers with an accuracy of 87.81%, marking a 7.6% improvement over the previous study. The proposed approach relies on NLP tasks and BERT for feature selection, and the results are evaluated using confusion matrices. The study emphasizes the importance of detecting and removing fake reviews to maintain the authenticity of online shopping platforms.

Manaskasemsak et al detected fake reviews through behavioural graphs integrating deep neural networks. The paper proposes two approaches called BeGP and BeGPX for detecting fake reviews and reviewers on online platforms. BeGP constructs a behavioural graph of reviewers based on their characteristic features and labelling patterns. It starts with a small subgraph of known fake reviewers and expands it by inducing other suspicious connected reviewers. BeGPX enhances BeGP by incorporating word embeddings and emotion analysis of review content using deep learning. Both approaches were evaluated on real datasets and outperformed other methods in identifying fake reviewers within the top rankings. BeGPX showed further improvement with limited labelled data.

III. METHODOLOGY

The following section of the research paper describes in detail the various classifiers, feature extraction techniques and algorithms used in the model proposed to predict fake reviews. Below is the flowchart of the Fake Product Review Detection Model.



Data Acquisition and Preprocessing:

In this research paper, the data is obtained and taken from the internet and is publicly available and accessible. It contains approximately forty thousand reviews with almost twenty thousand of each, fake as well as genuine.

Data Preprocessing:

Various steps are part of preprocessing the data. First of all is tokenization which is splitting the text into individual words or tokens. Then there is lowercasing of the word where all the text is converted to lowercase to maintain consistency in data. Eliminating punctuation marks like periods and commas, filtration of common words like 'the', 'is' and 'and' that doesn't carry any specific or valuable meaning.

Word2Vec:

Word2Vec is a famous and commonly used natural language processing technique which helps computers understand the meaning of any word in kind of similar to how a human does. It maps the words as vectors in a vector space, which are mathematical representations and help in capturing the meaning based on where a word is placed. Word2Vec also helps us capture the semantic meaning of words and is a powerful tool that helps convert words into vectors and capture the meaning of that word.

Models Used:

A) Logistics Regression

Logistic regression is a supervised learning algorithm used for binary classification tasks, where the output variable is categorical. It models the probability of a class which we usually represent with 1 using a logistic function, which keeps a check that the output is between 0 and 1.

B) K Nearest algorithm

KNN is a simple classification algorithm which works based on the similarity of data points. In KNN, the class of the unseen data is by the majority vote of the nearest neighbour.

C) Decision tree

Decision tree is a versatile supervised learning algorithm which is used for both classification and regression tasks. Here the tree-like structures classify data by asking a series of binary questions about the features. They offer interpretability and are efficient for handling large datasets.

D) Support Vector Machine

SVM is a supervised learning algorithm which is used for both tasks, regression as well as classification. In SVM an optimal hyperplane is found which separates the data points of different classes. It is capable of handling both linear and non-linear classification tasks and is well-efficient in high-dimensional tasks and binary classification problems.

Working:

Our model deploys a comprehensive approach which helps detect fake reviews by using natural language processing techniques and machine learning algorithms. Initially, the dataset is pre-processed heavily which includes the stop words removal, punctuation removal, lemmatization and stemming. The textual data is also converted to lowercase, and also missing values are handled for a cleaner input. After that, a Word2Vec feature extraction is used to capture the semantic meanings establish relationships between words and map them in vector spaces. These word embeddings are then used to calculate the average vector representation for each review, capturing the overall context of the text. The dataset is then split into training and testing data, and various classification algorithms like logistic regression, support vector machine, decision tree, k nearest neighbour and random forest are trained on the data with the help of the pipeline established. The trained model is then tested and evaluated on testing data with parameters such as accuracy, recall and F1 score, giving a deep perception of the algorithm performance on the fake review model.

IV. RESULTS AND DISCUSSION

the proposed model and technology would help the consumers with the filtering of spam and fraudulent reviews, providing the users with

better and more informed reviews and genuine opinions on products.

Further, this review can be modified using the inclusion of sentiment analysis and more real-time reviews with their positive and negative impact. This finally helps the user make better decisions about buying a certain product or service as well. Furthermore, it can also help genuine businesses and small ones get better sales and generate more revenue.

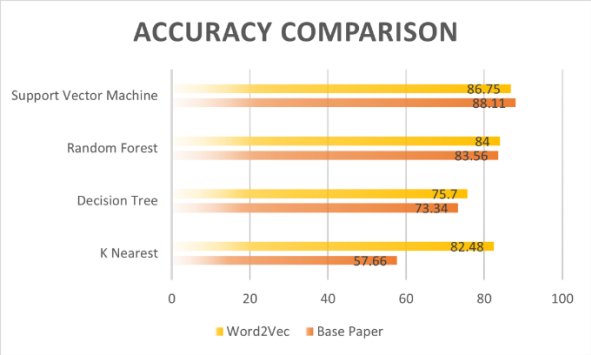
The classification reports of several classifiers, including precision, recall, and f1-score are displayed in the table below. When compared we see that the SVM classifier has the best precision of 0.89 which means that 89% of the time SVM is successful when predicting if a review is fake or not. The recall is which denotes that it detects 86 per cent of all fake reviews properly. An F1-Score of 0.87 summarises the model's predictability.

Features	Type	Precision	Recall	F1-Score
RFC	CG	0.81	0.89	0.85
	OR	0.88	0.78	0.83
SVM	CG	0.86	0.88	0.87
	OR	0.89	0.86	0.87
Decision Tree	CG	0.77	0.77	0.77
	OR	0.77	0.76	0.77
K Nearest	CG	0.84	0.82	0.83
	OR	0.82	0.84	0.83

The most frequent performance indicator for classification is accuracy. The highest accuracy is given by the SVM classifier, 86.70. In the future, the overall model accuracy can also be increased further and it will be helpful for the user or consumer to make a better judgment.

V. CONCLUSION

Below is a comparison chart of accuracy between our models and feature extraction method, compared to the previous similar work done by them. We have attained a better accuracy in the three of the algorithms using a better feature extraction technique.



REFERENCES

[1] Deshai, N., & Bhaskara Rao, B. (2023). Deep Learning Hybrid Approaches to Detect Fake Reviews and Ratings. *Journal of Scientific and Industrial Research*, *82*(1), 120–127. Retrieved from https://ischolar.sscldl.in/index.php/ijsir/article/view/220875

[2] Duma, R.A., Niu, Z., Nyamawe, A.S. *et al.* A Deep Hybrid Model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings. *Soft Comput* **27**, 6281–6296 (2023). https://doi.org/10.1007/s00500-023-07897-4

[3] N. Luo, H. Deng, L. Zhao, Y. Liu, X. Wang and Z. Tan, "Multi-Aspect Feature Based Neural Network Model in Detecting Fake Reviews," 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 2017, pp. 475-479, doi: 10.1109/ICISCE.2017.106.

[4] Jacob, M.S., Rajendran, S., Michael Mario, V., Sai, K.T., Logesh, D. (2020). Fake Product Review Detection and Removal Using Opinion Mining Through Machine Learning. In: Kumar, L., Jayashree, L., Manimegalai, R. (eds) Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications. AISGSC 2019 2019. Springer, Cham. https://doi.org/10.1007/978-3-030-24051-6_55

[5] Bo Guo, Zhi-bin Jiang, "What Is the Internet Water Army? A Practical Feature-Based Detection of Large-Scale Fake Reviews", *Mobile Information Systems*, vol. 2023, Article ID 2565020, 11 pages, 2023. https://doi.org/10.1155/2023/2565020

[6] Mir, A. Q., Khan, F. Y., & Chishti, M. A. (2023). Online Fake Review Detection Using Supervised Machine Learning And BERT Model. *ArXiv*. /abs/2301.03225