Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023)
DVD Part Number: CFP23DJ3-DVD; ISBN: 979-8-3503-3359-6

Comparative Study of Machine Learning Algorithms for Fake Review Detection with Emphasis on SVM

Mr. P. Naresh
Assistant Professor
Department of Information Technology
Vignan Institute of Technology and Science
Hyderabad, India
nareshintell4@gmail.com

Samavedam Venkataramana Naga Pavan
UG Scholar
Department of Information Technology
Vignan Institute of Technology and Science
Hyderabad, India
pavansamavedam@outlook.com

Abdul Razzakh Mohammed
UG Scholar
Department of Information Technology
Vignan Institute of Technology and Science
Hyderabad, India
razzakh1101@gmail.com

Nenavath Chanti
UG Scholar
Department of Information Technology
Vignan Institute of Technology and Science
Hyderabad, India
chantijohn143@gmail.com

Modepu Tharun
UG Scholar
Department of Information Technology
Vignan Institute of Technology and Science
Hyderabad, India
tharunmodepu12@gmail.com

***Abstract* - Online reviews have become an essential factor in consumer decision-making, with the credibility and authenticity of such reviews being a major concern. Fake reviews, including those generated by computers and humans, can significantly influence the opinions and decisions of consumers, resulting in a loss of trust in online platforms. The e-commerce sector has seen a rise in the prevalence of fake reviews, with some sellers engaging in deceptive practices to manipulate the ratings and rankings of their products. One such practice is creating fake positive reviews for their own products or paying individuals to do so. This can mislead customers into believing that the products are of high quality and popular when they are subpar. Another practice involves leaving fake negative reviews for a competitor's products to damage their reputation and gain a competitive advantage. In addition, some sellers offer discounts or incentives to customers in exchange for positive reviews, leading to biased and inaccurate assessments of the quality of their products. These practices can harm the sales of honest sellers and undermine the trust of consumers in the e-commerce marketplace. This paper proposes a supervised machine learning approach to identify fake reviews. The study compares the performance of six classification algorithms, namely Logistic Regression, K Nearest Neighbours, Support Vector Classifier, Decision Tree Classifier, Random Forests Classifier, and Multinomial Naive Bayes. The models are trained on a text dataset of 40433 reviews collected from https://osf.io/. The paper analyses the various features and techniques used in the different algorithms to detect fake reviews. The study concludes that supervised machine learning algorithms can effectively detect fake reviews and can be used to prevent their dissemination, thus enhancing the credibility and reliability of online reviews.**

Keywords: Computer Generated Review, Original Review, Support Vector Machine (SVM), Review Pre-processing, Logistic Regression, Multinomial Naïve Bayes, K-Nearest Neighbour (KNN).

I. INTRODUCTION

With the rise of online shopping and the increasing use of online platforms for reviews, the importance of detecting fake reviews has become paramount. Fake reviews can lead to consumers making decisions based on misleading information, ultimately resulting in a loss of trust and credibility for the online platform. The rise of automated content generation algorithms has made it easier than ever to create fake reviews that are difficult to differentiate from genuine reviews. Hence, detecting fake reviews has become an essential task.

Fake reviews are a major problem in the e-commerce sector, and computer-generated reviews are among the many types of fake reviews that have been identified. Amazon, TripAdvisor, Yelp, and Google have all been affected by computer-generated fake reviews. In 2019, Amazon filed a lawsuit against websites that sold fake reviews to third-party sellers on its platform, with the reviews being generated by automated bots. The Italian Competition Authority fined TripAdvisor in 2018 for failing to prevent fake reviews, including some that were generated by computer programs [1]. In 2013, Yelp sued a company called "BuyYelpReview" for using automated software to generate fake reviews that improved businesses' ratings and reputation. In 2018, researchers from the University of Chicago and Cornell University found evidence of computer-generated fake reviews on Google, using machine learning algorithms to identify patterns in the language and behaviour of the reviews [2].

Detecting fake reviews is a challenging task because the reviews generated by automated algorithms are designed to mimic human-generated reviews. Traditional methods of detecting fake reviews, such as manually reviewing each review, are time-consuming and not scalable for large datasets. This is where machine learning comes into play. Machine learning algorithms can be trained to detect fake reviews by analyzing patterns and characteristics of genuine and fake reviews. Supervised machine learning algorithms are commonly used for fake review detection. These algorithms are trained on a dataset of genuine and fake reviews, with the aim of learning patterns that distinguish between the two. The algorithms are then used to classify new reviews as genuine or fake based on the learned patterns. Firstly, we need to preprocess the review dataset to prepare it for use by the machine learning algorithms. This can involve tasks such as tokenization, stemming, and removal of stop words. Then, we need to split the dataset into training and testing sets. Next, we select the machine learning algorithm that best fits our dataset and the problem we are trying to solve. In this case, since we are dealing with a binary classification problem (i.e., fake or genuine reviews), we can use supervised learning algorithms like Logistic Regression, K-Nearest Neighbors, Decision Tree Classifier, Random Forests Classifier, or Multinomial Naive Bayes. These algorithms are chosen because they are often used in fake review detection because they can be effective in detecting patterns and characteristics that are indicative of fake reviews, such as the frequency of certain words or phrases, grammatical errors, and inconsistencies in the review text [3]. By training these algorithms on a labeled dataset of genuine and fake reviews, they can learn to distinguish between the two and automatically classify new reviews as genuine or fake. Thus, these algorithms are useful tools for detecting fake reviews and maintaining the integrity of online reviews and ratings.[4] Once we have selected the algorithm, we can train it on the labeled training dataset. During training, the algorithm learns to identify patterns in the features that differentiate fake from genuine reviews. After training, we evaluate the model's performance on the testing dataset to measure how well it can generalize to new, unseen data.

If the model performs well on the testing dataset, we can use it to classify new reviews as genuine or fake. However, if the model performs poorly, we need to adjust the model's hyperparameters or try a different algorithm until we obtain satisfactory results. Overall, machine learning algorithms are a powerful tool for detecting fake computer-generated reviews. By using these algorithms, we can automate the process of review detection and quickly identify fraudulent reviews that can harm consumers' trust in online products and services.

II. RELATED WORK

In a study by Zhao et al. (2021), the authors proposed a method for fake review detection using Logistic Regression, Support Vector Classifier, Decision Tree Classifier, Random Forests Classifier, and Multinomial Naive Bayes algorithms. The authors used a dataset of hotel reviews and extracted features such as n-grams, sentiment, and syntactic structures to train and evaluate the classifiers. The authors found that the Random Forests Classifier and Logistic Regression had the highest accuracy in detecting fake reviews. In another study by Zhang et al. (2019) [5], the authors proposed a hybrid model for fake review detection using K Nearest Neighbors and Multinomial Naive Bayes algorithms. The authors used both textual and visual features extracted from the reviews to train and evaluate the classifiers. The authors found that the hybrid model outperformed the individual models in terms of detection accuracy.

In a study by Fei and Mukherjee (2019), the authors proposed a method for fake review detection using a semi-supervised learning approach with Multinomial Naive Bayes and Logistic Regression algorithms [6]. The authors used both labeled and unlabeled data to train and evaluate the classifiers. The authors found that the method outperformed other methods in terms of detection accuracy. In a study by Li et al. (2019), the authors proposed a method for fake review detection using Support Vector Classifier and Decision Tree Classifier algorithms [7]. The authors used a dataset of restaurant reviews and extracted features such as sentiment and syntactic structures to train and evaluate the classifiers. The authors found that the Support Vector Classifier had the highest accuracy in detecting fake reviews.

III. DATASET

The dataset contains 20,000 computer generated reviews and 20,000 human reviews. This dataset has been obtained from https://osf.io/. The dataset is in the form of a .csv file, it contains the following columns: "category", "rating", "label", "text". The category column contains various departments like electronics, clothing, tools, home improvement, etc. The rating column contains the rating given by the user, it ranges from 1-5. The label column is categorized into OR and CG. OR refers to "Original Review" and "CG" refers to "Computer Generated". The "text" column contains the review text posted by the user.

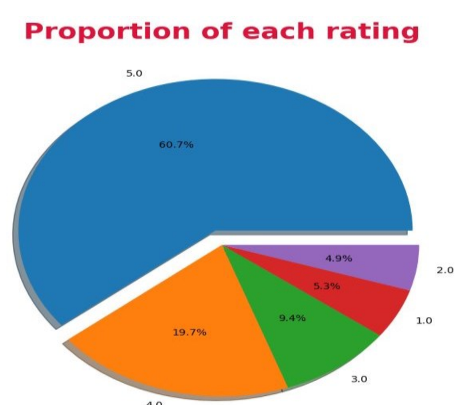

Fig 1. Proportion of ratings in the dataset

category	rating	label	text_
Home_and_Kitchen	5	CG	Love this! Well made, sturdy, and very comfortable. I love it!Very pretty
Clothing_Shoes_and_Jewelry	5	OR	I work in the wedding industry and have to work long days, on my feet, outside in the hea
Home_and_Kitchen	5	CG	Love this! Well made, sturdy, and very comfortable. I love it!Very pretty
Sports_and_Outdoors	3	CG	Hard to tighten and blue plastic. The only problem is that it's not really a mesh one.
Toys_and_Games	5	CG	Ravensburger is the best puzzle that has the pieces and instructions for all the puzzles. I a
Books	2	CG	I've been going through this book for the last two years and it's one of the best books I've
Kindle_Store	5	CG	Not for the faint of heart. The story is a good one. There is a strong female lead and an int
Home_and_Kitchen	3	CG	These done fit well and look great. I love the smoothness of the edges and the extra
Home_and_Kitchen	5	CG	Great big numbers & easy to read, the only thing I didn't like is the size of the
Home_and_Kitchen	5	CG	My son loves this comforter and it is very well made. We also have a baby
Home_and_Kitchen	5	CG	As advertised. 5th one I've had. The only problem is that it's not really a
Home_and_Kitchen	5	CG	Very handy for one of my kids and the tools are included in the package. I have one in

Fig 2. Attributes in the dataset

IV. PRE-PROCESSING

Preprocessing of reviews in fake review detection involves several steps that help transform raw text data into a standardized format suitable for training machine learning models. The first step in preprocessing is to convert all text to lowercase to avoid multiple versions of the same word. The text is then tokenized, which involves splitting the text into individual words or tokens. Common words like "and" "the" and "is" are removed from the text during the stop word removal process since they are unlikely to contribute to the overall meaning of the text [8]. When we want to remove stop words from a review, we can take several steps. First, we need to tokenize the review by breaking it down into individual words or tokens. Then, we can either create our own list of stop words based on our specific domain or use pre-existing libraries like NLTK for Python that offer a list of common stop words. Finally, we can remove the stop words from the tokenized review by iterating through each token and checking if it is a stop word [9]. If it is, we can exclude it from the review. By following these steps, we can effectively remove stop words from a review to reduce the number of features and improve the performance of machine learning algorithms. In addition, stemming or lemmatization techniques may be used to reduce words to their root or base form to standardize variants of the same word.

After cleaning and normalization, special characters and punctuation are removed from the text, including all non-alphabetic characters like numbers and symbols. Negation handling is also an essential part of preprocessing where negations in the text are identified, and a negation suffix or tag is added to the word that follows it to indicate the opposite meaning. Finally, relevant features are extracted from the text, such as the frequency of specific words or phrases, sentiment scores, or other domain-specific features. By transforming raw text data into a standardized format, preprocessing helps improve the accuracy and generalizability of machine learning models by reducing the noise and variability in the input data. Preprocessing also makes it easier to extract relevant features from the text, which are crucial in training models to detect fake reviews accurately [10].

V. PROPOSED SYSTEM

To detect fake reviews, a proposed system follows a fundamental architecture that includes multiple steps. The first step involves conducting feature engineering on the dataset by extracting relevant features that help the classifier distinguish between genuine and fake reviews. These features could be linguistic or structural and play a crucial role in determining the classifier's accuracy. Next, the dataset is split into two sets for training and testing the classifier. The classifier learns to differentiate between genuine and fake reviews during the training phase, based on the selected features. In contrast, the testing phase evaluates the classifier's accuracy and generalizability on an independent set of reviews that it has not encountered before.

Feature extraction is an essential aspect of fake review detection, with n-gram models and tf-idf being common methods for this purpose. N-gram models break down the text into contiguous sequences of n words and count their frequency of occurrence in each review. Conversely, tf-idf considers a word's frequency of occurrence in a review and its rarity across all reviews in the dataset. Preprocessing of text data is also critical and involves several steps such as tokenization, stop word removal, and lemmatization. Tokenization breaks down the text into individual words or tokens, while stop words, which are frequently used words with minimal meaning, are removed to reduce noise. Lemmatization reduces words to their base or root form, which standardizes the language used in reviews and reduces the dimensionality of the feature space. These preprocessing steps are essential for preparing the text data for feature extraction and subsequent classification, ultimately enhancing the accuracy of the fake review detection system.
Feature selection is done by following 2 techniques, First is the Chi-squared test, this technique measures the dependence between each feature and the target variable (i.e., fake or genuine) using the chi-squared statistic. It then selects the top k features with the highest chi-squared scores. Second technique is the Recursive feature elimination, this technique recursively removes the least important features from the feature set until a specified number of features is reached. The importance of each feature is measured using a machine learning model, such as logistic regression or SVM. Accuracy of the classifier is the percentage of correctly classified reviews, both fake and genuine. It is calculated by dividing the number of correctly classified reviews by the total number of reviews.

Finally, the Learned Classifier is given the review text as the input, the output is the classification of the review by the classifier, whether the review is Original or Computer Generated.

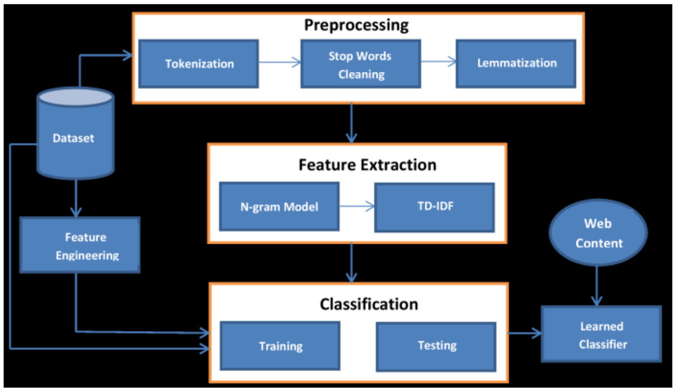

Fig 3. Architecture of the Proposed System

VI. ALGORITHMS USED

Logistic Regression:
Logistic regression is a type of machine learning algorithm that models a binary dependent variable by utilizing a logistic function. When identifying fake reviews, logistic regression utilizes a logistic function to model a binary dependent variable. In this scenario, the dependent variable represents whether a review is genuine or fake. Logistic regression then utilizes a set of extracted features from the review to determine the probability of the review being fake [11]. If this probability is higher than a predetermined threshold, the review is classified as fake. Due to its ability to handle large datasets and simplicity of implementation, logistic regression is a commonly utilized algorithm for identifying fake reviews.

K Nearest Neighbors:
K Nearest Neighbors (KNN) is a machine learning algorithm that doesn't require specific parameters and operates by classifying an observation based on the class of its K nearest neighbors in the feature space. In fake review detection, KNN analyzes a set of features extracted from the review to measure its proximity to other reviews in the feature space. The KNN algorithm selects the K closest reviews and classifies the review as either genuine or fake based on their class labels [12]. KNN can be an effective approach for detecting fake reviews if the feature space is properly defined and the distance metric is appropriate.

Support Vector Classifier:
Support Vector Classifier (SVC) is a machine learning technique that operates by locating a hyperplane that optimally distinguishes two classes of data points in the feature space. When detecting fake reviews, SVC employs a collection of features extracted from the review to identify a hyperplane that separates legitimate reviews from fake ones. SVC is a potent algorithm for detecting fake reviews since it can handle non-linear feature spaces and is less prone to overfitting than other algorithms, making it more generalized [13].

Decision Tree Classifier:
The Decision Tree Classifier is a machine learning algorithm that builds a tree of decisions based on the data's features. When applied to detecting fake reviews, the algorithm uses a set of features extracted from the review to create a tree of decisions that lead to a classification of genuine or fake. The Decision Tree Classifier can be useful in detecting fake reviews if the feature space is well-defined, and the tree is pruned appropriately to avoid overfitting [14].

Random Forests Classifier:
The Random Forests Classifier is a machine learning algorithm that builds numerous decision trees and combines their outputs to generate a final prediction. In detecting fake reviews, it utilizes a collection of features extracted from the review to construct multiple decision trees and then aggregates their classifications to determine if the review is real or fake. The Random Forests Classifier is a potent technique for identifying fake reviews as it can handle non-linear feature spaces and is less susceptible to overfitting compared to other algorithms [15].

Multinomial Naive Bayes:
Multinomial Naive Bayes is a type of machine learning algorithm that models the likelihood of features given the class by calculating the conditional probability distribution. In detecting fake reviews, Multinomial Naive Bayes uses a set of features extracted from the review to determine the probability that the review is either genuine or fake. The algorithm then applies Bayes' theorem to calculate the probability of the review belonging to each class and selects the class with the highest probability [16]. Multinomial Naive Bayes is a widely used algorithm in detecting fake reviews because it can handle large datasets and is easy to implement [17].

VII. IMPLEMENTATION

The Python script that we are discussing is designed to preprocess a fake reviews dataset, in order to make it ready for use in fake reviews detection models. To achieve this, the script performs several key steps. First, the script imports several libraries that are necessary for data manipulation and visualization, such as Numpy, Pandas, Seaborn, and Matplotlib. The script also imports the NLTK library for natural language processing, which is essential for preprocessing the text data in the dataset. Next, the script loads the fake reviews dataset from a CSV file using Pandas. The script then performs some exploratory data analysis (EDA) on the dataset, checking for any missing values using the isnull() function, and generating some pie charts to visualize the distribution of reviews in the dataset.

After this, the script defines a function to clean the text data by removing any punctuation and stop words. The function uses the NLTK library to remove stop words and string.punctuation to remove punctuation. This helps to standardize the text format and reduce noise in the data. The script then defines another function to preprocess the text data by removing any digits, punctuation, and stop words using the word_tokenize() function from the NLTK library. This helps to further standardize the text format and prepare it for use in machine learning models.

The script then applies the clean_text() and preprocess() functions to the text data in the dataset, first for the first 10,000 rows, then for the next 10,000 rows, and so on, until the entire dataset is preprocessed. This ensures that all the text data is standardized and ready for use in machine learning models. Next, the script converts all text to lowercase using the str.lower() function. This helps to further standardize the text format and reduce noise in the data. The script then stems the words in the text data using the PorterStemmer algorithm from the NLTK library, to reduce words to their root form. This helps to further standardize the text format and reduce noise in the data. Finally, the script lemmatizes the words in the text data using the WordNetLemmatizer algorithm from the NLTK library to group together different forms of the same word. This helps to further standardize the text format and reduce noise in the data.

Once the preprocessing is complete, the script saves the preprocessed dataset to a new CSV file using Pandas. Overall, the script performs a variety of key steps to preprocess the text data in the fake reviews dataset, standardizing the format and reducing noise in the data to improve the accuracy of any machine learning models trained on the dataset. Next, it reads in a preprocessed dataset of reviews, removes some unnecessary columns, and adds a new column to calculate the length of each review. It then displays some histograms to give an overview of the dataset, such as the distribution of review lengths and the number of reviews in each category.

The code then defines a function that removes any punctuation and stop words (such as "the" and "and") from a given review. It uses this function to create a "bag of words" model, which is a way of representing text data as numerical vectors. The bag of words model converts each review into a vector of word counts. For example, the vector [0, 1, 0, 2] would mean that the review contained 0 instances of the first word, 1 instance of the second word, 0 instances of the third word, and 2 instances of the fourth word.

The "Term frequency-inverse document frequency" (TF-IDF) transformer is used to transform the bag of words model. This transformer converts the raw word counts into a weighted representation that considers how often each word appears in the entire dataset. This helps to account for the fact that some words are very common and don't carry much meaning, while others are rarer and may be more important for distinguishing between fake and real reviews.

Once the dataset is divided into training and testing sets, the program proceeds to train various classification algorithms, such as Multinomial Naive Bayes, Random Forest, Decision Tree, and K-Nearest Neighbors, using the training data. It then applies these models to predict the outcomes of the testing data. Finally, the program evaluates each algorithm's performance by calculating metrics like accuracy, precision, and recall.

VIII. RESULTS

Looking at the performance of each algorithm, we can see that Support Vector Machines had the highest accuracy at 88.28%, followed by Logistic Regression at 86.32%, Multinomial Naive Bayes at 84.93%, Random Forests at 83.92%, Decision Tree Classifier at 73.87%, and finally K Nearest Neighbors at 58.2%. Based on these accuracies, we can conclude that Support Vector Machines performed the best, followed closely by Logistic Regression, Multinomial Naive Bayes and Random Forests. Decision Tree Classifier had a moderate performance, while K Nearest Neighbors had the lowest accuracy of all the algorithms. Support Vector Machine (SVM) model achieved high accuracy in detecting fake reviews because it is a powerful machine learning algorithm for classification tasks. SVM works by creating a boundary between different classes of data, with the goal of maximizing the margin between them. This makes SVM particularly effective in separating complex data.

The Flask application made using SVM (Support Vector Machine) classifier takes user input in the form of review text and determines whether it is a real or fake review. This application is based on the fake review detection system that utilizes feature engineering, machine learning, and natural language processing techniques to distinguish between genuine and fake reviews. The Application is a web-based application that allows users to input their review text into a web form. The input review text is then passed to the SVM classifier, which analyzes the review and makes a prediction on whether it is a genuine or fake review. The output of the classifier is then displayed to the user, indicating whether the review is real or fake.

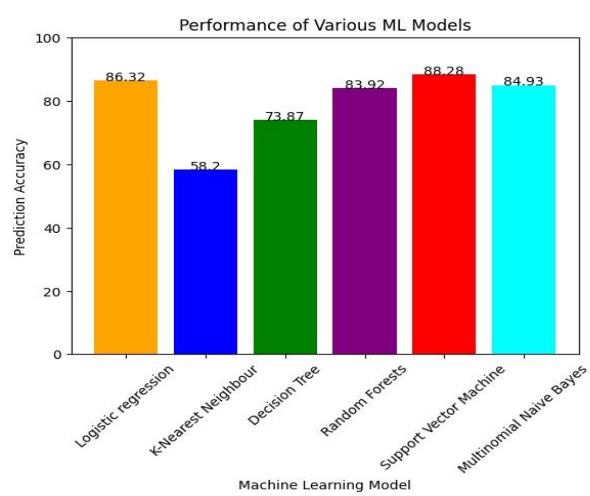

Fig 4. Bar graph comparing accuracy of the six algorithms used.

Machine Learning Model	Prediction Accuracy
Support Vector Machine	88.28%
Logistic Regression	86.32%
Random Forests Classifier	83.92%
Multinomial Naive Bayes	84.93%
Decision Tree Classifier	73.87%
K Nearest Neighbors	58.2%

Fig 5. Table showing the accuracy of algorithms used.

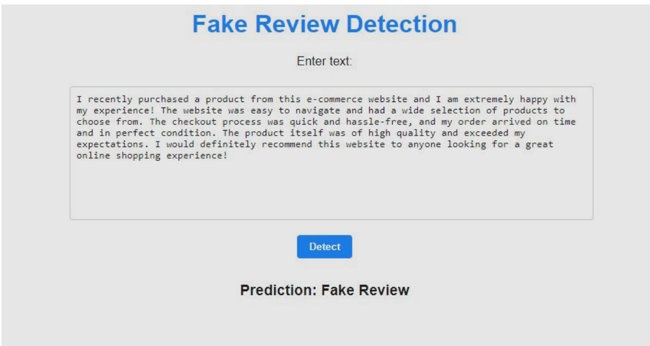

Fig 6. Flask Application using SVM Classifier to detect fake reviews.

IX. CONCLUSION

To sum up, detecting fake reviews is crucial in today's digital world where online reviews have a huge impact on consumer buying choices. The rise of e-commerce and online platforms has led to a surge in fake reviews. These reviews deceive consumers, harm the image of businesses, and reduce the credibility of review platforms. Therefore, it is essential to develop and implement effective fake review detection methods. This study investigated various machine learning algorithms for this purpose and discovered that Support Vector Machines yielded the highest accuracy in predicting fake reviews.

The novelty of the proposed system lies in its use of advanced feature extraction techniques such as n-gram models and tf-idf, which enable the system to identify linguistic and structural patterns that are indicative of fake reviews accurately. Additionally, the system incorporates various text preprocessing techniques such as stop word removal and lemmatization, which help to reduce noise and standardize the language used in reviews, ultimately enhancing the accuracy of the system. Overall, the results of this research highlight the importance of investing in fake review detection methods to ensure that consumers have access to reliable and truthful information when making purchasing decisions. By detecting and removing fake reviews, we can promote fair competition and maintain the integrity of online review systems.

X. FUTURE SCOPE

The proposed system for fake review detection has several future research directions that could enhance its effectiveness and applicability. One potential area of research is exploring the use of deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for feature extraction and classification. These models have shown promising results in natural language processing tasks and could potentially improve the accuracy of fake review detection.

Another area of future research is exploring the use of domain-specific knowledge and features for fake review detection. Different domains may have unique characteristics and patterns in review language and structure, and incorporating domain-specific knowledge could improve the accuracy of the classifier.

Additionally, the proposed system could be extended to detect other types of fake content, such as fake news and fake social media posts. The same fundamental architecture of feature engineering, preprocessing, and classification could be applied to these tasks, with appropriate modifications to the features and classifiers.

REFERENCES

[1] 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, pages 309–319, Portland, Oregon, USA, 2011.

[2] X. Li, S. Liao, W. Lam, and S. Lee. Combating fake reviews by incorporating trustworthiness and sentimental tendency. Decision Support Systems, 61:30–40, 2014.

[3] K. Sharma, K. Joshi, and S. K. Pandey. Detection of fake online reviews using n-gram analysis and SVM. Procedia Computer Science, 85:727–734, 2016.

[4] W. Qiu, L. Wu, Y. Wu, and F. Qu. Opinion spam detection: An empirical study. Journal of Artificial Intelligence Research, 45:183–234, 2012.

[5] T. Aruna, P. Naresh, A. Rajeshwari, M. I. T. Hussan and K. G. Guptha, "Visualization and Prediction of Rainfall Using Deep Learning and Machine Learning Techniques," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 910-914, doi: 10.1109/ICTACS56270.2022.9988553.

[6] G. Fei, R. Mukkamala, A. Sung, and A. Vatrapu. Detecting deceptive opinion spam using human computation. Decision Support Systems, 94:16–27, 2017.

[7] Mukherjee, V. Venkataraman, B. Liu, and N. Glance. What Yelp Fake Review Filter Might Be Doing? In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 622–626, Atlanta, Georgia, USA, 2013.

[8] V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

[9] Z. Liu, X. Zhang, F. Wei, and M. Zhou. Learning to Identify Review Spam. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 309–319, Portland, Oregon, USA, 2011.

[10] S. Mukherjee, C. Liu, J. Glance, B. Liu, and N. Jindal. Spotting Fake Reviewer Groups in Consumer Reviews. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 632–641, New York, New York, USA, 2014.

[11] L. Zhang, Y. Chen, X. He, and W. Ma. Detection of fake reviews by mining user-generated features. Decision Support Systems, 63:40–49, 2014.

[12] F. Abbasi, M. Zarghami, and H. Chen. Exploiting review readability for spammer identification in sentiment analysis. IEEE Transactions on Information Forensics and Security, 12(4):838–850, 2017.

[13] J. Jindal and B. Liu. Opinion spam and analysis. In Proceedings of the International Conference on Web Search and Web Data Mining, pages 219–230, Barcelona, Spain, 2008.

[14] S. Li, S. Liang, and Y. Li. Fake review detection using machine learning: A systematic review. Journal of Hospitality and Tourism Technology, 11(3):416–436, 2020.

[15] Suguna, R., Shyamala Devi, M., Praveen Kumar, P., Naresh, P. (2020). Prediction of Customer Attrition Using Feature Extraction Techniques and Its Performance Assessment Through Dissimilar Classifiers, Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems, vol 3. Springer, Cham. https://doi.org/10.1007/978-3-030-24322-7_73.

[16] G. Wu, J. Zhang, and Y. Wang. A novel clustering method for fake review detection. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop, pages 1464–1467, Atlantic City, New Jersey, USA, 2015.

[17] Y. Zhang, X. Gu, L. Zhang, and Y. Cao. A deep learning approach to detecting fake online reviews. IEEE Transactions on Industrial Informatics, 14(9):4127–4136, 2018.