International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

Deceptive Consumer Review Analysis

I. Ravindra Kumar[1], Swetha Shivannagari[2], Rachakonda Rishika [3], Vegesna Sri Sowmya [4], Sankati Sreeja[5]

Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering &Technology
Vignana Jyothi Nagar,Pragathi Nagar, Nizampet,
Hyderabad, Telangana 500090

Date of Submission: 01-11-2024 Date of Acceptance: 10-11-2024

ABSTRACT: Online consumer review plays an important role in purchasing online products. Availability of various versions and models of a single product makes it difficult to choose. So, most of the customers rely on reviews to purchase the products. These deceptive reviews are being created majorly in two ways both human-crafted and AI-generated. But in today's world most of the reviews are being deceptive either to defame the product or to make fake promotions. To solve this issue many organizations started to rely on manual labor which is a time consuming, biased and costly process. To overcome these problems, there is a need for an automatic model to detect deceptive consumer reviews. While creating this model we plan to use various Machine Learning algorithms like SVM, Multinomial Naive Bayes, Logistic Regression, Decision Trees, Random Forests and KNN to ensure the integrity and accuracy of the system. Among the above algorithms used for our models we got to observe that logistic regression seems to produce the best results for fake or real review prediction with highest accuracy (88%) and for the AI generated or human generated model we have got to see that the Support Vector Machine (SVM) produces results with highest accuracy (85%). This further led us to combine the Logistic Regression model for real or fake review detection and SVM for AI/human generated review detection which helps to find whether a review is fake or real, if at all it is fake then whether it is human generated, or AI generated.

KEYWORDS: Consumer reviews, Exploratory Data Analysis, Machine Learning, Natural Language Processing, Preprocessing

I. INTRODUCTION

In the rapidly expanding world of online shopping, building and maintaining consumer trust is more critical than ever. Online reviews have become a significant factor in deciding purchasing decisions, with buyers relying heavily on the experiences and opinions shared by others. However, the surge in online reviews has brought about a serious issue: the prevalence of fake reviews. Sellers, in their quest to boost their reputations and attract more customers, often resort to unscrupulous methods, including the use of both human-written and AI-generated fake reviews. These deceptive practices not only mislead consumers but also disturb the truthfulness of online shopping sites.

Detecting fake reviews is therefore not just a concern for individual e-commerce platforms but a matter that impacts the overall credibility of the entire online shopping ecosystem. Recognizing the gravity of this problem, our project is dedicated to contributing to ongoing research by developing effective methods to identify and weed out both AI-generated and human-written fake reviews. We are employed a comprehensive approach that combines various factors such as language features, AI probability scores, and detailed reviewer characteristics to enhance the accuracy and reliability of our fake review detection system.

Our Literature survey has two main parts, first for fake review detection, then for detecting if a fake review is human generated or computer generated. To get a complete idea of the current scenario 21 papers have been referred to and the best of qualities have been picked up.

Our methodology involves analyzing the linguistic patterns and stylistic elements of reviews to spot inconsistencies that may indicate fraudulent activity. Additionally, we leverage advanced algorithms to assess the likelihood that a review has been generated by an AI, as opposed to a genuine human experience.

Ultimately, our goal is to create a more trustworthy online marketplace by effectively identifying and minimizing the impact of fake reviews. By developing and implementing these innovative detection methods, we hope to foster a more

transparent and reliable environment for consumers, where they can make informed purchasing decisions based on authentic and trustworthy reviews. Our efforts are aimed at not only protecting individual shoppers but also enhancing the overall integrity of online commerce, thereby contributing to a healthier and more credible digital marketplace.

To ensure the consistent production of accurate and reliable information, a review text analysis approach is employed. This approach harnesses the power of various algorithms, capitalizing on their unique strengths to enhance model performance and deliver robust insights into influential factors. By using diverse methodologies, such as Support Vector Machines, K-Nearest Neighbors, Decision Trees, Random Forests, Logistic Regression, and Multinomial Naive Bayes, the system achieves a comprehensive understanding of complex datasets. This framework not only facilitates the development of a reliable system but also ensures the ability to distinguish between human-written and AI-generated content, thereby upholding online authenticity and fostering consumer trust in reviews and information shared online.

Further this document contains literature survey of 21 papers, then theoretical background of content, proposed approach, methodology, exploratory data analysis, modeling, and results.

II. LITERATURE SURVEY

[1] It uses NLP model and neural network model to detect deceptive reviews. The algorithms like Random Forest, SVM, Bidirectional Encoder Representations and Transformers reduce chances of overfitting whereas the accuracy of the model is depended on how features have been selected after dimensionality reduction.

[2] The system uses Natural Language Processing and Machine Learning Techniques, that uses both supervised learning algorithms like Decision tree, Naive Bayes, Rule-Based Classifier, Bayesian Networks and unsupervised learning approaches like Twice Clustering, K-Mean Clustering. This system achieves huge accuracy in detecting deceptive reviews and captures sentiment using Natural Language Processing Techniques. Nevertheless, the system is not evaluated on a large, diverse dataset of reviews and lacks comparative analysis with other works.

[3] The paper presents a fake review detection system that utilizes both supervised and unsupervised learning techniques, including decision trees, naive Bayes, rule-based classifiers, Bayesian networks, twice clustering, and k-means clustering. The system provides high accuracy and NLP algorithms provides sentiment effectively. Although the evaluation is limited as it does not work on the large and diverse dataset of reviews.

[4] The model utilizes K-Nearest Neighbors, Decision Tree, Random Forest, and Logistic Regression with bigram and trigram techniques. It excels by considering both textual and behavioral features for improved accuracy and compares the performance of various classifiers and language models. However, the limited dataset may affect generalizability across different domains, and future work directions are not clear, requiring further research to enhance the model's robustness and applicability.

[5] This paper used machine learning and performed training on an Online product selling platform. By using logistic regression, it identifies repetitive language, strange phrasing, and extreme positivity/negativity, common in fakes. Consider various factors like reviewer ID, rating, purchase verification, and sentiment for a complete picture. It needs improvements in a few areas. In this model some complex fakes may slip through the cracks and requires continuous updating and adjustments to stay effective.

[6] Predictive model for detecting fake reviews used Simple Logistic Regression Stochastic Gradient Descent, K nearest neighbors, Support Vector Machine Decision trees: Simple DT, Random Forest, Gradient boosted tree, XGBoost . By using three different types of corpus the model classified the reviews. But it is a complex and resource intensive model.

[7] Enhancing NLP techniques for fake reviews used the rough set classifier, decision tree and random forest. This model improved the decision making and enhanced the user experiences by using text of the reviews, rating and usernames. This model leads to mistakes and inaccurate results in algorithms that affect the adaptability in the system. The system may require significant amounts of resources, such as time, money, or computational power, making it costly and challenging to maintain.

[8] Using Bert model, they trained labeled hotel dataset that contains hotel name, polarity, source and text of review. Along with Bert using naive bayes and Support Vector Machine Enhanced Accuracy Levels and made user friendly model. As models use labeled data, that can be time consuming and costly to create and the process itself is complex.

[9] Fake Review Detection System Using Machine Learning model uses Support Vector Machine, Naive Bayes. The system offers a wide range of

features, covering various aspects to enhance performance and adaptability. It includes features focused on text analysis, metadata and reviewer perceptive that provides a well-rounded approach. However, the system performance is dependent on feature selection that creates oversampling to balance data can create challenges like overfitting.

[10] A deep learning approach for detecting fake reviewers using classification and regression trees (CART), support vector machine (SVM), and naive Bayes (NB) is built in 2023 by using text of the review rating and time of reviews. By using features like local dependency, behavior sensitive extraction and context aware attention improves data analysis. The combination of these features and text offers detailed outcomes. However, this adds complexity and can be costly to create and compute and data imbalance might affect performance.

[11] The model was developed on a dataset of restaurant reviews by using n-gram model and max features. They have used two different feature extraction techniques, which are then coupled with five distinct machine learning classification algorithms. They have stated among various experiments, passive aggressive classifiers got highest accuracy in finding the fake review and implementing various deep learning techniques enhanced the detection process by improving accuracy and robustness over traditional machine learning methods. But focusing solely on restaurant reviews may limit the generalizability of the model to other domains or types of deceptive content. Data augmentation introduces noise or bias into the dataset if not carefully implemented.

[12] The paper discusses detecting fake online reviews using semi-supervised and supervised learning methods, such as the Expectation-Maximization (EM) algorithm, Support Vector Machines (SVM), and Naive Bayes classifiers. The model improves the performance by using extra information and combining with supervised algorithms like svm, naive bayes for different types of reviews. Yet, it can be sensitive to parameters like gamma for svm. The iterative nature of the Expectation-maximization algorithm used in semi-supervised learning may introduce additional complexity to the model.

[13] According to A deceptive reviews detection model: Separated training of multi-feature learning and classification detection of deceptive reviews are mainly of traditional methods and intelligent models. This paper used a feature fusion strategy uses three independent models: Text CNN , Bidirectional Gated Recurrent Unit (GRU), and the Self-Attention for local semantics, temporal semantic features, and weighted semantic features of reviews and concatenated them together to form final model .By splicing together features extracted from different sources, the model may benefit from a richer representation of the input data. Implementing this system may require significant computational resources and may increase the risk of overfitting especially when correct care is not taken

[14] The paper uses Convolutional Neural Networks, Long Short-Term Memory networks, and CNN-LSTM combinations. The CNN and CNN-LSTM models exhibit overfitting yet still capture essential patterns, whereas the LSTM model underfits. The paper suggests using larger or more varied datasets, trying Generative Adversarial Networks (GANs) and Gated Recurrent Units (GRUs), and applying advanced methods like BERT. The paper highlights the potential of traditional machine learning for smaller datasets and advocates for the use of cloud platforms to address computational limitations. Although the models promise, further improvements are necessary for optimal performance.

[15] The paper describes a framework for detecting deceptive reviews using combination of coarse and fine grained features including Latent Dirichlet Allocation (LDA) topic modeling, a 2-layered Backpropagation Neural Network, TextCNN, LSTM, and BiLSTM. It used two balanced and unbalanced datasets from Yelp. The combination of different feature types and advanced techniques helps improve the detection of fake reviews by capturing various patterns in the data. However, managing multiple complex models can be challenging, requiring powerful computers and skilled personnel, which may complicate scaling the system for larger tasks.

[16] The paper analyzes deceptive online reviews from a linguistic perspective using the LIWC (Linguistic Inquiry and Word Count) tool and Negative Binomial Regression, along with categorizing reviewers' motivations. The advantages include easier identification of deceptive reviewers through categorized reviews and a straightforward mechanism that relies on specific language features. However, the study has limitations, such as not accounting for reviews created by bots and needing additional factors for accurate detection. Additionally, the focus on the presence rather than the order of words may overlook important contextual information.

[17] Fast Detection of Deceptive Reviews by Combining the Time Series and Machine Learning works by focusing on suspected time intervals. They captured suspected time intervals and for

each interval a co-training two-view semi supervised learning algorithm was performed to classify the nature of reviews based on linguistic cues, metadata, and user purchase behaviors. This model offers to work with large unlabeled dataset which can improve the learning performance. However, the authors have used limited data in the experiments due to inaccessibility of some data on Taobao website.

[18] The paper, Detection of fake opinions using time series eliminates costly comparisons by using time series models. This method only needs basic information that's available on most review websites, so it can be used in many different places. Additionally, the model addresses spam reviewers' motivations, though this aspect may still have limitations.

[19] This model aims to distinguishes human-written texts from bot-generated texts without prior knowledge about the bot by using K-means and Wishart clustering algorithms. The GPT3 for bot text and literacy books for human text is used with word2vec which creates numerical representations. The method effectively identifies various bot types, from simple RNNs to advanced GPT bots, and requires minimal labeled data. However, it relies on specific features such as word order and entropy, which may not be effective against bots that manipulate other aspects of text. Additionally, the method lacks transparency, making it challenging to understand the reasons behind the classification of certain texts as bot generated.

[20] AI vs. Human Differentiation Analysis of Scientific Content Generation developed a framework and used logistic regression to detect AI-generated texts, achieving high accuracy. It found that AI texts often lack depth and insight but still align well with real scientific knowledge. As NLG models get better, focusing on meaning and context will be key for detection. The method is effective and understandable but requires significant resources and may struggle with advanced text features. It is not accurate and needed more research to accomplish the goals.

[21] AI generated review Detection paper is generated in 2023 by using algorithms like -K Nearest Neighbours, Logistic Regression, Random Forest, SVM, Multilayer perceptron, AdaBoost, BERT. This provides high performance and robustness. It offers clear transparency and interpretability in its processes that enhances its effectiveness. However, the approach is resource-intensive, requiring significant computational power and time. Additionally, it shows limited contribution from large language models (LLMs), which might constrain its applicability in contexts where LLMs could provide substantial benefits.

A. Theoretical Background

Consumers rely heavily on reviews when deciding to purchase a product or service. Studies show that 80% of consumers will change their minds due to negative reviews, and 87% will make a purchase based on positive reviews. This makes consumers vulnerable to misleading online reviews.

Detecting fake reviews involves classifying them as either fraudulent or truthful. Traditional methods can easily identify some types of spam manually, but untruthful reviews are challenging to detect because they often look like genuine reviews. Automated detection is necessary for these cases.

Deceptive opinion spam refers to fictitious reviews designed to appear authentic. Some deceptive reviews are more harmful than others, particularly negative reviews that can significantly impact a business. Therefore, it's crucial to focus on detecting these harmful reviews.

B. Applied Methodologies

Several approaches have been explored to detect fake reviews. Traditional methods involve identifying duplicate reviews, which are considered fake if they appear multiple times across different user IDs or products (Jindal & Liu, 2008).

In the field of research, the most used classifiers are Support Vector Machines (SVM), Naive Bayes, Decision Trees, Random Forests, and Logistic Regression. Ensemble methods and neural networks have also been explored, but less frequently. Additionally, some researchers have proposed using semi-supervised learning and combining textual and behavioral features to improve detection accuracy.

Previous research has mainly focused on the hospitality industry, such as restaurants and hotels, and some on e-commerce. Existing algorithms are often domain-specific, meaning they work well for the language and context of the domain they were trained in but may not perform as well in other areas.

Quantitative data like ratings have been the focus of much research because it's easier for machines to process. However, actual consumers prefer detailed textual reviews, which provide richer information. Qualitative reviews can offer insights for service providers to improve operations and meet customer expectations better.

There has also been research on fake news detection, which has some similarities with fake review detection. Fake news tends to have more

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

text, providing more features for training models. Methods like Glove word embeddings have been useful for preprocessing text into numerical tokens that preserve semantic meaning, improving model accuracy

Transfer learning, a technique where knowledge gained from solving one problem is applied to a related problem, has also been explored. This approach helps overcome the issue of limited data by leveraging large datasets used by other organizations, leading to more reliable outcomes

While the methods used for detecting fake news and fake reviews differ, it might be beneficial to explore some techniques from fake news detection for application in fake review detection. Using pretrained word embeddings could be one such method to enhance the accuracy of fake review detection.

C. Proposed Approach

For review prediction model gather a labeled dataset with balances data , clean up dataset using explorative data analysis techniques on whole dataset , drop irrelevant columns or biased columns then finally perform explorative data analysis on review text by finding total wors, punctuations, stopwords, lowercase , uppercase words and finally preprocess by Spelling is corrected, tokenization, removing stopwords, punctuations, special characters, Lowercasing, Stemming, removing top 3 common and rare words. Vectorize using count vectorization methos, TFDIF method and apply Machine learning techniques, namely SVM, Logistic regression, Multinominal Naye Bayes.

For Source Identification model gather a labeled, balanced dataset, clean up data set using explorative data analysis like review prediction model. After cleaning data and stemmeries using potter stemmer, lemmatize using wordnetlemmatiser. Use pipeline to streamline various process for faster and smooth execution use Count vectorizer and TFIDF vectorizer and K Nearest Neighbors, Decision Tree Classifier, Random Forest Classifier, Support Vector Classifier, Multinomial Naive Bayes, Logistic Regression Machine learning techniques to generate model

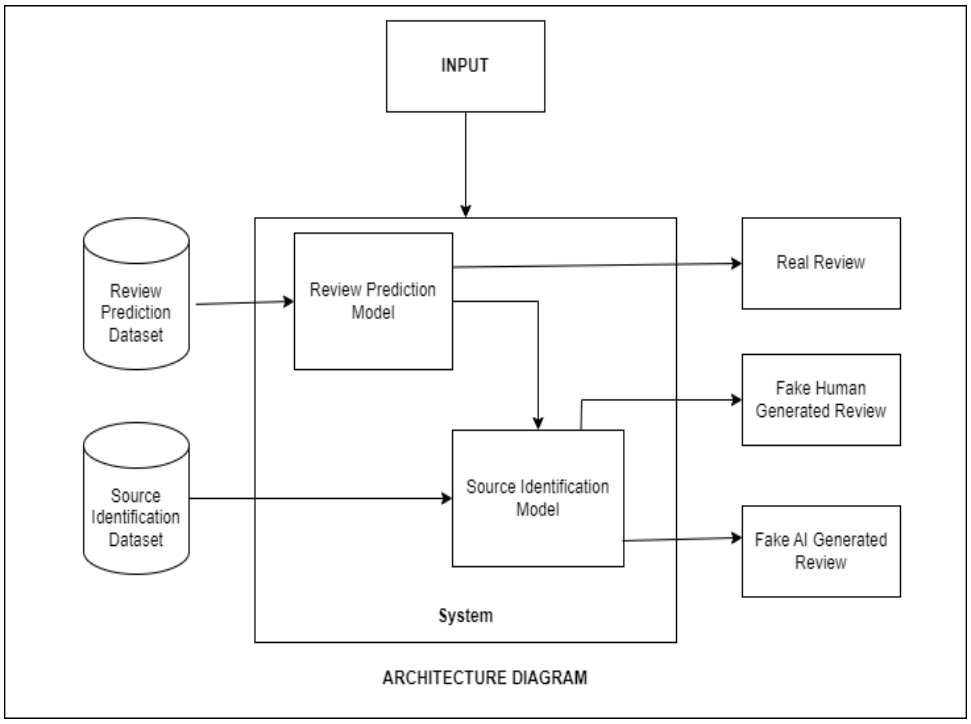
III. METHODOLOGY

This research applies machine learning and natural language processing (NLP) techniques to analyze the textual components of consumer reviews, aiming to extract valuable insights from the language used. The analysis focuses primarily on text, as it provides rich semantic and contextual information that can enhance understanding beyond numerical ratings. By exploring how text tokenization impacts the predictive capabilities of various classifiers, the study aims to develop a model of deception detection in reviews. The methodology involves a comprehensive approach to data cleaning, exploration, and preprocessing to prepare the text data for analysis and model building.

The initial phase of the research involves thoroughly understanding the dataset, which includes columns such as review_rating, review_date, review_title, review_text, and verified_purchase. Unwanted rows are removed, null values are handled, and duplicates are eliminated to ensure the dataset is clean and reliable. Exploratory Data Analysis (EDA) is performed on textual data to assess various aspects, such as word count, character count (including spaces), stopword usage, punctuation, and uppercase characters. This analysis provides insights into the structure and patterns within the text, informing further preprocessing steps.

At last, we have selected Logistic Regression, SVM and Multinomial Naive Bayes to predict the review as fake or not. And to classify the fake review as AI generated or Human generated, we have selected Logistic Regression, SVM, Decision Tree, Random Forest, K -Nearest Neighbors and Multinomial Naive Bayes algorithms.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252



ARCHITECTURE DIAGRAM

The architecture diagram represents a sophisticated system designed to classify reviews into three categories Real Review, Fake Human Generated Review, and Fake AI Generated Review. At the core of this system are two datasets and two models. The Review Prediction Dataset and Source Identification Dataset provide the necessary data for the models to make accurate predictions. The Review Prediction Model utilizes the Review Prediction Dataset to assess whether an input review is genuine or fake. If the review is classified as real, it is output directly as a Real Review.

If the review is fake, the Source Identification Model takes over to determine whether the fake review was generated by a human or an AI. This model relies on the Source Identification Dataset to make this classification. The system flow begins with an input review that is processed by the Review Prediction Model. If the review is genuine, it is classified accordingly. However, if the review is fake, the Source Identification Model steps in to identify its origin, categorizing it as either a Fake Human Generated Review or a Fake AI Generated Review.

In summary, this architecture is designed to efficiently and accurately classify reviews by leveraging two specialized models and datasets. By focusing on both the authenticity of the review and the source of fake reviews, the system provides a comprehensive solution for identifying and categorizing reviews. This dual-model approach allows for a nuanced understanding of review authenticity, ensuring that businesses and consumers can trust the information they receive.

For the first model from the above models, we have chosen Logistic Regression for predicting fake review or not since it gave highest accuracy as 85% and for the second model, we have chosen SVM as it showed highest accuracy of 88%.

Data Understanding

In the process of training and testing the model we considered two different data sets; one for review prediction and the other for identifying the source of generation of the review. The following are the two different datasets used.

Review prediction model dataset

As the below image depicts there are 32 columns more than 2501 rows in the dataset we used to train and test the review prediction model.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

```
Data columns (total 32 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   report_date             2501 non-null   object
 1   online_store            2501 non-null   object
 2   upc                     2501 non-null   float64
 3   retailer_product_code   2501 non-null   object
 4   brand                   2501 non-null   object
 5   category                2501 non-null   object
 6   sub_category            2501 non-null   object
 7   product_description     2501 non-null   object
 8   review_date             2501 non-null   object
 9   review_rating           2501 non-null   int64
 10  review_title            2396 non-null   object
 11  review_text             2501 non-null   object
 12  is_competitor           2501 non-null   int64
 13  manufacturer            2501 non-null   object
 14  market                  2501 non-null   object
 15  matched_keywords        0 non-null      float64
 16  time_of_publication     0 non-null      float64
 17  url                     1654 non-null   object
 18  review_type             2501 non-null   object
 19  parent_review           2501 non-null   object
 20  manufacturers_response  0 non-null      float64
 21  dimension1              2501 non-null   object
 22  dimension2              2501 non-null   object
 23  dimension3              2310 non-null   object
 24  dimension4              0 non-null      float64
 25  dimension5              0 non-null      float64
 26  dimension6              0 non-null      float64
 27  dimension7              2499 non-null   object
 28  dimension8              2501 non-null   object
 29  verified_purchase       2501 non-null   bool
 30  helpful_review_count    2501 non-null   int64
 31  review_hash_id          2501 non-null   object
dtypes: bool(1), float64(7), int64(3), object(21)
```

Among all the data collected about various reviews which are classified either as true or false, there are 55.7% reviews which are true reviews, and the remaining 44.3% reviews are fake reviews.

Generation source data set

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

The dataset that we used for training and testing the generation source detection model is a huge dataset containing 40433 rows and four columns of data. In this dataset the data is labelled either as "CG" or "OR", where CG represents computer generated texts which are AI generated and OR represents original reviews which are human generated.

	category	rating	label	text_
0	Home_and_Kitchen_5	5.0	CG	Love this! Well made, sturdy, and very comfor...
1	Home_and_Kitchen_5	5.0	CG	love it, a great upgrade from the original. I...
2	Home_and_Kitchen_5	5.0	CG	This pillow saved my back. I love the look and...
3	Home_and_Kitchen_5	1.0	CG	Missing information on how to use it, but it i...
4	Home_and_Kitchen_5	5.0	CG	Very nice set. Good quality. We have had the s...

Among the various kinds of reviews collected for training and testing the model there are 20215 reviews labelled as "CG" representing computer generated reviews and the remaining 20216 reviews are labelled as "OR" representing the human generated reviews.

label	count
CG	20215.0
OR	20216.0

Data Preprocessing

Text preprocessing is a crucial step that involves several transformations to prepare the data for modeling. Spelling correction is applied to standardized word usage, while tokenization breaks down the text into individual words or tokens. Stopwords, punctuation, and special characters are removed to focus on meaningful content. Text is converted to lowercase to ensure uniformity, and stemming is used to reduce words to their base forms. Additionally, the most common and rare words are identified and removed, allowing the analysis to focus on words that provide the most significant insights into deceptive reviews. Following preprocessing, machine learning models are trained using the cleaned and processed text data.

IV. ANALYSIS OF EXISTING TEXTS

A. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process, it helped us to understand more about the data that we have used to build the model. It is useful in Understanding Data Structure and Characteristics, Identifying Patterns and Relationships, Detecting Anomalies and Outliers, Data Cleaning and Preparation, Guiding Model Selection and Feature Engineering. We have selected the best features from the dataset by performing feature engineering with the help of data visualization techniques.

B. Evaluation

Transformation of text into numerical representation is crucial for identifying deceptive reviews. For this purpose, we have used two natural language processing methods TF-IDF and Count Vectorization. TF-IDF transforms text data into numerical representations by assessing the importance of words within documents relative to a corpus, using metrics like Term Frequency (TF) and Inverse Document Frequency (IDF). his results in a TF-IDF score for each term that highlights rare but significant words. Whereas Count vectorization is a simpler method that counts the frequency of each word, creating feature vectors based on these counts. It performs well with methods that can handle high dimensional data.

The performance of both methods is evaluated using metrics such as accuracy, precision, recall, F1-score, and Area Under the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

Curve. After Extracting the features using Count vectorization and TF-IDF vectorization, model is trained. These steps are performed on different datasets where one is used to detect whether the reviews are real or fake then on other datasets to identify whether the fake review is human generated, or computer generated.

C. Modeling

Cleaning of data will give us the data ready for model training and building. After cleaned data is produced, we started testing models for fake or real review prediction using various techniques like Logistic regression, SVM, Multinomial NB. Other than these models we have also used models like Decision trees, Random forests, KNN for AI/human generated review prediction. The models tested are as follows:

Decision Trees:

The decision trees are like flowcharts where each branch in the tree represents a rule of the decision making, internal nodes relate to attributes whereas leaf nodes represent classes. The algorithm then divides the data sets based on some criterion. In other words, the Decision tree algorithm employs Attributes Selection Measures such as ID3, Gini index and Gain ratio to select the attribute for splitting the data.

Basic algorithm for decision trees :

1. start with whole training set
2. select attribute or feature satisfying criteria that results in the “best” partition.
3. create child nodes based on partition.
4. Repeat process on each child using child data until a stopping criterion is reached.

Random Forests:

Random forest can be explained as an ensemble learning method which involves combining several decision trees for improving performance and robustness. It is simply many decision trees that were trained on random subsets of both training data and features. When being used for classification, Random Forest predicts class with maximum votes among all individual trees. For regression, it predicts the average outputs from all trees.

Basic algorithm for random forests:

1. Randomly select K features from total m features where k << m
2. Among the K features, calculate the node “d” using the best split point
3. Split the node into daughter nodes using the best split
4. Repeat the 1 to 3 steps until 1 number of nodes has been reached
5. Build forest by iterating steps 1 to 4 for n number of times thereby creating n number of trees

K-Nearest Neighbors (KNN):

KNN is a method that predicts the output using its k-nearest input values. In case of classification, it considers the closest neighbors to the input and returns the most occurring class among them while for Regression, it computes average or weighted average of target values of those k-nearest neighbors to give out a prediction. This technique depends on computing distances between data points to look for close ones.

Basic algorithm for KNN:

1. Select optimal k value
2. Calculate the distance between each data point and sample input.
3. Find the nearest neighbors.
4. Classify the new instance based on the majority class among its K nearest neighbors or by computing average or weighted average of the target values of the k-nearest neighbors to make the prediction.

Logistic Regression:

Logistic regression is a kind of supervised machine learning algorithm that predicts the probability of a binary outcome, event or observation. It provides a binary or dichotomous output with two possible results: yes/no, 0/1, true/false.

Basic algorithm for Logistic regression:

1. Select K features randomly from m total features.
2. Initialize weights and bias parameters.
3. Train the model using the selected features to predict the probability of each class.
4. Iteratively adjust model parameters to minimize logistic loss function.
5. Perform steps 1 through 4 n times to create forest consisting of n different Logistic Regression models with varying feature subsets.

Support Vector Machines (SVM):

SVMs determine a boundary that is optimal for separating the different classes in data. The hyperplane or surface of separation, as it’s sometimes called, is chosen in such a way as to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

maximize the margin between the nearest points of two separate classes. Thus, making this model robust and accurate.

Basic algorithm for SVM:

1. Randomly select K features from a total of m features.
2. Use the selected features to calculate the optimal hyperplane using a chosen kernel function and regularization parameter C.
3. Split nodes into daughter nodes based on the best split point until l nodes are reached.
4. Repeat steps 1 to 3 for n iterations to build a forest of n SVMs with varied feature subsets.

Multinomial Naive Bayes:

Basic algorithm for Multinomial NB:

Multinomial Naive Bayes has been widely used for text classification. It assumes independence between all words or other features (like words) in a document. Thus, given input characteristics, it estimates probabilities of each class and assigns a class with highest probability.

1. Choose randomly K features from m feature set totally.
2. Based on selected attribute independence assumption estimate class odds.
3. Each instance gets classified by maximum probability among classes calculated before.
4. Repeat steps 1 to 3 for n iterations to build a forest of n MNB models with varied feature subsets.

At last, we came to find that for fake or real review prediction logistic regression mode gives the best accurate results and for AI/human generated review detection SVM model gives the best results.

V. RESULT AND DISCUSSIONS

To find out if a review is fake, we used three models: Logistic Regression, Naive Bayes, and Support Vector Machine (SVM). We used the Amazon dataset and a fake review dataset for training and testing. The data preprocessing involved steps like lemmatization, stemming, stop word elimination, punctuation removal, and converting text to lowercase.

Based on the accuracy and F1 scores, Logistic Regression performed the best among the three models, with an accuracy of 85%. Therefore, we selected Logistic Regression as the preferred model for fake review detection.

For the task of determining whether a fake review is human-generated or computer-generated, we evaluated six machine learning models: Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree Classifier, Random Forest Classifier, and Multinomial Naive Bayes.

The SVM model outperformed the others with an accuracy of 88%, making it the best choice for determining whether a fake review is human-generated or computer-generated.

With these two models we developed a system that will predict whether a review is fake or not and categorize the review as computer generated or human generated.

In the system we finally after combining the models using pickle, first we will be cleaning the given input review using below methods like stemmer.

INPUT REVIEW	OUTPUT
Smells great, easy one-handed application	Real Review
Great quality wipes that are strong to use and don't tear apart like other wipes. A must buy over other brand names and I'll be purchasing more in the near future. Very simple with simple's product line.	Fake Review This review is human generated
Very flimsy for blackout curtains. Not an easy task to put together.	Fake review This review is AI generated

In the above table basically, there are three kinds of reviews given as inputs, where in which each review is classified as real or fake. If in case the review is fake it is further classified as AI generated or human generated.

In the first case the review is classified as a “Real review “by using the logistic model. Hence there is no further requirement to check whether it is AI generated or human generated.

In the second case the review is classified as a “Fake review "after testing it with the logistic

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

regression model. Hence the review is further verified using the Support Vector Machine (SVM) model and it is classified as a human generated review.

In the third case the review is classified as a “Fake review "after testing it with the logistic regression model. Hence the review is further verified using the Support Vector Machine (SVM) model and it is classified as an AI generated review.

```
# Single input text
input_text = ['Smells great, easy one-handed application']
data={"text":input_text}
```

```
if prediction[0] == True:
    print("Real Review")
else:
  print("Fake Review")
  if(prediction1[0]=="CG"):
    print("This review is AI generated")
  else:
    print("This review is human generated")
```

```
Real Review
```

In the first case the review is classified as a “Real review “by using the logistic model. Hence there is no further requirement to check whether it is AI generated or human generated.

```
# Single input text
input_text = ['''Great quality wipes that are strong to use
 and don't tear apart like other wipes. A must buy over other brand names
  and I'll be purchasing more in the near future. Very simple with simple's
            product line.''']
data={"text":input_text}
```

```
if prediction[0] == True:
    print("Real Review")
else:
  print("Fake Review")
  if(prediction1[0]=="CG"):
    print("This review is AI generated")
  else:
    print("This review is human generated")
```

```
Fake Review
This review is human generated
```

In the second case the review is classified as a “Fake review "after testing it with the logistic regression model. Hence the review is further verified using the Support Vector Machine (SVM) model and it is classified as a human generated review.

```
# Single input text
input_text = ['''Very flimsy for blackout curtains. Not an easy task
to put together.''']
data={"text":input_text}
```

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 24-36 www.ijaem.net ISSN: 2395-5252

```
if prediction[0] == True:
    print("Real Review")
else:
  print("Fake Review")
  if(prediction1[0]=="CG"):
    print("This review is AI generated")
  else:
    print("This review is human generated")


```

```
Fake Review
This review is AI generated
```

In the third case the review is classified as a "Fake review "after testing it with the logistic regression model. Hence the review is further verified using the Support Vector Machine (SVM) model and it is classified as an AI generated review.

VI. CONCLUSION

In this project, our aim was to develop a robust framework for detecting and categorizing fake reviews, which are prevalent in online platforms and can significantly influence consumer decisions. We utilized two distinct datasets: one comprising Amazon product reviews and another containing a diverse range of reviews categorized as either computer-generated (CG) or original (OR). Our methodology involved rigorous data preprocessing steps, including exploratory data analysis (EDA), to clean and transform the data. Techniques such as stemming, lemmatization, punctuation removal, and stop-word elimination were crucial in preparing the text data for machine learning model training.

For the primary task of distinguishing between real and fake reviews, we evaluated multiple machine learning algorithms including Logistic Regression, Naive Bayes, and Support Vector Machine (SVM). Logistic Regression emerged as the most effective model, achieving an accuracy of 85%, demonstrating its robust performance in binary classification tasks. This model's capability to predict whether a review is genuine or deceptive provides a foundational layer of trustworthiness in online review systems.

Furthermore, we addressed the challenge of categorizing fake reviews into human-generated or AI-generated categories. Leveraging models such as SVM, Decision Trees, Random Forests, K-Nearest Neighbors (KNN), Multinomial Naive Bayes, and Logistic Regression, SVM outperformed others with an accuracy of 88%. This success highlights SVM's ability to discern subtle differences between human and AI-generated content, crucial for identifying sophisticated deceptive practices.

By integrating these models into a unified system, we enable real-time detection and classification of reviews, thereby safeguarding consumer trust and fostering a more reliable online marketplace.

VII. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

A. Limitations

The models were trained on specific datasets from Amazon and fake review datasets. These models might not work as well on reviews from different platforms or in different contexts.

This system considers reviews with large number of characters as deceptive reviews, and it is one of the factors in decision making but in real time there may be deceptive reviews with a smaller number of characters.

B. Recommendations For Future Works

Ensemble Methods:

Combining multiple models might improve accuracy. Deep learning models offer high accuracy and robustness but require significant computational resources and are complex to maintain. Hybrid approaches aim to combine the best of both worlds, offering comprehensive detection capabilities but at the cost of increased complexity.

Advanced NLP Techniques:

Using advanced models like BERT or GPT could find more detailed patterns in the text.

To create a trustworthy online marketplace, ongoing research and development of innovative fake review detection methods are essential. By continuously improving detection techniques, leveraging new data, and integrating advanced models, the impact of fake reviews can be minimized, thereby enhancing consumer trust and the overall integrity of e-commerce platforms.

REFERENCES

[1] Abhijeet A Rathore, Gayatri L Bhadane, Ankita D Jadhav, Kishor H Dhale, Jayshree D Muley, "Fake Reviews Detection Using NLP Model and Neural Network Model," ISSN: 2278-0181, Vol. 12 Issue 05, May-2023

[2] P. Aishwarya Sri, R. Vamshidhar Reddy, "A Fake Review Detection System Using NLP and Machine Learning Techniques," ISSN 2229-5518, International Journal of Scientific & Engineering Research Volume 12, Issue 8, August-2021.

[3] Ioana-Ruxandra Stăncioi, Ștefan Trăușan-Matu, "Fake reviews detection techniques", posted: 2021.

[4] Ahmed M. Elmogy1 , Usman Tariq 2 , Ammar Mohammed 3, Atef Ibrahim, "Fake Reviews Detection using Supervised Machine Learning," Vol. 12, No. 1, 2021.

[5] Jayalakshmi. L1 , Sneha. S2 , Subha Ilakiya. P3 , Kavi Bhaarathy. DA4 ,Bhavani.N 5, "Fake Review Detection on Using Machine Learning on Online Product Selling Platform," ISSN: 2321-9653, Volume 10 Issue VII July 2022.

[6] Doris Macean, "PREDICTIVE MODELS FOR DETECTING FAKE REVIEWS VIA WORD EMBEDDINGS,"

[7] Ms. Rajshri P. Kashti1, Dr. Prakash S. Prasad2, "Enhancing NLP Techniques for Fake Review Detection," p-ISSN: 2395-0072, Volume: 06 Issue: 02 | Feb 2019.

[8] . Chandaka Babi, M. Sai Roshini, P. Manoj, K. Satish Kumar, "Fake Online Reviews Detection and Analysis Using Bert Model,"

[9] Aishwarya M. Kashid1, Ankita K. Lalwani2, Samiksha S. Gaikwad3, Rajal A. Patil4, R. G. Sonkamble5, S. S. More6, " Fake Review Detection System Using Machine Learning", ISSN (Online): 2581-5792, Volume-1, Issue-12, December-2018.

[10] Dong Zhang a, Wenwen Li b, Baozhuang Niu a, Chong Wu c, "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information", Volume 166, March 2023, 113911.

[11] Anusuya Baby, "Unmasking Falsehoods in Reviews: An Exploration of NLP Techniques", 10.48550/arXiv.2307.10617, July 2023.

[12] Rakibul Hassan, Md. Rabiul Islam, "Detection of fake online reviews using semi-supervised and supervised learning", 10.1109/ECACE.2019.8679186, 04 April 2019.

[13] Ning Cao a 1, Shujuan Ji a 2, Dickson K.W. Chiu b 3, Maoguo Gong a c 4, "A deceptive reviews detection model: Separated training of multi-feature learning and classification", https://doi.org/10.1016/j.eswa.2021.115977, Volume 187, January 2022, 115977.

[14] Lay Acheadeth, "Deceptive review detection using Deep Learning", June 2023.

[15] Ning Cao, Shujuan Ji, Dickson K.W. Chiu, Mingxiang He, Xiaohong Sun, "A deceptive review detection framework: Combination of coarse and fine-grained features," Volume 156, 15 October 2020.

[16] Wen Zhang, Qiang Wang, Jian Li, Zhenzhong Ma, Gokul Bhandari & Rui Peng, "What makes deceptive online reviews? A linguistic analysis perspective", 01 November 2023.

[17] Minjuan Zhong, Zhenjin Li, Shengzong Liu, Bo Yang, "Fast Detection of Deceptive Reviews by Combining the Time Series and Machine Learning", 10.1155/2021/9923374, 2021(6):1-11.

[18] Atefeh Heydari, Mohammadali Tavakoli, Naomie Salim, "Detection of fake opinions using time series", https://doi.org/10.1016/j.eswa.2016.03.020, Volume 58, 1 October 2016.

[19] Vasilii Gromov, Quynh Nhu Dang, "Spot the Bot: Distinguishing Human-Written and Bot-Generated Texts Using Clustering and Information Theory Techniques", posted: 19 Nov 2023.

[20] Yongqiang Ma, Jiawei Liu, Fan Yi, Qikai Cheng, Yong Huang, Wei Lu, Xiaozhong Liu, "AI vs. Human -- Differentiation Analysis of Scientific Content Generation", 24 Jan 2023.

[21] Jiwei Luo, Guofang Nan, Dahui Li, Yong Tan, "AI-Generated Fake Review Detection", 20 Nov 2023.