A PROJECT REPORT

ON

”Automated Fake Review Detection Using Machine Learning ”

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY (SPPU),
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

FINAL YEAR OF ENGINEERING
(Computer Engineering)

SUBMITTED BY

1. Tamanna Maner	**35**
2. Prathamesh Joshi	**29**
3. Atharv Gadekar	**17**
4. Dnyaneshwar Rede	**15**

UNDER THE GUIDANCE OF
Prof.Nilufar Zaman

DEPARTMENT OF COMPUTER ENGINEERING
TRINITY ACADEMY OF ENGINEERING
Kondwa Annex, Pune - 411048.
SEMESTER I [AY: 2023-24]

DEPARTMENT OF COMPUTER ENGINEERING
TRINITY ACADEMY OF ENGINEERING
SEMESTER I [2023-24]

CERTIFICATE

This is to certify that the project work entitled

”Automated Fake Review Detection Using Machine Learning”

Submitted by

1. Tamanna Maner	**35**
2. Prathamesh Joshi	**29**
3. Atharv Gadekar	**17**
4. Dnyaneshwar Rede	**15**

is a bonafide work carried out under the supervision of Prof. Nilufar Zaman and it is submitted towards the partial fulfilment of the requirements of Savitribai Phule Pune University (SPPU), for the award of the degree of Bachlors of Engineering (Computer Engineering).

Date:
Place: Pune

(Prof.Nilufar Zaman) **Project Guide**	**(Dr. M. B. Wagh)** **HOD**	**(Dr. N. J. Uke)** **Principal**

Acknowledgement

I have great pleasure and sense of satisfaction in presenting this report on ”Automated Fake Review Detection” for Project Stage - I. This report would not be possible without the help and support of Prof.Nilufar Zaman. We take great delight in expressing our appreciation and thanks to everyone who contributed to the project’s timely completion. We would like to thank Dr. M. B. Wagh, our Head of Department and mentor, for his unwavering support and inspiration. With great pleasure, we would like to convey our sincere appreciation and debt of gratitude to the faculty in our Department of Computer Engineering as well as to Trinity Academy of Engineering, Pune for their helpful advice, ongoing supervision, and unwavering support during the course of this project. For allowing us to embark on this project, Dr. N. J. Uke, Principal of Trinity Academy of Engineering in Pune, deserves our sincere gratitude.

Tamanna Maner
Prathamesh Joshi
Athrav Gadekar
Dnyaneshwar Rede
BE COMPUTER
Batch: 2024

Abstract

In the age of digital commerce and information sharing, online reviews have become an integral part of the decision-making process for customers and consumers. However, Increasing the growth of fake reviews has raised concerns about the authenticity and reliability of these assessments. This Project presents an overview of the methods and techniques used for the detection of fake reviews, with a focus on the evolving landscape of deceptive practices and the countermeasures developed to combat them.

The abstract encompasses a broad scope of fake review detection, covering various aspects such as text analysis using NLP, and machine learning approaches. It highlights the significance of this research in maintaining the trust and transparency of online review platforms and the impact of fake reviews on businesses and consumers. The project also discusses emerging challenges and future directions in the field of fake review detection, emphasizing the need for ongoing research and innovation to stay ahead of increasingly sophisticated deceptive strategies.

Keywords: NLP, Ml Algorithm, Reviews, Transparency

Contents

List of Figures

1. Introduction

In this online world , online review platforms are widely used by people to share their opinions on products and local businesses. These reviews provide valuable impact into the experiences of others, particularly for items whose quality becomes apparent only after use. Online reviews play a pivotal role in helping consumers make informed choices, whether it's picking a restaurant or buying a product online. Reading product reviews before making a purchase has become a common practice among potential customers. They typically seek out reviews from other customers when considering a product. If the reviews are predominantly positive, it increases the likelihood of them buying the product[1]. Conversely, if the reviews are mostly negative, they tend to look for alternative products. In the realm of online shopping, discrepancies between the information provided about products and the actual products received by consumers have made it necessary for consumers to rely on a substantial number of reviews to make informed decisions.

The rise of fake reviews is a big problem on the internet. People write these fake reviews for different reasons, like making money, trying to look better than others, or beating their competitors. This report is like a journey to understand how we can find these fake reviews. It's an important task in today's connected world where we rely so much on online reviews to make decisions.As a result, product reviews not only influence consumers' purchasing intentions but also impact the interests of businesses. Favorable reviews can attract more potential customers, while unfavorable reviews can deter potential customers. To maximize profits, some unscrupulous merchants resort to hiring professional writers to create fake positive reviews for their own products, thereby increasing their products' popularity to draw in potential customers. Simultaneously, they may craft fake negative reviews about their competitors. The system we are going to design will identified the fake reviews among the reviewers using their features.

Fake reviews erode trust in online platforms and contribute to skepticism among consumers. When the lines between genuine and fabricated feedback blur, it becomes challenging for consumers to make informed decisions. Detecting fake reviews means finding out if the reviews people write online are honest or if they're pretending to be real. It's like having a detective to check if what someone says about a product or service is true or if they're just saying nice or mean things for the wrong reasons. This checking helps us trust what we read and decide better about what to buy or use.

We explore the evolving landscape of online reviews, the motivations behind fake reviews, and the various techniques employed by perpetrators to create deceptive content. Moreover, we present an overview of the methodologies and technologies used to detect fake reviews, including natural language processing (NLP)[1], machine learning algorithms, and data mining approaches. Additionally, we discuss the ethical implications of fake review detection and the need for responsible, transparent, and fair practices in mitigating this issue. These deceptive narratives are carefully crafted to deceive potential buyers,manipulate perceptions, and bolster the reputation of a business or product. Fake reviews come in various forms, ranging from overly positive endorsements designed to inflate a product's rating to maliciously negative critiques aimed at tarnishing a competitor's image. The consequences of fake reviews are profound, as they not only mislead consumers but also erode trust in the entire review system, ultimately harming businesses and undermining the authenticity of the online marketplace.

As we delve deeper into the world of fake review detection, this report aims to shed light on the complexities of the problem and provide valuable insights into the ever-evolving battle against deceptive practices in the online review landscape. By the end of this report, readers will have a comprehensive understanding of the challenges posed by fake reviews and the strategies available for tackling this pervasive issue, thereby safeguarding the trust and reliability of online review platforms for consumers and businesses alike.

1.0.1 Motivation

In the process of developing a model for detecting fake reviews using features derived from deception theories, feature engineering plays a crucial role. It involves the identification, manipulation, selection, and extraction of the most valuable features from raw data, ultimately simplifying the model and leading to improved results in identifying deceptive behavior within fake reviews With the increasing transition of businesses to online platforms in recent years, there has been a growing emphasis on the development of precise algorithms for identifying online fake reviews. This shift is vital to safeguard businesses and individuals from the harmful effects of misinformation and spam. Reviews come in various forms, each with its own unique impact on people's emotions and attraction towards products and services.

- Positive reviews: Positive reviews typically generate feelings of trust and excitement, increasing attraction.
- Negative reviews: Negative reviews may initially lower attraction due to disappointment or skepticism, but can lead to long-term improvements.
- Neutral reviews: Neutral reviews offer balanced information without strong emotional sway, and detailed reviews are trusted for their depth and can enhance attraction.• Emotional reviews: Emotional reviews, whether positive or negative, can attract readers through their authenticity, while expert reviews are valued for their authority.
- User-generated reviews: User-generated reviews create a sense of community and reliability, fostering attraction.

However, fake reviews can lead to skepticism and decreased attraction, and long-term user reviews can bolster attraction by showcasing sustained satisfaction over time. Transparent, honest, and informative reviews tend to have a more positive influence on attraction and decision-making. reviews. Considering behavioral features is highly expected to enhance the performance of the presented fake review detection approach. Also, using neural network models to perform this task would be equally beneficial for detecting fake reviews from large datasets. Detecting fraudulent online reviews primarily constitutes a binary classification problem. Numerous researchers are actively engaged in developing methods for automatically identifying fake reviews through the application of diverse machine learning techniques. The majority of these studies are rooted in supervised learning, with a few also exploring semi-supervised and clustering approaches. In general, these investigations predominantly revolve around two primary aspects: the reviewers and the reviews themselves.

2. LITERATURE SURVEY

This section deals with the work done already in the Online fake review detection area. In Past few years the Online fake review detection system has gained popularity due to increase in the use of e-commerce websites. The aim of this study is to gain an understanding of the diverse models and algorithms employed in the detection of fraudulent or deceptive reviews concerning various.

In [2], We established an initial index system that comprises multiple features, including textual aspects, sentiment attributes of reviews, and reviewer behavior traits. This framework serves to provide a comprehensive overview of the process[2]. Additionally, we created corresponding algorithms to extract each feature from the reviews. Following this, reviews were manually categorized. Subsequently, we employed seven classifiers, utilizing the initial sample set generated by the algorithm, and selected the most accurate classifier to classify new reviews.One notable innovation in our method is the continuous expansion of the sample set. New reviews' features and categorization labels are added to the original sample set as fresh samples, allowing it to grow naturally. Our proposed approach for detecting fraudulent reviews achieved an accuracy rate of 84.45, which is 3.5 percentage points higher than the baseline methods, as demonstrated by experimental results on reviews from the Yelp shopping website. Furthermore, the baseline precision increased by 5.3 percentage points compared to the latest deep learning model. Statistically, [2]the Friedman test indicates that the Support Vector Machine (SVM) and Random Forest (RF) classifiers are the most effective.

In [3],gives a summary of the Covid-19 outbreak and how lockdowns, social distancing, and other preventative measures caused a spike in online buying all around the world. The competition amongst organisations for online selling has intensified due to the growing importance of online buying and the widespread use of e-commerce. emphasises the importance of internet reviews in promoting or defaming a company. Because product reviews play a crucial role in consumers' decision-making, the detection of fraudulent or phoney reviews has become a hot topic. In this study, we used machine learning and text classification methods to propose a fake review detection model. Support Vector Machine, K-Nearest Neighbour, and logistic regression (SKL)[3] are some of the classifiers we used. A bigram model that identifies fraudulent reviews based on the quantity of pronouns, verbs, and sentiments was employed. Using the Yelp and TripAdvisor datasets, our suggested methodology for identifying fraudulent online reviews performs better than other cutting-edge methods with accuracy scores[3] of 95 and 89.03, respectively.

In [1]Propose Online reviews that are fake share a few traits. They are initially defined as reviews posted online by people based only on their fantasies and lack of experience. The second essential feature of fake reviews is their capacity to deceive customers. Thirdly, there are two primary methods for creating fraudulent reviews: computer-generated and human-generated. Fourth, phoney reviews can be posted by various online retailers, platforms, or customer types. Fifth, phoney reviews originate from various cultures and speak multiple languages. Sixth, people cannot tell the difference between phoney and real reviews just by reading the text of online reviews. To create a cohesive theoretical framework[1], we extracted significant deception constructs from the various deception theories. We chose characteristics that could be measured from the review texts and the reviewer's actions and could be used to describe the derived constructs. After extracting the chosen features, our fake reviews detection model was empirically validated using three well-known Yelp review datasets.

a comprehensive model was developed to identify fraudulent online reviews. The authors meticulously selected various criteria, encompassing specificity, quantity, non-immediacy, affect, uncertainty, informality, consistency, source credibility, and deviation in behavior. These constructs were em- ployed, utilizing both verbal and non-verbal attributes, to empirically validate the proposed model. For the empirical validation, data was extracted from Yelp datasets, and subsequently, four machine learning algorithms—Logistic Regression, Na¨ıve Bayes, Decision Tree, and Random Forest—were trained using the extracted features. Interestingly, [1]findings indicated the greater significance of non-verbal features over verbal ones. Moreover, the amalgamation of features from both categories led to enhanced pre- dictive performance. Significantly, the theory-based model developed in this research surpassed many existing fake review detection models in terms of performance. Notably, it exhibited a high degree of interpretability and retained a low level of complexity. The evaluation of machine-learning classifiers commonly relies on metrics such as the F1-score (F1), accuracy (A), recall (R), and precision (P). Precision (P) measures the proportion of correctly classified fake reviews among all reviews labeled as fake. It assesses the classifier's ability to avoid misclassifying genuine reviews as fake. Recall (R) quantifies the ratio of correctly classified fake reviews out of all the fake reviews present in the dataset, indicating the classifier's effectiveness in identifying all fake reviews. The F1-score (F1) represents the harmonic mean of recall and precision. Accuracy (A) calculates the ratio of reviews correctly classified as either truthful or fake among all the reviews assessed.

In [4]Explains Firstly, the results of our study are not entirely generalizable, necessitating additional research and follow-up studies. Second, the application of the presumably cutting-edge Amazon fake review detection mechanism defines the dataset's labelling. Finally, because certain datasets are missing, we are unable to provide a cross-domain check, such as identifying phoney news or phoney stores, to validate the results. Their results indicate that this system's accuracy is 0.7313. the limitation of [4] is There are some limitations and a need for further research. First, our study does not provide fully generalizable results, thus requiring further investigation and follow-up studies. Second, the labeling of the dataset is defined by the application of the amazon fake review detection mechanism, which is assumed to be state-of-the-art. Last, we lack to deliver a cross-domain check like to identify fake news or fake shops to validate the results, due to missing datasets.

In [5], E-commerce websites contain real and fake reviews of both kinds. The accuracy level of the Amazon dataset for fake reviews is 83.35 1 when using the machine learning open AI model.Just two top k nucleus sampling algorithms are used. In this Paper, Methodlogy uses six main steps:

1. Generate example reviews using two distinct language models.
2. Assess the generated example reviews using both quantitative measures and qualitative evaluations.
3. Decide on the superior language model for the development of a fabricated review dataset, which incorporates authentic human-authored reviews.
4. Train machine learning classifiers to distinguish artificially generated reviews from genuine ones.
5. Employ a team of crowd workers to annotate a subset of the original and fabricated reviews.
6. Contrast the accuracy of crowd workers with that of the classification algorithms using statistical tests. Our methodology revolves around synthetic review generation, which involves the creation of fake reviews based on existing genuine reviews.

Linguistic features encompass a spectrum that includes basic methods like word or n-gram frequency counting to more advanced techniques that depend on distributional semantics. For algorithms to be more widely applicable, they must be trained to represent each language. Social media reviews in general as well as reviews of e-commerce products in particular. The search algorithms indicate that the accuracy is 96.64.

In this[6], we present the results of binary classification achieved through the application of machine and deep learning techniques, along with the introduction of a novel dataset of fraudulent reviews. Additionally, we employed various methods, including bagging, Random Forest, Support Vector Machines (SVMs), and Multi-Layer Perceptrons (MLPs), each optimized with hyperparameters, as well as other ensemble learning-based approaches. Our findings illustrate that individual classifiers can achieve an accuracy of up to 68.2 in the case of MLPs, utilizing document embedding from Doc2Vec and hyperparameter optimization[6]. However, when employing an ensemble of MLPs, ensemble learning-based classifiers can reach an accuracy of up to 77.3 .

In our research [7], we delved into the significance of 15 linguistic features for the categorization of written reviews into either fake or reliable. Employing methods such as RF feature importance, Recursive Feature Elimination (RFE), Boruta, and Analysis of Variance (ANOVA), we consistently identified the number of adjectives, redundancy, and pausality as the most crucial features in this classification task. These features were used in [7] classification experiments with seven different classifiers, and it was the Multi-Layer Perceptron (MLP) that demonstrated the highest accuracy, achieving 79.09 accuracy while utilizing only four features.

The authors introduce in [8], a sentiment analysis model that exhibits a remarkable capacity for extracting sentiment with a high degree of accuracy. The model is employed to delineate the unique characteristics of sentiment scores in both fake and genuine reviews. Furthermore, the researchers undertake a comprehensive investigation to assess the impact of utilizing probabilistic sentiment scores derived from their model in the development of a counterfeit review detection classifier. The study [**?**]presents a comparative analysis of the classification results obtained from the probabilistic sentiment scores generated by their sentiment model and the binary sentiment scores present in the dataset. The dataset utilized for proposed sentiment analysis model and fake review detection model comprises 1600 hotel reviews, equally divided between positive and negative sentiments. Additionally, the dataset is tagged with labels for deceptive and truthful reviews. This dataset is characterized by its balanced composition, containing an equal number of deceptive and truthful reviews. Notably, the dataset exclusively features the review text and omits any user information or rating scores.

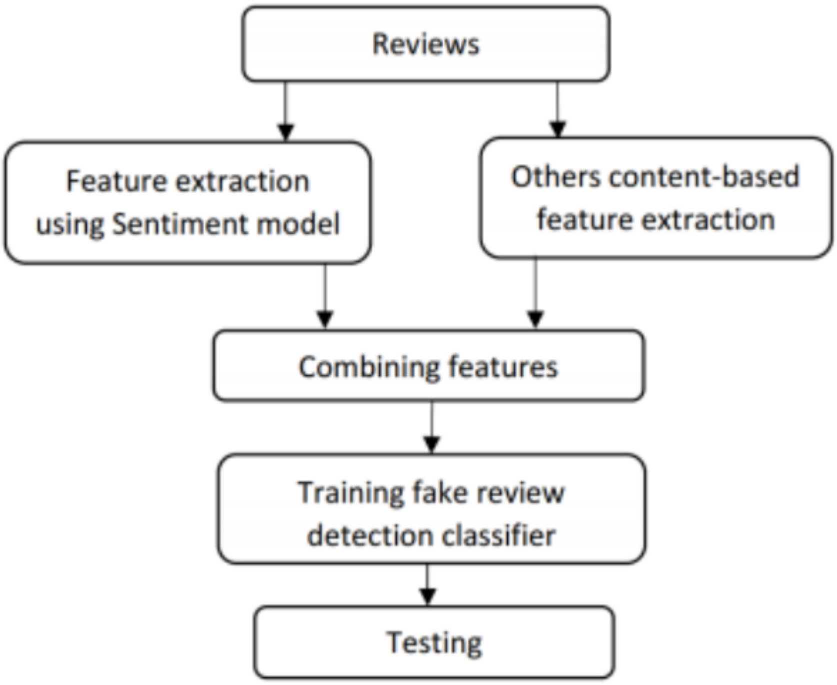

Reviews
Feature extraction using Sentiment model
Others content-based feature extraction
Combining features
Training fake review detection classifier
Testing

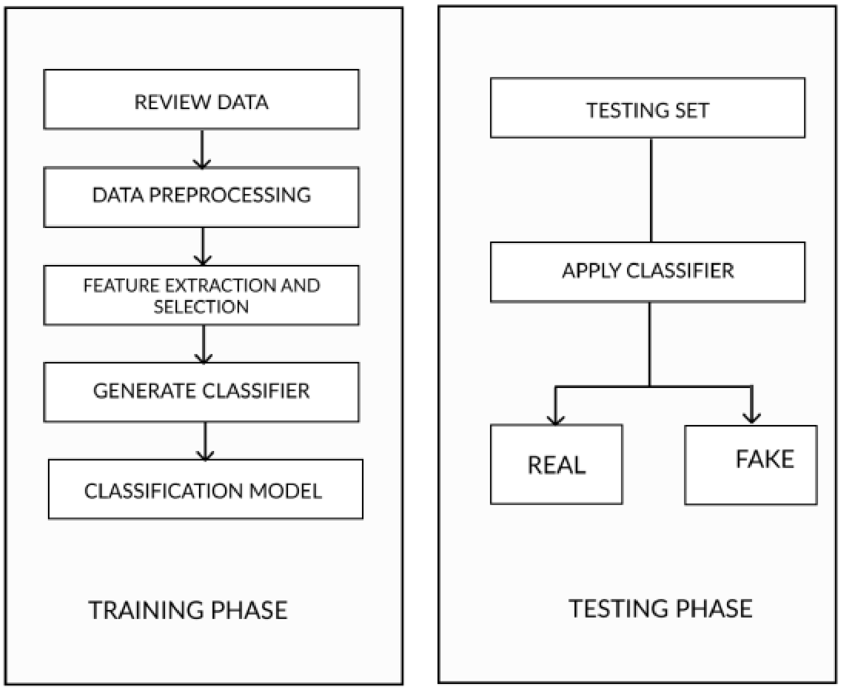
Figure 2.1: Proposed fake review classification model

in this [9] a supervised approach is adopted, leveraging Yelp's filtered reviews for training. It's worth not- ing that existing supervised learning approaches primarily rely on pseudo fake reviews rather than those filtered by a commercial website. Recent studies have demonstrated the high performance of supervised learning using linguistic n-gram features, achieving approximately 90percent accuracy in the detection of crowdsourced fake reviews generated through Amazon Mechanical Turk (AMT). These established research methods are subjected to testing and evaluation on real-life Yelp data. Surprisingly, while be- havioral features exhibit strong performance, linguistic features prove to be less effective. To explore this discrepancy, a novel information-theoretic analysis is introduced to uncover the specific psycho lin- guistic distinctions between AMT reviews and Yelp reviews, essentially differentiating crowd sourced from commercial fake reviews. The researchers uncover an intriguing observation. Their analysis and experimental findings lead to the hypothesis that Yelp's filtering is justifiable, and its filtering algorithm appears to have a connection with unusual spamming behaviors. The paper investigated fake re- views on Yelp.com. Linguistic methods that worked well for crowd sourced fake reviews showed lower effectiveness for real commercial Yelp reviews. In contrast off [9], behavioral features indicated Yelp's filter relies on behavioral patterns. In it confirmed Yelp's filtering reliability and highlighted that linguistic features are less effective in real-life scenarios. Crowd sourced fake reviews may not reflect real-life fake reviews accurately.

This study [10] concentrates on evaluating the effectiveness of ensemble learning within the realm of detecting review spam. It makes use of specific features extracted from actual and semi-real-life datasets, and it assesses various performance metrics, including Precision, Recall, F-Measure, and the Receiver Operating Characteristic (RoC). The research introduces an ensemble learning module (ELM) in conjunction with the Chi-Squared feature selection technique, surpassing all other methods with a Precision score of 0.851. Ensemble learning is a machine learning approach that combines predictions from multiple individual models (learners) to produce a more accurate and robust final predictions of multiple individual models (learners) to create a more accurate and robust final prediction. The proposed Ensemble Learning Module (ELM)[10] is structured into two tiers. In the first tier, three classifiers are employed: Decompos- able Naive Bayes (DMNB), J48 decision tree, and LibSVM support vector machine. DMNB is based on Naive Bayes and suited for text classification, while J48 constructs decision trees for classification, and LibSVM employs support vector machines. In the second tier, the meta-classifier, Logistic Regression (LR), refines the predictions made by the Tier 1 classifiers. This multi-tiered approach leverages the strengths of various classifiers and the combining power of LR to enhance the overall classification per- formance, particularly in the context of review spam detection with selected features. In Experiment Setting , the process involves selecting reviews, preprocessing data, utilizing full feature sets, dividing the dataset into training and testing subsets through 10-fold cross-validation, training the classifier, evaluating model performance, and recording classification results. Their proposed ELM outperforms others by 0.842 precision, 0.834 recall, 0.832 F-measure and 0.908 RoC. One limitation of this study pertains to the utilization of imbalanced datasets, as in practical scenarios, spam reviews are typically fewer in number than genuine ones, resulting in dataset imbalances.

In their research paper[11], the authors delve into the realm of machine learning techniques(such as supervised learning, unsupervised learning and semi-supervised learning techniques) for the identifi- cation of fake reviews. They categorize these techniques based on behavioral, linguistic, and rela- tional features in order to discern the authenticity of reviews.Linguistic features play a crucial role in the identification of fake reviews, as they rely on distinctive writing styles and language patterns. These lin- guistic and textual features encompass N-gram characteristics, parts-of-speech (POS) attributes, features derived from Linguistic Inquiry and Word Count (LIWC), and stylistic elements. The paper [11]surveys pre- vious methodologies used to detect fake reviews, taking into account the types of data employed, which encompass labeled data (such as supervised learning), unlabeled data (as seen in unsupervised learning),and partially labeled data (as exemplified in semi-supervised learning). The authors provide an overview of the existing work in fake review detection, primarily focusing on machine learning-based techniques. They emphasize the significance of key attributes like features and classifiers in this context, while also addressing the major challenges inherent in fake review detection. Overall, the paper comprehensively explores the landscape of fake review detection through machine learning, encompassing various ap- proaches and challenges. Some of the mentioned challenges in different fake review detection are as below:

1. Characteristics such as ratings and references to brand names can be challenging for both humans and machines to discern.
2. It becomes challenging to detect rating patterns when there is only a single review available for a specific product.
3. When counterfeit reviews are intentionally crafted to mimic authentic ones, distinguishing between genuine and fake reviews becomes a formidable task



[11]

Figure 2.2: Phrases of ML

In this study [12], the authors have introduced the utilization of the BERT (Bidirectional Encoder Representation from Transformers) model to extract word embeddings from textual data, particularly reviews. This implementation is aimed at enhancing the accuracy of existing methods for classifying or detecting fake reviews. Word embeddings are derived through several fundamental techniques, including Support Vector Machines (SVM), Random Forests, Naive Bayes, and other methodologies. NLP[12] is used to create some review features which are not directly related to data or text provided. We used BERT (Bidirectional Encoder Representation from Transformers) to perform NLP tasks and other text processing such as feature selection.The dataset considered is the gold standard mock review data corpuses. The pro- posed system approach in this is:

1. The dataset is loaded into a BERT (Bidirectional Encoder Representation from Transformers) to generate word embeddings, which are large vector representations of the text words in the dataset.

2. The input is then loaded into the classification models for their training. Training and testing data are divided in the ratio: 80:20, i.e.: 80percent for training and the rest for testing the model.

3. Results are evaluated using a confusion matrix representation for Precision, Recall, F1score and Accuracy.

4. The best performing classifier is then saved and later used to detect user reviews as fake or real. The results indicate that the SVM classifiers outperforms the others in terms of accuracy and f1-score with an accuracy of 87.81percent and f1-score of 0.88 [12]

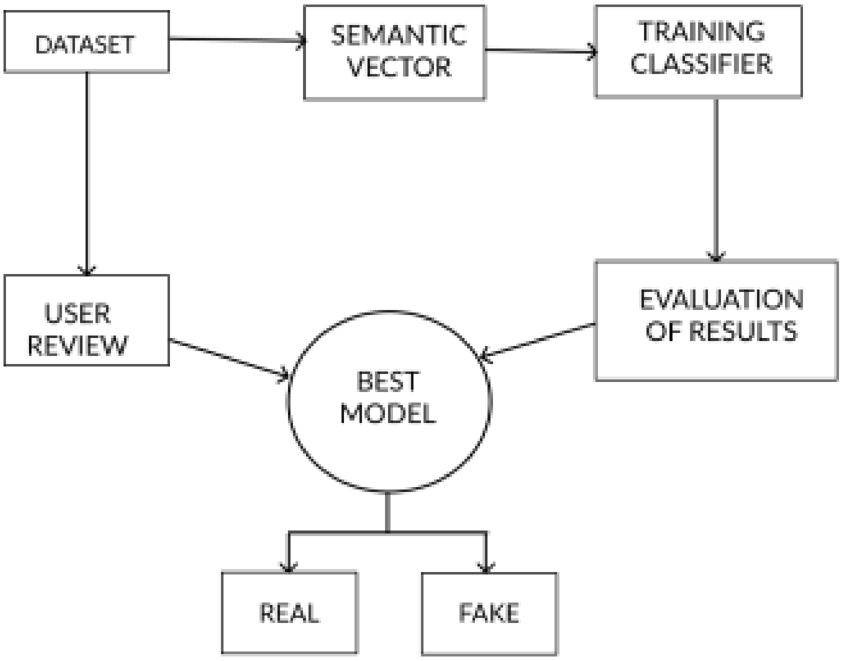


Figure 2.3: working flow of proposed model

This research [13] employs data preprocessing, feature extraction, and classification techniques for the identification of fraudulent product reviews, and the findings have been made publicly available. The process commences with data preparation to eliminate any redundant or irrelevant noise signals. After preprocessing, positive and negative assessments are utilized to generate lists of keywords. The subsequent phase involves feature extraction, where relevant features are extracted from the preprocessed data. Notably, the CNN-BiLSTM algorithm proves to be the most efficient approach for detecting counterfeit products.Figure shows the proposed workflow diagram [13]. In the data preprocessing phase of an online fake review detection system, several essential steps are taken to prepare textual data for analysis. This process typically begins with the removal of punctuation and special characters to ensure that the text is clean and consistent. Next, stop words, which are common and non-informative words like ”the,” ”and,” and ”is,” are removed to focus on meaningful content. Tokenization follows, breaking down text into individual words or tokens. Finally, stemming is applied to reduce words to their base or root form, which helps in grouping similar words together. These steps collectively enhance the quality of the text data, making it ready for feature extraction and subsequent machine learning analysis. The study [13]utilized the YelpZip dataset, which encompasses both legitimate and deceptive reviews, collected from a total of 1,035,038 reviewers. This dataset was employed to assess the hypotheses presented in this research. Each dataset was partitioned into three segments: a training set (60 percent), a validation set (10 percent), and a testing set (30 percent) The training set comprises spam and non-spam reviews that have been annotated, whereas the validation set contains reviews that have been used to fine-tune the model. The testing set includes unlabeled data that can be assumed to be false or true based on

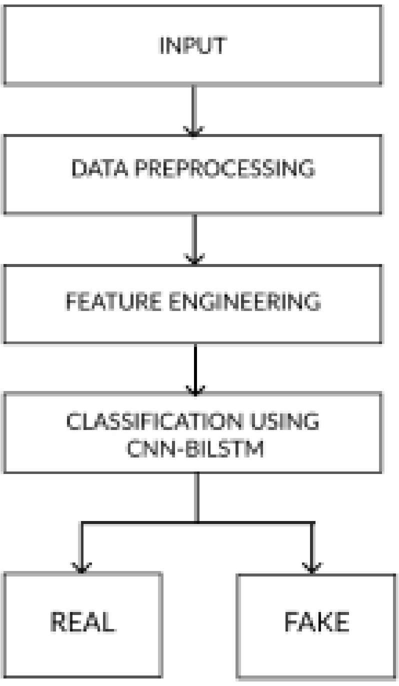


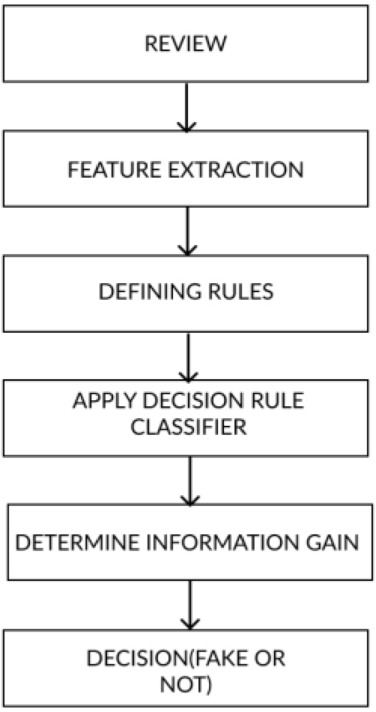
Figure 2.4: Proposed Workflow

past experience.

In this paper [14], the authors collected web reviews related to a specific product and extracted additional information about the reviewers. Their goal was to identify fake reviewers using a decision tree classifier and Information Gain. They confirmed the significance of features in influencing the decision through Information Gain analysis. The research[14] included extensive experiments conducted on a comprehensive dataset of web-extracted reviews, demonstrating the effectiveness of the proposed approach.

The researchers employed the Decision Trees classifier technique to identify fake reviews, utilizing six distinct criteria, including star ratings, responses, replies, profile usefulness, profile status, and template conditions. They used the Web Harvey crawler to extract online product reviews and extracted potential features from these reviews. Decision rules were established for the Decision Rule Classifier, which employed various criteria to distinguish fake reviews within the dataset.

The dataset consisted of 200 online reviews from web users for the product ’LED TV’ available on the website ’www.amazon.com.’ The proposed approach achieved an impressive success rate of 96 percent.



[14]

Figure 2.5: Block Diagram of Proposed Method

This [15]offers an overview of the researchers' study, which seeks to develop a machine learning model for distinguishing between genuine and fake reviews within Yelp's dataset. They systematically applied and compared various classification techniques in machine learning to determine the most effective approach. Succinct explanations are provided for each classification technique to elucidate the reasons behind their varying performance. Notably, the highest level of success was attained through the utilization of the XGBoost classification technique, yielding an impressive F-1 score of 0.99 in predictions. The dataset from Prof. Rayana from Stony Brook University is used. The dataset contains reviews of restaurants and hotels in Chicago. they used only restaurant reviews in the experiments. There are 61541 reviews from 33502 reviewers. In training the model of [15], the researchers incorporated a comprehensive set of features, which included both fundamental and intricate elements. These encompassed attributes such as the usefulness, funniness, and coolness ratings, the length of reviewer reviews, the similarity between reviewer reviews, the extreme rating ratio of reviewers, the deviation of ratings from the mean of reviews, and the maximum number of reviews a reviewer posted in a single day. The paper applied four distinct machine learning algorithms: logistic regression, support vector machine, Gaussian Naive Bayes, and XGBoost, using these extracted features.Among the four machine learning algorithms [15], the results obtained with XGBoost exhibited the highest predictive performance. However, it's worth noting that SVM had the longest training time for their models compared to the others. The result of machine learning prediction using XGBoost classification algorithm achieved 0.99 in F-1 score average in prediction while using the other algorithms, the maximum F-1 score is 0.78 on average.

In this [16], numerous methods for detecting fake profiles and online social bots are examined. The research delves into the multi-agent perspective of online social networks and discusses the relevance of machine learning techniques in profile creation and analysis. Social networks are the social connection build over online network among the people having similar interests or have common backgrounds etc. Bot detection in online social networks (OSN) encompasses three primary techniques: content-based bot detection, network graph-based detection, and a combination or hybrid approach. Initially, efforts to detect bots were primarily concentrated on examining the relationships among users.Al-Jarrah's survey on efficient machine learning highlights the importance of processing big data efficiently for better prediction accuracy. The POST Method considers People, Objectives, Strategies, and Technology as key factors in using Online Social Networks (OSNs). In fake profile detection, machine learning techniques analyzing user metadata offer scalability advantages over graph-based methods, with supervised learning taking various forms like logistic regression and Na¨ıve Bayes.The utilization of social engineering techniques and the proliferation of fake profiles have posed significant threats to data security and the manipulation of information for ulterior purposes. The paper [16]provides an extensive review of methods for identifying fake profiles within established social networks. Additionally, it also explores the influence of machine learning methods on bot detection, contributing to a comprehensive understanding of these issues.

In this paper[17] , a unified framework known as Reliable Fake Review Detection (RFRD) is introduced. This framework explicitly incorporates the temporal patterns of users' review behavior into a probabilistic generative model. Additionally, RFRD addresses users' underlying review credibility and the highly skewed review distributions of objects. It includes experiments conducted on two Yelp datasets(from two cities: Charlotte (CLT) and Phoenix (PHX)), showcasing the effectiveness of the RFRD framework proposed by the researchers. It[17] discusses two key points: First, a fake review might go undetected because it can mimic normal behavior, making it challenging to identify a camouflaged account. Second, a genuine review could be mistakenly labeled as fake if it appears within a burst sequence of fake reviews. Consequently, techniques focused solely on individual factors may not provide reliable detection. The RFRD framework systematically examines user and object characteristics, incorporating their temporal behaviors. Additionally, it has the capability to discern users' concealed review credibility and the infrequent review patterns of objects. Through a combination of explicit and implicit factors, the RFRD model takes into account the inter-relationships between these elements to effectively identify spam reviews. It used two evaluation metrics:

1. Classification-based metrics: Precision(Per), Recall(Recc), Accuracy(accu), F1-score (F1).

2. Rank-based metrics: Average Precision (AP), Area Under the curve (AUC).Baselines: Singular Value Decomposition (SVD), TruthFinder (TF), Average Log (AL), HITS, FRAUDEAGLE

The provided text is an abstract or summary of a research study [18]that reviews existing literature on spam review detection in online reviews. The study evaluates methods, techniques, and metrics used in spam review detection, identifies feature extraction techniques, and discusses future directions in the field.This practice is known as review spamming. In the past few years, a variety of methods have been suggested in order to solve the issue of spam reviews. In this study, the researchers carry out a comprehensive review of existing studies on spam review detection using the Systematic Literature Review. It emphasizes the interdependencies between feature extraction and detection accuracy. This research [18]aims to contribute to the understanding and improvement of spam review detection methods.

3. Current Trends and Work

The comprehensive methodologies implemented in these research papers portray the breadth and depth of intricate techniques utilized to confront the complex challenge of detecting fake reviews within online platforms, emphasizing the ongoing evolution aimed at enhancing the precision, reliability, and robustness of these detection systems.

1. Sentiment Analysis: Understanding Emotions, Analyzing the sentiment and emotional tone within reviews to spot overly positive or negative sentiments that might indicate fakeness.

2. Linguistic Analysis: Identifying Anomalies: Assessing the linguistic patterns, grammar, and vocabulary usage within reviews to detect unnatural language that might signal manipulation.

3. Metadata Examination: Review Context Analysis: Assessing reviewer behavior, timestamps, and frequency of reviews to identify irregularities or patterns associated with fake reviews.

4. Machine Learning Models: Classification Algorithms: Employing models like Naive Bayes, Support Vector Machines (SVM), or Neural Networks to classify reviews as genuine or fake based on extracted features and patterns.

5. Ensemble Learning: Combined Models: Utilizing multiple models or ensemble methods to improve detection accuracy, integrating various algorithms to make collective predictions.

6. Text Preprocessing: Cleaning and Preparing Data: Removing noise, tokenization, stemming, and converting text into a format that machine learning algorithms can process.

7. Anomaly Detection: Identifying Unusual Patterns: Employing anomaly detection techniques to spot unusual behaviors or patterns within reviews that deviate from the norm.

8. Real-time Monitoring: Immediate Detection: Systems capable of real-time analysis, swiftly identifying and flagging potentially fake reviews upon submission.

9. Feature Engineering: Extracting Useful Features: Deriving relevant features from reviews, such as sentiment, word frequencies, or metadata, to feed into machine learning models. These methodologies are employed individually or in combination to develop systems capable of accurately differentiating between genuine and fake reviews. Continuous advancements in these methodologies aim to address the evolving tactics used by those generating fake reviews and to enhance the accuracy and efficiency of detection systems.

4. Proposed Solution and its Design

4.1 Proposed Solution

Detecting and mitigating fake reviews using Machine Learning (ML) represents a powerful and evolving approach in the ongoing battle against deceptive practices in online review platforms. This proposed solution involves a multi-stage process, starting with the collection and preprocessing of a comprehensive dataset containing both genuine and fake reviews, enabling the ML model to learn and distinguish between the two categories. Feature engineering plays a pivotal role, with the creation of a diverse feature set that combines text-based characteristics like sentiment analysis and keyword frequency with metadata such as reviewer history. To boost the accuracy of fake review detection, a comprehensive approach that combines innovative techniques and strategic methodologies is essential. Here's a proposed solution incorporating a range of strategies:

Feature extraction: Feature extraction is the process of transforming raw data into a set of relevant characteristics or attributes, called features, that can be used to represent the data in a more meaningful way. In this project, we have collected reviewer's Id and reviewer's text for the main columns or feature.

Feature Encoding: Feature encoding is a critical step in machine learning, particularly when dealing with categorical data or textual information that algorithms can't directly process. It involves converting categorical or text data into a numerical format that machine learning models can effectively interpret and learn from.

Machine Learning Algorithm:

We will apply some of the following algorithms to classify the reviews :

- Naive Bayes Classifier: Naive Bayes is a simple but powerful classification algorithm based on Bayes' theorem with the assumption of independence among predictors (also known as features or attributes). Despite its ”naive” assumption of feature independence, it has shown to be effective in many real-world applications, particularly in text classification and spam filtering.
- Support Vector Machines (SVM): Classifies data by finding the hyperplane that best differentiates classes. Also effective for high-dimensional spaces, ideal for text-based classification tasks.

- Random Forest:Ensemble learning method that constructs multiple decision trees and merges their predictions.
- Logistic Regression: It is suitable for binary classification tasks. It also estimating the probability that a given review belongs to a particular class or not.
- Neural Networks:Deep learning models such as Multilayer Perceptrons (MLPs) or Convolutional Neural Networks (CNNs) can learn intricate patterns from data. Effective for complex feature representations, but might require computational resources.

4.2 Design of Proposed Solution

4.2.1 Architecture

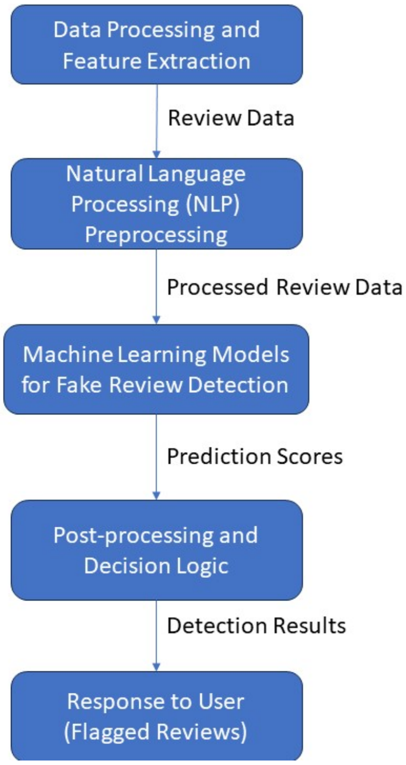

Data Processing and Feature Extraction
Review Data
Natural Language Processing (NLP) Preprocessing
Processed Review Data
Machine Learning Models for Fake Review Detection
Prediction Scores
Post-processing and Decision Logic
Detection Results
Response to User (Flagged Reviews)

Figure 4.1: System Architecture

The system architecture for fake review detection typically involves several interconnected components and processes. Initially, the architecture begins with a data collection module, acquiring review data from various sources like e-commerce platforms or review websites. This data undergoes preprocessing and feature extraction, identifying essential elements such as sentiment analysis, text structure, language patterns, and more. These features are then fed into the machine learning models, which are trained to distinguish between genuine and fake reviews. These models, which could include decision trees, support vector machines, or neural networks, form the core of the detection system. The evaluation and tuning module continuously refines the models, adjusting hyperparameters and optimizing feature selection to enhance accuracy and reduce misclassifications. Once trained and refined, the models are integrated into the detection system, which analyzes incoming reviews in real-time. Additionally, there might be a feedback loop to continuously improve the system based on newly identified fake review patterns. The architecture should be scalable, allowing for handling increasing review volumes, and robust enough to adapt to new forms of fraudulent behavior. Finally, the system typically includes a reporting and analytics module to generate insights on the system's performance, helping in continual improvement and transparency.

4.3 Requirement Specification of Software and Hardware

4.3.1 Hardware Resources Required

PROCESSOR :- PENTIUM –IV

SPEED :- 2.4 GHZ

RAM :- 4 GB(MIN)

HARD DISK :- 80 GB

KEY BOARD :- STANDARD WINDOWS KEYBOARD

MOUSE :- TWO OR THREE BUTTON MOUSE

MONITOR :- LCD/LED

4.3.2 Software Resources Required

Programming Language: Python
Machine Learning Libraries: Scikit-learn,NLTK (Natural Language Toolkit), Matplotlib, Seaborn
Data Collection and Analysis: Web Scraping Tools (such as Beautiful Soup, Scrapy, etc.) for collecting review data.
Machine Learning Models:Classification models
Using tools: Git and GitHub for version control and collaboration with a team.
Google Collab.

4.3.3 Constraints and Assumptions

Assumption:

We consider Yelp Dataset to represent for training and testing the model. It should consists a diverse range of reviews, both genuine as well as fake, covering various Hotels and Restaurants. By applying simple and efficient classification algorithms we classify them as true or fake .The accurcay of classification algorithm might be above 89.40.

Constraint:

Reviews might lack sufficient context or metadata, making it difficult to assess the authenticity of the review. Information such as reviewer history, purchase behavior, or other contextual data can be vital for accurate detection but may not be always available.

4.3.4 Required Features and Functionalities

o Data preprocessing

o Feature Extraction

o Metadata dataset

o Classification algorithms

5. Milestones for Dissertation Stage - II

A milestone is a specific project task with no duration, serving as a marker to denote a significant accomplishment within a project. Milestones provide a means to gauge the project's progress, especially for those who may not be acquainted with the specific tasks being carried out. Stage 2 milestones for a fake review detection project using machine learning include:

Data collection and preprocessing, feature engineering, model selection and development, model evaluation and tuning, prototype testing and validation, and documentation and reporting. These stages involve tasks like organizing data, extracting relevant features, developing and refining machine learning models, testing prototypes. Finally, documenting the entire process, including data sources, preprocessing steps, model architectures, and evaluation results, culminates this stage, leading to a comprehensive report summarizing findings, challenges faced, and the models' effectiveness.

5.1 Expected outcomes

ML models can enhance the accuracy of distinguishing between genuine and fake reviews, reducing the likelihood of misclassifications. Higher accuracy ensures better reliability in identifying fraudulent reviews.Through model optimization and fine-tuning, the system aims to decrease false positives (genuine reviews classified as fake) and false negatives (fake reviews classified as genuine), thereby enhancing precision and recall rates. Effective ML models streamline the review detection process, enabling quicker identification and flagging of suspicious reviews. This efficiency can significantly reduce the workload for manual review verification.The system's design should be scalable, allowing it to handle a growing volume of reviews without compromising performance. It should effectively adapt to increasing data and maintain its accuracy. Successful fake review detection contributes to enhancing user trust in online platforms by maintaining the integrity of reviews, fostering a more reliable and transparent environment for consumers.

Conclusion

We introduce increasing the accuracy of fake review detection involves a combination of techniques of strategies and consideration that encompasses the human aspect problem. Some points are consider for expanding the scope of fake review detection and improving accuracy. Gather the information of amazon dataset of fake reviews includes various features of reviews characteristics and ensuring the dataset is accurately labelled with clear distinctions between real and fake reviews. Utilize the techniques like NLP (Part of speech, semantic analysis). Incorporate contextual information such as the reviewer's profile, their purchase history, and the timing of their reviews relative to their interaction with the product and services

REFERENCES

1. [1] Received 30 November 2022, accepted 4 December 2022, date of publication 8 December 2022, date of current version 14 December 2022. Digital Object Identifier 10.1109/ACCESS.2022.3227631 Fake Online Reviews: A Unified Detection Model Using Deception Theories MUJAHED ABDULQADER , ABDALLAH NAMOUN , (Member, IEEE), AND YAZED ALSAAWY Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia Corresponding author: Mujahed Abdulqader (mujahid.kamal2013@gmail.com).IEEE

2. [2] Received September 8, 2020, accepted September 27, 2020, date of publication October 5, 2020, date of current version October 16, 2020. Digital Object Identifier 10.1109/ACCESS.2020.3028588 Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training. IEEE

3. [3] Received January 17, 2022, accepted February 10, 2022, date of publication February 18, 2022, date of current version March 10, 2022. Digital Object Identifier 10.1109/ACCESS.2022.3152806 The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection. IEEE HINA TUFAIL 1, M. USMAN ASHRAF 2 , KHALID ALSUBHI 3 , AND HANI MOAITEQ ALJAHDALI 4

4. [4] Aug 10th, 12:00 AM Fake Review Detection - The Value of Domain-Specificity René Theuerkauf Martin-Luther University Halle-Wittenberg, rene.theuerka halle.de Ralf Peters Institute for Information Systems and Operations Research, ralf.peters@wiwi.uni-halle.de.AISel

5. [5] Creating and detecting fake reviews of online products Joni Salminen a,b,, Chandrashekhar Kandpal c, Ahmed Mohamed Kamel d,Soon-gyo Jung a Bernard J. Jansen a Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar b Turku School of Economics at the University of Turku, Turku, Finland c Jaypee Institute of Information Technology, Noida, India d Cairo University, Cairo, E

6. [6] Fake Reviews Detection through Ensemble Learning Luis Gutierrez-Espinoza1, Faranak Abri1, Akbar Siami Namin1, Keith S. Jones2, and David R. W. Sears31Department of Computer Science,2Department of Psychological Sciences, 3Performing Arts Research Lab Texas Tech University Luis.Gutierrez-Espinoza, faranak.abri, akbar.namin, keith.s.jones, david.sears@ttu.edu

7. [7] Fake Reviews Detection through Analysis of Linguistic Features Faranak Abri1, Luis Felipe Gutierrez ´1, Akbar Siami Namin1, Keith S. Jones2, and David R. W. Sears31Department of Computer Science, 2Department of Psychological Sciences, 3Performing Arts Research Lab1,2,3Texas Tech Universityfaranak.abri, luis.gutierrezEspinoza, akbar.namin, keith.s.jones, david.sears@tt

8. [8] Rakibul Hassan and Md Rabiul Islam. Impact of sentiment analysis in fake online review detection. In 2021 International conference on information and communication technology for sustainable development (ICICT4SD), pages 21–24. IEEE, 2021.

9. [9]Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie Glance. What yelp fake review filter might be doing? In Proceedings of the international AAAI conference on web and social media, volume 7, 2013.

10. [10] Faisal Khurshid, Yan Zhu, Zhuang Xu, Mushtaq Ahmad, and Muqeet Ahmad. Enactment of ensem- ble learning for review spam detection on selected features. International Journal of Computational Intelligence Systems, 12(1), 2019.

11. [11] Nidhi A Patel and Rakesh Patel. A survey on fake review detection using machine learning tech- niques. In 2018 4th international Conference on computing Communication and automation (IC- CCA), pages 1–6. IEEE, 2018.

12. [12] Abrar Qadir Mir, Furqan Yaqub Khan, and Mohammad Ahsan Chishti. Online fake review detection using supervised machine learning and bert model. arXiv e-prints, pages arXiv–2301, 2023.

13. [13] Raghavendra Reddy and UM Ashwin Kumar. Convolutional neural networks-bidirectional long short term memory based fake review classification. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pages 1069–1073. IEEE, 2022.

14. [14] KS Sanjay and Ajit Danti. Detection of fake opinions on online products using decision tree and information gain. In 2019 3rd International Conference on Computing Methodologies and Com- munication (ICCMC), pages 372–375. IEEE, 2019

15. [15] Andre Sihombing and Alvis Cheuk Ming Fong. Fake review detection on yelp dataset using clas- sification techniques in machine learning. In 2019 international conference on contemporary com- puting and informatics (IC3I), pages 64–68. IEEE, 2019.

16. [16] Vijay Tiwari. Analysis and detection of fake profile over social network. In 2017 International Conference on Computing, Communication and Automation (ICCCA), pages 175–179. IEEE, 2017.

17. [17] Xian Wu, Yuxiao Dong, Jun Tao, Chao Huang, and Nitesh V Chawla. Reliable fake review detection via modeling temporal and behavioral patterns. In 2017 IEEE International Conference on Big Data (Big Data), pages 494–499. IEEE, 2017.

18. [18] Spam Review Detection Techniques: A Systematic Literature Review Naveed Hussain 1,2, Hamid Turab Mirza 1,* , Ghulam Rasool 1 , Ibrar Hussain 2 and Mohammad Kaleem 3 Received: 20 February 2019; Accepted: 4 March 2019; Published: 8 March 2019.

Details of Publication

1. Acceptance Notification and Review Result of Your Paper - JETIR526647 — JETIR (ISSN:2349-5162) — www.jetir.org — editor@jetir.org Your Email id: tamannamaner2003@gmail.com Track Your Paper Link Track Your Paper https://www.jetir.org/trackauthorhome.php?arid=526647.

2. Paper accepted in International Conference on Research based Paper:Unique Paper ID: DYPATIL-ICIEST /Nov-2023/P3002.

Details of Certifications

1. Certification in Avishkar College Level Poster Competition(21/09/2023).
2. Internal Smart India Hackathon Competition -2023(21/09/2023).

Bibliography

[1] M. Abdulqader, A. Namoun, and Y. Alsaawy, "Fake online reviews: A unified detection model using deception theories," *IEEE Access*, vol. 10, pp. 128622–128655, 2022.

[2] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi, and X. Xiao, "Fake review detection based on multiple feature fusion and rolling collaborative training," *IEEE access*, vol. 8, pp. 182625–182639, 2020.

[3] H. Tufail, M. U. Ashraf, K. Alsubhi, and H. M. Aljahdali, "The effect of fake reviews on e-commerce during and after covid-19 pandemic: Skl-based fake reviews detection," *Ieee Access*, vol. 10, pp. 25555–25564, 2022.

[4] R. Theuerkauf and R. Peters, "Fake review detection-the value of domain-specificity," 2023.

[5] J. Salminen, C. Kandpal, A. M. Kamel, S.-g. Jung, and B. J. Jansen, "Creating and detecting fake reviews of online products," *Journal of Retailing and Consumer Services*, vol. 64, p. 102771, 2022.

[6] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. Sears, "Fake reviews detection through ensemble learning," *arXiv preprint arXiv:2006.07912*, 2020.

[7] F. Abri, L. F. Gutierrez, A. S. Namin, K. S. Jones, and D. R. Sears, "Fake reviews detection through analysis of linguistic features," *arXiv preprint arXiv:2010.04260*, 2020.

[8] R. Hassan and M. R. Islam, "Impact of sentiment analysis in fake online review detection," in *2021 International conference on information and communication technology for sustainable development (ICICT4SD)*, pp. 21–24, IEEE, 2021.

[9] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?," in *Proceedings of the international AAAI conference on web and social media*, vol. 7, pp. 409–418, 2013.

[10] F. Khurshid, Y. Zhu, Z. Xu, M. Ahmad, and M. Ahmad, "Enactment of ensemble learning for review spam detection on selected features," *International Journal of Computational Intelligence Systems*, vol. 12, no. 1, pp. 387–394, 2019.

[11] N. A. Patel and R. Patel, "A survey on fake review detection using machine learning techniques," in *2018 4th international Conference on computing Communication and automation (ICCCA)*, pp. 1–6, IEEE, 2018.

[12] A. Q. Mir, F. Y. Khan, and M. A. Chishti, "Online fake review detection using supervised machine learning and bert model," *arXiv preprint arXiv:2301.03225*, 2023.

[13] R. Reddy and U. A. Kumar, "Convolutional neural networks-bidirectional long short term memory based fake review classification," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1069–1073, IEEE, 2022.

[14] K. Sanjay and A. Danti, "Detection of fake opinions on online products using decision tree and information gain," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 372–375, IEEE, 2019.

[15] A. Sihombing and A. C. M. Fong, "Fake review detection on yelp dataset using classification techniques in machine learning," in *2019 international conference on contemporary computing and informatics (IC3I)*, pp. 64–68, IEEE, 2019.

[16] V. Tiwari, "Analysis and detection of fake profile over social network," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 175–179, IEEE, 2017.

[17] X. Wu, Y. Dong, J. Tao, C. Huang, and N. V. Chawla, "Reliable fake review detection via modeling temporal and behavioral patterns," in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 494–499, IEEE, 2017.

[18] N. Hussain, H. Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Spam review detection techniques: A systematic literature review,‖ appl," 2019.