M. Tech. Dissertation Preliminary Report

titled

DETECTING FRAUDULENT REVIEWS USING MACHINE LEARNING

Submitted in partial fulfilment towards the award of the degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING WITH SPECIALIZATION IN DATA SCIENCE

by

Ms. Aditi Das
P23DS008

Supervisor

Dr. Krupa N. Jariwala, SVNIT, Surat

2024 – 2025

Department of Computer Science and Engineering
SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY, SURAT.

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY, SURAT

DECLARATION

I hereby declare that the work being presented in this dissertation preliminary report entitled "Detecting Fraudulent Reviews Using Machine Learning" by me i.e. Ms. Aditi Das, bearing Admn. No: P23DS008 and submitted to the Department of Computer Science And Engineering at Sardar Vallabhbhai National Institute of Technology, Surat; is an authentic record of my own work carried out during the period of July 2024 to December 2024 under the supervision of Dr. Krupa N. Jariwala. The matter presented in this report has not been submitted by me in any other University/Institute for any cause.

Neither the source code there in, nor the content of the project report have been copied or downloaded from any other source. I understand that my result grades would be revoked if later it is found to be so.

____

(Aditi Das)

C E R T I F I C A T E

This is to certify that the dissertation preliminary report entitled " Detecting Fraudulent Reviews Using Machine Learning ", prepared and presented by Ms. Aditi Das, bearing Admn. No: P23DS008 of MTech.- II, Semester - III in Computer Science And Engineering with Specialization in Data Science, at Department of Computer Science and Engineering of the Sardar Vallabhbhai National Institute of Technology, Surat is satisfactory.

Certified By

____

Dr. Krupa N. Jariwala
Assistant Professor,
Department of Computer Science and Engineering
Sardar Vallabhbhai National Institute of Technology,
Surat – 395007, Gujarat
India

____

Jury's Signature

____

PG Incharge,
M.Tech in Computer Science and Engineering,
with Specialization in Data Science,
SVNIT - Surat

____

Head,
Department of Computer Science and Engineering,
SVNIT - Surat

Acknowledgments

I am deeply grateful to Dr. Krupa N. Jariwala, Assistant Professor of Computer Science and engineering, SVNIT, Surat; for her invaluable guidance and support throughout this project. As a supervisor, her insightful feedback, encouragement, and expertise have been instrumental in shaping this work and pushing me to strive for excellence. Her mentorship has not only enhanced my technical skills but also instilled in me a deeper understanding of research methodologies.

I am also thankful to Dr. Mukesh A. Zaveri, Head of the Department, for providing the infrastructure and resources necessary to conduct my research. His support and the department's facilities created a conducive environment for carrying out this project successfully.

This accomplishment would not have been possible without the guidance and support of all these individuals.

Aditi Das

P23DS008

Abstract

Now detecting fake reviews has become the most challenging task in upholding the integrity of online platforms. This work explores machine learning-based pipelines for effective classification of a piece of text as fake versus genuine reviews. Three independent pipelines are implemented, using both CountVectorizer and TF-IDF for feature extraction along with Random Forest, Support Vector Classifier (SVC), and Logistic Regression classifiers. The preprocessing step includes tokenization, removing punctuation and digits, stop words removal, stemming, and lemmatization for normalizing text data. From the experimental results, it can be seen that SVC produces the highest accuracy in prediction with 88.26%, which is because of the ability to identify the optimal hyperplanes in the high-dimensional feature space. The logistic regression model, which produces 86.66% accuracy, performs very well in the case of linearly separable data but not so good for the non-linear patterns. Random Forest achieves 84.35% accuracy, with its slightly lower performance linked to overfitting tendencies in sparse, high-dimensional text data. The study is focused on the importance of strong preprocessing and feature extraction techniques, especially TF-IDF, in improving model performance.

Further work will include multisourced data curation, hyperparameter optimization using GridSearchCV, and embedding techniques such as BERT for higher accuracy and scalability. This work contributes to developing robust and scalable solutions against fake reviews, with great implications for e-commerce, social media, and other online applications.

Keywords: *Fake review detection, machine learning pipeline, Support Vector Classifier, preprocessing of text, CountVectorizer, TF-IDF, Logistic Regression, Random Forest, BERT embeddings, GridSearchCV, multisourced data curation.*

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The online reviews have been very vital for the digital economy since they contribute to consumer choice decisions and impact product and service reputations. Some platforms such as Amazon[1], eBay[2], Yelp[3], and TripAdvisor[4] rely on reviews produced by users to generate trust, offer customer information, and increase sales. The dependency, however, has increased on the rise of fake reviews meant to deceive consumers and distort market dynamics. Fake reviews damage the reputation of a company, cause a financial loss to it, and mislead the consumer in purchasing decisions. Unless this problem is tackled, e-commerce platforms might lose the trust of customers and legitimacy in online transactions will come down.

Fraudsters have now become smarter. They have employed sophisticated tools to perpetrate their frauds. This has, therefore, become an area of much research for identifying fake reviews. Nearly totally ineffective against the adaptive tactics of fake review writers, early methods based on algorithms or keyword matching are useless. More current false reviews use sophisticated means of operation, like apparently true AI-generated language. The processes have to scale to thousands of pieces of information at a time in order to be useful, with minimal time requirements due to the size and real-time nature of this dataset.

To offer some more solid basis for reliable detection frameworks, this thesis tackles the problem of review-falseness detection, expands upon recent solutions, and uncovers vital gaps in existing methods. Advanced machine learning and natural language processing tools are applied to develop practical solutions for the enhancement of trust and transparency in online review systems. The following sections further explain why the research was undertaken, what problems it covers, the contributions needed for carrying out this study, and describing the structure of the report.

[1] Amazon website
[2] eBay website
[3] Yelp website
[4] TripAdvisor website

1.1 Motivation

Online reviews have changed the way people interact with products and services. Since trust in traditional advertising has dwindled, online reviews have become the foundation for consumer choice. The platforms such as Amazon, Yelp, and TripAdvisor rely on user-generated reviews to influence consumer decisions, boost sales, and develop brand reputations. However, with this dependence, the platforms have also become fertile ground for exploitation through fake reviews. These reviews are made with deceptive purposes of rating inflation, or even sabotage of a competitor, which not only harms an individual consumer but also leads to financial burdens for businesses and distorts fair competition, causing a discredit of the platforms.

The sophistication in fake reviews increases the sense of urgency in dealing with them. Fake reviews surfacing lately are more easily traced as these often contain superfluous enthusiasm and duplicate patterns. Nowadays, with the advancement in natural language generation tools power through AI, reviews get generated mimicking real reviews on tone, structure, and content. For example, AI-based bots can now produce customized reviews of context that avoid the normal and traditional detection mechanisms. This raises questions about the need for continuous and evolving research and technology towards responding to the challenge. Besides, spotting reviews that are fake helps respond to the wider objectives for having digital literacy, safe protection of consumers, and trust in e-commerce platforms.

The ethical implications of fake reviews add a huge weight to this research. There are economic implications, but the loss of the moral fabric of trust and fairness in digital transactions is much more severe. As the volume of e-commerce continues to increase, safeguarding the integrity of online commerce becomes paramount. This dissertation is expected to contribute towards the above efforts by examining advanced methodologies for enhancing the effectiveness and scalability of fake review detection systems.

1.2 Problem Description

The problem of fake reviews is complex and multifaceted, as well as constantly developing, manufactured with the intent to deceive, either by overstating the virtues of a product or unfairly and exclusively badmouthing competitors. Indeed, fake reviews are quite indistinguishable from authentic feedback, because fraudulent actors use tactics that closely mirror true user behavior. Modern fake reviews are increasingly generated using advanced AI tools capable of producing human-like text. This trend has also made the traditional techniques used for detection, like keyword spotting or rule-based systems, fairly in-effective. A variety of sources and methods by which fraudsters operate is one of the challenges in fake review detection. A paid individual might be employed to write fabricated testimonials; a click farm could churn out a large number of reviews; or, by using AI, a bot could create content on their own. All sources of such content are different and thus require a multi-pronged approach by detection systems. For example, human-crafted fake reviews rely on nuances of words, while bot-crafted reviews could only repeat patterns that human eyes tend to miss out. Scalability presents another level of sophistication for this problem. Platforms handling millions of reviews on a daily basis create a situation that cannot be manually moderated. Therefore, automated detection systems should not compromise accuracy or have massive false positives and handle data volumes in real time. Furthermore, the adversarial nature of the problem ensures that improvement in detection systems leads fraudulent actors to alter their tactics for the next cycle of evolution.

Ethical considerations also clearly define the problem. Detectors must ensure that actual reviews are not mistakenly classified as fraudulent and the consequences that this may have for genuine users and genuine businesses. Systems must be compliant with privacy regulations such as GDPR, which limits the scope to which user behavior may be analyzed. This thesis defies these limitations through proposing an advanced methodology which allows for effectiveness while preserving ethical and practical constraints.

1.3 Contribution

The primary aim of this dissertation is to design machine learning pipelines that are able to identify whether a review is either real or fake. The pipelines attempt to overcome imbalanced datasets and small variations between the false and the actual reviews using effective techniques of feature extraction, multi-classifiers, and powerful preprocessing procedures. The objectives are:

1. **Pipeline Development and Comparison:** CountVectorizer and TF-IDF are used to convert the review text into a structured numerical format to enhance classifier input. Three different pipelines were developed, and their performances were compared using Random Forest, Support Vector Classifier (SVC), and Logistic Regression in order to identify the model that would be used. The performance of the classifications is measured using metrics such as accuracy, precision, recall, and F1-score in order to obtain relevant data to identify fake reviews.

2. **Comprehensive Preprocessing Techniques:** It presents a pretty powerful preparation pipeline for the standardization and cleaning of textual input in machine learning applications. The procedures may include tokenization, deletions of punctuation or numbers

from text, removal of stop words, stemming, and finally lemmatization. With this process, the input data were cleaned and standardized. These made the models significantly able to classify reviews with increased precision.

3. **Experimental Analysis and Insights:** The study demonstrates, through rigorous experimentation, that SVC has the highest accuracy of 88.26%, which shows that it is the most effective model for fake review detection amongst the classifiers tested. This is followed by Logistic Regression with an accuracy of 86.66% and Random Forest with an accuracy of 84.35%. The outcomes are useful in determining strengths and weaknesses of each model and the robustness of SVC in high-dimensional spaces and Random Forest's susceptibility to overfitting in text-based tasks.

4. **Recommendations for Future Improvements:** The study offers some guidance for the future development of the methods:

   - **Advanced Embedding Techniques:** Using the most current embeddings, such as BERT, may allow the models to sense more complex relationships between the context, and hence help distinguish between real reviews and fake ones.
   - **Multisourced Data Curation:** It adds reviews from various sources such as social media and e-commerce sites to the dataset in order to reduce biases and improve model robustness.
   - **Advanced Deep Learning Methods:** Even deeper architectures involving LSTMs or transformers would continue to push performance with learning sequential and contextual patterns on the reviews.
   - **Hyperparameter Optimization:** Using tools such as GridSearchCV or more advanced optimization techniques can fine-tune model parameters to optimize their effectiveness and scalability.

5. **Applicability in Real-World Scenarios:** One step of the process is building a test application that replicates how it would be used, allowing users to input content for reviews and to then decide whether it is real or fake. This proves that the proposed solutions may be applied in social media and e-commerce sites and other areas that are plagued by fake reviews.

With these contributions incorporated, this work not only forms a solid foundation for machine learning-based false review detection but also opens avenues for further development in the field. The methods and results discussed here can be applied to related fields where textual classification is important.

1.4 Report Outline

This report is split into six chapters, all of which describe a different area of research. Below are the outlined contents of what each of the chapters cover:

1. **Chapter 1 - Introduction:** This chapter summarizes the study, which discusses motivation in doing the research, problem identification that the research aims to answer, contributions made, and the structure of the report. It sets the groundwork for a holistic understanding of the worth of web fake review detection and the purposes of this work.

2. **Chapter 2 - Theoretical Background and Literature Survey:** This paper provides a conceptual framework and extensive reviews of some relevant literatures.It elaborates on current methods, and, in depth, compares differing techniques, followed by review over datasets, which was applicable in analogous studies. This chapter is essential for grounding the justification made in terms of how methods and methodologies were selected.

3. **Chapter 3 - Design and Analysis:** The purposes of the proposed algorithms and the workings of the three machine learning pipelines designed for the detection of false reviews are explained in this chapter. All the preprocessing procedures, techniques of feature extraction, and the pipeline structure that would serve as the framework for their use and evaluation are conceptually analyzed.

4. **Chapter 4 - Implementation Methodology:** This chapter explains the process of implementing it, from dataset preprocessing to feature extraction using CountVectorizer and TF-IDF and the training of machine learning models in three pipelines: Random Forest, Support Vector Classifier, and Logistic Regression. It also deals with the hardware and software environment in which it is built, along with the test application developed to evaluate the system.

5. **Chapter 5 - Performance Results and Analysis:** This chapter shows experimental results of a comparison in performance between three pipelines in terms of precision, recall, and F1-score metrics. The result is broken down to the point of determining the strengths and weaknesses of each pipeline and, therefore, which model will stand out to perform the best.

6. **Chapter 6 - Conclusion and Future Work:** The last chapter summarizes the conclusion of the research, with the proposed pipeline being efficient enough in spam review detection, but there are comments made regarding further improvement, namely embedding techniques that are more complex, multisourced data curation, and optimization strategies, that might help in the efficiency and scalability of the system.

Such an outline of the report will ensure that chapters are read with crystal clarity concerning the logical and coherent process undertaken in the study.

Chapter 2

Theoretical Background & Literature Survey

The growing issue within the digital economy is the constant presence of false reviews. Since purchase decisions and a company's reputation are linked with the feedback from customers, a malicious activity which affects the client's trust, market dynamics, and financial and reputational losses of businesses, fake reviews can be defined as such. This chapter discusses the theory and practice of fake review detection and their important methods and evolution over time.

Types of Fake Reviews:

There are two types of fake reviews, namely promotional reviews used to artificially inflate the value of a product and negative ones aimed at damaging competitors. Such reviews exploit consumer reliance on peer comments and thus require their detection and mitigation in order to ensure that the integrity of the online marketplace is sustained. Such reviews have implications beyond the purchase decision of an individual. Distorted market competition and erosion of consumer confidence in digital platforms are just a few. It must, therefore, require hardy methods to strike between accuracy and scalability with sensitivity towards an evolving deception in place.

1. **Methodologies and Techniques of Fake Review Detection:** Machine learning techniques have become the foundation of fake review detection, which rely on supervised, unsupervised, and semi-supervised algorithms to classify reviews as being genuine or fake. Techniques such as SVM, Random Forest, Logistic Regression, and most recently deep learning models like CNNs and LSTMs have shown great promise. Supervised models require labeled datasets, whereas unsupervised and semi-supervised strategies, like PU learning are challenges related to limited labelled data. Scalable and versatile models such as SGD have been shown to be highly adaptable using iterative optimization for strong classification.

2. **Natural Language Processing (NLP):** NLP tools allow the extraction of linguistic signals in a text that can point to such deception. Techniques to extract features from text are TF-IDF, n-grams, and semantic embeddings that measure the strength of textual attributes. Some of the preprocessing techniques involved include tokenization, removal of stop words, stemming to improve model performance through reducing noise and ensuring uniform representation of data. Modern advanced NLP techniques incorporate pre-trained language models, BERT and Longformer, for capturing contextual nuances with massive improvements in classification accuracy.

3. **Behavior Analysis:** Apart from textual content analysis, review behavior analysis provides another layer of detection. Features like review frequency, redundancy, sentiment intensity, and timestamp patterns reflect anomalous activities pointing toward fake reviews. For modeling the local dependencies of behavioral data, CNNs have been used for obtaining enhanced accuracy in detection.

2.1 Literature Survey

2.1.1 Literature Survey on Methodology

- Balakrishnan et al. [1] presented the robust framework in detecting fake reviews based on using an algorithm in achieving high accuracy and scalability via stochastic gradient descent, thereby fully incorporating preprocessing steps into a proper procedure that included text cleaning and tokenization, removing of stop words, as well as stemming or lemmatizing. Techniques like Bag of Words, n-grams, and semantic embeddings extracted contextual as well as linguistic features of reviews. Among the tested classifiers, the best performance was that of Stochastic Gradient Descent (SGD), with an accuracy of 86.47%, surpassing Random Forest, SVM, and KNN. Its nature of optimization by iteration ensured fine-tuning of the parameters for reducing classification error. The architecture performed at its best in precision, recall, and F1-score, classifying the fake review from genuine ones. Future directions include the incorporation of advanced feature engineering with TF-IDF, pre-trained models like BERT, and real-time review detection capabilities, thus proposing a scalable and adaptive solution for combating deceptive practices in e-commerce.

- Shunxiang et al. [2] introduced the Sentiment Intensity and PU Learning (SIPUL) model, which is a semi-supervised framework for fake review detection. It overcomes some of the challenges posed in handling deceptive reviews and streaming data. The model categorizes the reviews into "strong" and "weak" sentiment sets by exploiting the property that fake reviews usually have overly exaggerated sentiments. Using a Positive-Unlabeled (PU) learning algorithm, the model iteratively identifies reliable negative examples and trains a binary classifier, such as SVM, to detect fake reviews. It adapts to streaming data with continuous updating and control of size over datasets to avoid overfitting. The same model, validated on openly available datasets like Yelp, showcases its ability to distinguish actual from fraudulent reviews with huge boosts in classification accuracy following a sentiment-driven division of the data. Although the model has been successful, it needs manual adjustment for the thresholds of sentiment and as of now only supports file-based implementation. The future work will center on extending the model to network-based streaming, ensuring scalability and enhanced adaptability for diversified e-commerce environments.

- Devika et al. [3] put forward a framework for detecting phony reviews using sentiment analysis and natural language processing. The study uses machine learning classifiers such as Support Vector Machines (SVM) and Naive Bayes to overcome the critical problem of online review manipulation, with tools such as TextBlob for sentiment scoring and TF-IDF for feature extraction. Preprocessed raw text by feature extraction that includes word density, length of the review, and scores for sentiment is a combination step where there will be classification of review: either fake or genuine; fake or genuine are split up into positive or negative in nature for their sentiments. Suspicious activity alerts for administrators send proactively block malicious users improving on trustworthiness in online markets. The framework was found to enhance the accuracy in identifying spam reviews, appropriately inferring sentiment polarity and subjectivity for reviews. Future work includes adding automated spam detection mechanisms and eliminating product-dependent components to expand the suitability of the system for use in many different applications while maintaining feedback accuracy and reliability for the consumer.

- Thilagavathy et al. [4], presented a machine learning framework that can detect and remove fake reviews from e-commerce-based websites, thereby strengthening the trust and credibility in online environments. TF-IDF vectorization has been used to extract features using the supervised learning algorithms Naive Bayes and SVM. Preprocessing applied to the dataset includes tokenization, stop-word removal, and stemming to classify hotel reviews as trustworthy or untrustworthy. The study finds that combining TF-IDF with Naive Bayes and SVM for the text classification is a very effective approach for fake reviews, achieving high accuracy. There were 1,144 spam entries in the analysis dataset of 5,853 reviews. Applications include user-generated content websites and e-commerce, where reviews are crucial in influencing consumers' purchasing decisions. The approach would be more beneficial to social networks and other content sites if scaled to larger datasets and procedures to identify sources of spam propagation were developed.

- Dinesh et al. [5], discussed a complete framework for customer review analysis through opinion mining and NLP techniques that focuses on sentiment classification and spam detection. The paper underlines the importance of identifying spam and fraudulent reviews to ascertain the reliability of feedbacks on e-commerce websites. Their system combines lexicon-based approaches, IP monitoring, and ontology for spam detection and uses machine learning algorithms such as Naive Bayes and SVM for sentiment analysis. Preprocessing operations like removing stop words, stemming, and handling punctuation make the data ready for efficient classification. The framework categorizes reviews as neutral, negative, or positive and, thereby, helps businesses in their decision-making process while making it easier for customers to find reliable input. It thus enhances the identification of fraudulent reviews by combining supervised and unsupervised algorithms with leading-edge strategies such as IP monitoring. Future research will scale this methodology to larger datasets and social networks, further improving performance and adaptability in a variety of e-commerce scenarios.

- Dhamdhere et al. [6] proposed the system to detect the fake reviews on Amazon based on a supervised machine learning approach by especially the SVM. It emphasizes upon the fact that what makes online review so crucial for deciding for consumers and how the presence of fake reviews can make any business lose its money. The methodology used web scraping to scrape data, preprocessing to clean up text using tokenization, stop-word removal, and lemmatization, and the sentiment analysis to determine polarity. Feature extraction techniques were used in order to reduce the complexity of the data and allow for efficient model training. SVM achieved an accuracy of 80%, making it a very effective model for classifying reviews as fake or genuine. This research focuses on the opportunities presented by the system toward increasing the transparency of online buying. The study further envisages future applications on the Yelp and Flipkart platforms so that a system to detect fake reviews can reach out to a wider level.

- Biruntha et al. [7] discussed the use of sentiment analysis in distinguishing reviews, focusing on its application in identifying whether the sentiments are positive, negative, or neutral, using NLP and machine learning techniques. The study was carried out with respect to web-based opinion mining, and in particular business applications of the same in better decision-making through usage of customer's opinion. In their proposed system, they use a dictionary-based approach with the WorldNet unsupervised classification, incorporating supervised learning method including a hybrid SVM+NB model. The preprocessing steps include URL and punctuation removal, stop-word filtering, and stemming to identify the root words. It used POS tagging to fine-tune the accuracy of sentiment classification. The results show that the SVM+NB model outperformed the previous techniques by a wide margin in precision and sentiment classification, hence its applicability in processing complex datasets. The paper underlines the growing demands for automated sentiment evaluation methods in the economy driven by data and points towards integrating linguistic, behavioral, and advanced techniques of machine learning for fake or misleading review detection.

- Elmogy et al. [8] proposed a machine learning framework for the detection of fake reviews based on both review content and reviewer behavior features in the paper "Fake Reviews Detection using Supervised Machine Learning". The study presents the role of engineered features toward better detection accuracy and shows a three-phase approach including data preprocessing, feature extraction, and user behavior analysis. On Yelp preprocessing steps such as tokenization, removing of stop words and lemmatization was carried out. Feature extraction methods included TF-IDF, which made use of both bi-grams and tri-grams, sentiment classification, and cosine similarity that were performed. Other behaviors of the reviewer to detect the spammers included time-stamp frequency, redundant reviews, and usage of punctuation. The classifiers were Logistic Regression, SVM, and KNN. Among them, KNN performed the best with a high accuracy of 87.87% when user behavior features are added. The precision and recall performance is better than others. Thus, it indicates the need for adding user behavior features along with content features. The results also indicate further work that could extend feature engineering to include review frequency and temporal patterns among other attributes for better model performance.

- Zhang et al. [9] proposed an advanced deep learning framework to detect fake reviewers by leveraging both behavioral patterns and textual information. This paper proposes a behavior-sensitive feature extractor using one-dimensional convolution filters to analyze reviewer behavior data while emphasizing the relationship among behavioral features using Spearman's rank correlation. The work demonstrated that behavioral features depend locally, and improved the detection performance. They applied the context-aware attention mechanism, together with Longformer, CNN, BiLSTM, and attention layers, in order to process textual information; this way, they ensured the capturing of valuable text representations, reducing the dimensionality by emphasizing critical textual elements. Their work was tested along two major hypotheses: whether features of behavior could enhance detection and how the fusion of behavioral and textual features could enhance performance. With the YelpZip and YelpNYC datasets, their methodology surpassed benchmarks such as logistic regression, random forests, and transformer models, achieving an accuracy of 85.05% and F1-score of 0.7664. This study opens up future research opportunities that could include richer online features and datasets from diverse platforms for anomaly detection.

- Ozbay and Alatas [10] developed a comprehensive model is presented that uses a two-step approach consisting of text mining and applying 23 supervised AI algorithms to detect fake news within online social media. This model addresses the issues of unstructured social media data through preprocessing to structured form with tokenization, stopword removal, and stemming processes followed by feature extraction through term frequency weighting and the creation of the DTM. The experimental evaluation with three different datasets BuzzFeed Political News, Random Political News, and ISOT Fake News indicated vast variability in algorithm performance. Decision Tree performed at the top across datasets with a mean accuracy of 74.5%, whereas ZeroR, CVPS, and WIHW attained perfect recall. It underlines the capability of effective structured text representation to be integrated with AI algorithms, showing a highly scalable framework for detecting fake news across several datasets. It further draws the possibility of more explorations on ensemble methods and novel algorithms to enhance the precision in detection.

- Singh and Tanwar [11] provided a machine learning framework for identification of fake reviews on home appliances. Methodology Preprocessing of review dataset included elimination of noisy or irrelevant data; count vectorization and TF-IDF transformations were further applied to convert the text data into suitable representations that could be fed to the machine learning models. In the research, three classification algorithms are used: Random Forest, Support Vector Machine (SVM), and Logistic Regression. Of these, SVM proved to be the best, reaching a higher accuracy of 88.11%, whereas Logistic Regression stood at 86.05% and Random Forest at 83.56%. The authors believe that proper text preprocessing techniques, as well as feature transformation methods, make an effective fake review detection system reliable. It shows machine learning pipelines in this context as practical resources for binary classification.

2.1.2 Comparison of Methodology

Authors	Dataset Used	Methods	Contributions	Disadvantages	Remarks
Balakrishnan et al. [1]	Fake reviews dataset by Joni Salminen (CC-By Attribution 4.0 International)	SGD, RF, SVM, KNN	Achieved 86.7% accuracy in fake review detection with SGD, outperforming RF, SVM, and KNN	Limited to the dataset used; may not generalize well to other datasets	Further testing and validation needed for real-world application
Shunxiang et al. [2]	Ott, YelpChi, YelpZIP	Sentiment Intensity and PU Learning (SIPUL)	Proposed a model that continuously learns from streaming data to detect deceptive fake reviews	Requires continuous training data updates, increasing time cost due to the sentiment division step	Effective in detecting fake reviews in large datasets, though performance may decline with data overload
Devika et al. [3]	Not specified	NLP, Sentiment Analysis, TF-IDF, Naive Bayes, SVM	Proposed a framework combining NLP and sentiment analysis for fake review detection	Manual data labeling is resource-intensive, with uncertain credibility of labeled data	Proposed system can filter fake reviews and separate spammers from non-spammers
Thilagavathy et al. [4]	Reviews of 20 Hotels in Chicago hotel dataset	Naive Bayes, Support Vector Machine, TF-IDF Vectorizer	Developed a model to detect fake reviews using machine learning and NLP techniques	Limited dataset used for training, which may affect accuracy	The model can be expanded for greater accuracy and authenticity in reviews

Authors	**Dataset Used**	**Methods**	**Contributions**	**Disadvantages**	**Remarks**
Dinesh et al. [5]	Product review datasets from e-commerce sites	SVM, Naive Bayes, Back-propagation, POS tagging, Stemming, Lexicon-based approach	Proposed a framework for opinion mining using NLP techniques, integrated spam detection module	High complexity and time-consuming spam detection	The system helps in analyzing customer feedback and detecting fake reviews
Dhamdhere et al. [6]	Amazon reviews	Support Vector Machine (SVM)	Developed a system to detect fake reviews on Amazon using SVM, achieving 80% accuracy	Limited to Amazon dataset; may not generalize to other platforms	Focuses on both polarity ratings and classifiers for fake review identification
Biruntha et al. [7]	Twitter API, Amazon, eBay	Naive Bayes, Maximum Entropy, SVM, Unsupervised Dictionary-based Approach	Demonstrated sentiment analysis using various NLP tools, proposed a system for detecting fake reviews	Limited by the quality of the dataset, potential for misclassification	Focus on improving sentiment analysis accuracy and detecting fake reviews

Authors	**Dataset Used**	**Methods**	**Contributions**	**Disadvantages**	**Remarks**
Elmogy et al. [8]	Yelp dataset of restaurant reviews	KNN, Naive Bayes, SVM, Logistic Regression, Random Forest	Proposed an ML approach using textual and behavioral features to detect fake reviews. KNN (K=7) achieved the highest f-score of 82.40%	The study is limited to the Yelp dataset and may not generalize to other datasets	The inclusion of behavioral features increased the f-score by 3.80%
Zhang et al. [9]	Yelp	Behavior-sensitive feature extractor, Context-aware attention mechanism	Proposed an end-to-end framework using behavioral and textual features, achieving state-of-the-art results in fake reviewer detection	Requires extensive computational resources	Significant step towards automated fake reviewer detection, reducing human labor costs

Authors	Dataset Used	Methods	Contributions	Disadvantages	Remarks
Ozbay and Alatas [10]	BuzzFeed Political News, Random Political News, ISOT Fake News	BayesNet, JRip, OneR, Decision Stump, ZeroR, SGD, CVPS, RFC, LMT, LWL, CvC, WIHW, Ridor, MLP, OLM, SimpleCart, ASC, J48, SMO, Bagging, Decision Tree, IBk, KLR	Proposed a two-step method combining text mining and supervised AI algorithms for fake news detection. Evaluated 23 algorithms on three datasets	Limited to the datasets used; may not generalize to other datasets	The Decision Tree algorithm showed the best performance in terms of accuracy, precision, and F-measure
Singh and Tanwar [11]	Home appliances review dataset	Random Forest, Support Vector Classifier, Logistic Regression	Developed a model to identify fake reviews using sentiment ratings and n-grams	Requires substantial labeled data for training	The proposed methods performed better in terms of accuracy compared to previous studies

Table 2.1: Comparison of Different Methodologies

Chapter 3

Design and Analysis

3.1 Goals of Proposed Algorithm

The primary goal of this dissertation is to build pipelines that utilize machine learning for classifying reviews as either real or fake. It would seek to utilize powerful preprocessing procedures, multiple classifiers, and efficient feature extraction techniques to go past the barriers such as imbalanced datasets and slight differences between the fake and real reviews. The goals are stated below:

1. Conversion of review text into the structured numerical format using CountVectorizer and TF-IDF for improved input to classifiers.

2. Three different pipelines were made and tested with Random Forest, Support Vector Classifier, and Logistic Regression to know which one worked best.

3. Using the metrics such as accuracy, precision, recall, and F1-score, classifiers are evaluated to get the desired data required for the detection of fake reviews.

3.2 Workflow

The workflow begins with data preprocessing, where raw text is tokenized, cleaned of punctuation, digits, and stop words, converted to lowercase and processed through stemming and lemmatization. The pre-processed text is then passed through three pipelines. The overview of the dataset used, preparation of the pre-processed review text and the working process of the pipelines are discussed in the following subsections.

3.2.1 Overview of the Dataset Used

This section gives a detailed description of the dataset used (Home Appliances Review) to train and test the fake review detection pipelines. The dataset has 40,431 records and the following attributes:

- **category:** This is the category of the home appliance being reviewed.
- **rating:** It is a number value (1 to 5) that measures consumer satisfaction.
- **label:** This is an attribute that tells whether the review is fake or actual.
- **text_:** This is the text content of the review.

Key features:

- **Balanced Labels:** The dataset will be useful for the training and testing of an unbiased model with equal fake and actual review data.
- **Multilingual Text Data:** The text_ attribute of the dataset is of different linguistic styles that can be computed for feature extraction and other related operations.
- **Domain-Related Knowledge:** As per the home appliances, it gives the freedom to explore some specific kinds of review patterns in domains.

Figure 3.1 is an example of the data set with a structure of a record and attributes that forms the basis of the next step of preprocessing and feature extraction.

	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...

Figure 3.1: Sample of Home Appliances Review Dataset

The structured nature of this dataset supports tasks such as sentiment analysis, opinion mining, and spam detection, in addition to fake review classification.

3.2.2 Preparation of the Pre-processed Dataset

Dataset Preprocessing is the stage of the methodology of fake reviews detection which is a series of transformations applied to text data in order to prepare it for further analysis and model input. The process is followed step by step so that the text will be clean, normalized, and optimized for feature extraction. The next step-by-step description of the process is done according to the following flowchart.

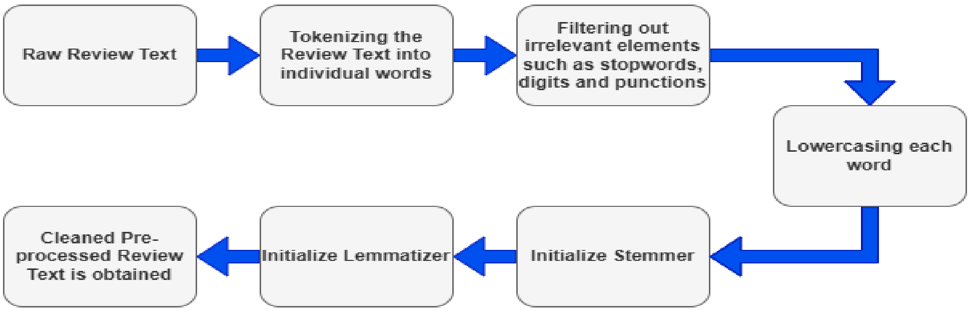

Figure 3.2: Flowchart of Preprocessing the Raw Review Text

1. **Tokenizing the review text:** Tokens are created through input text tokenization into words. The text string is being converted into a list of words for easy manipulation and analysis of components. Tokenizing the text "This product is amazing!" into ['This', 'product', 'is', 'amazing', '!'] is an example.

2. **Filter Out Stopwords, Digits, and Punctuation:** These words like "is", "the", "and", are the very common occurring non-informative words which can be removed using pre-defined stopwords list from NLTK or SpaCy. Another thing which needs to be removed here is the numerical digits and punctuations as they do not contribute semantic meaning but only add noise to the data.

3. **Lowercasing the text:** Now, the text is converted into the lowercase. This is done in order to avoid counting as different words. For example, words like "Product" and "product" are treated differently if not lowercased.

4. **Initialize Stemmer:** Use a stemming method such as Porter Stemmer, to take the words down to their base form. For instance, the words "run," "runs," and "runner" will be brought down to the base "run." That reduces dimension and permits clustering along conceptual lines.

5. **Initialize Lemmatizer:** For example, lemmatization would bring words back to their roots, dictionary forms, but now with context in mind. Thus, "better" lemmatizes to "good." Unlike stemming, it would ensure that the word resulting is valid.

With such preprocessing steps, the dataset would then be cleaner and standardized and even easier for machine learning models to process toward improving the performance of algorithms on fake review detection.

```
Original Text: Very nice set. Good quality. We have had the set for two months now and have not been
Preprocessed Text: nice set Good quality set two months
```

Figure 3.3: Original Text & Preprocessed Text

```
0        love well made sturdi comfort love pretti
1    love great upgrad origin 've mine coupl year
2          pillow save back love look feel pillow
3             miss inform use great product price
4             nice set good qualiti set two month
Name: text_, dtype: object
```

Figure 3.4: Final Preprocessed Text

3.2.3 Functioning of Fake Review Detection Pipeline

Pipelines automate the process of preprocessing and training the model in a sequential manner. Three core stages are: text vectorization, transformation of features, and using the machine learning model. In this way, workflow will be efficient and reproducible. Here is the step-by-step explanation of all three pipelines:

1. **Text Representation (CountVectorizer):** The CountVectorizer transforms the preprocessed textual data into a matrix of token counts where each unique word in the dataset becomes a feature. For example, the sentences "The product is great" and "Great product indeed" generate a vocabulary matrix with word frequencies like:

The	**product**	**is**	**great**	**indeed**
1	1	1	1	0
0	1	0	1	1

Table 3.1: CountVectorizor Matrix

2. **Feature Transformation (TfidfTransformer):** Further transformations done on the matrix by this class compute the TF-IDF score for each word, meaning that the weight for a term is adjusted in function of importance, providing a larger value to the word when it is particular for one review and minimizing common words like "product" or "good". Such a transformation is able to bring forward the normalized representation of textual features where local relevance balances the global importance.

3. **Model Training:** The same classification model is used on different pipelines for fake review detection.

    - **Pipeline 1: Random Forest Classifier:** This is tree-based ensemble learning. Thereby, bagging combines several decisions of multiple decision trees using a bagging procedure with a result of strong reductions in variance and avoiding over-fitting.
    - **Pipeline 2: Support Vector Classifier (SVC):** In the transformed feature space, it finds the best fit hyperplanes to classify and thus distinguish between fake and authentic reviews. It usually demonstrated good performance on larger dimensional data.
    - **Pipeline 3: Logistic Regression:** Probabilistic linear model using sigmoid function to predict the chance of a review being spam. This algorithm helps solve many binary classification problems.

    The system chains together the components into pipelines in a manner that ensures all processing of preprocessing, feature extraction, and classification flows are smooth. Each pipeline would offer different insights for a comparison of the performances by accuracy, precision, and other measures.

```
(0, 7280)       1
(0, 18374)      2
(0, 18652)      1
(0, 23822)      1
(0, 29216)      1
(0, 33201)      1
(1, 7862)       1
(1, 13595)      1
(1, 18374)      1
(1, 19706)      1
(1, 21894)      1
(1, 32098)      1
(1, 34165)      1
(2, 3723)       1
(2, 11616)      1
(2, 18278)      1
(2, 18374)      1
(2, 23093)      2
(2, 26326)      1
(3, 13595)      1
(3, 15837)      1
(3, 19846)      1
(3, 23845)      1
(3, 23975)      1
(3, 32179)      1
:     :
(40430, 28447)          1
(40430, 28741)          1
(40430, 28921)          1
(40430, 29380)          1
(40430, 29509)          1
(40430, 29802)          1
(40430, 29852)          1
(40430, 30345)          1
(40430, 30366)          1
(40430, 30371)          1
(40430, 30432)          1
(40430, 30478)          1
(40430, 30596)          2
(40430, 30720)          1
(40430, 30802)          2
(40430, 31590)          1
(40430, 32179)          2
(40430, 32885)          3
(40430, 33074)          1
(40430, 33117)          7
(40430, 33148)          1
(40430, 33683)          1
(40430, 33803)          1
(40430, 33806)          4
(40430, 34230)          3
```

Figure 3.5: CountVectorized Review Texts

```
(0, 13595)    0.22849485246633358
(0, 15837)    0.5617597836165477
(0, 19846)    0.5603228913884096
(0, 23845)    0.3623513042969684
(0, 23975)    0.34242957418905806
(0, 32179)    0.2640034550694301
```

Figure 3.6: TfIdf Value of Initial Vectorized Review Text

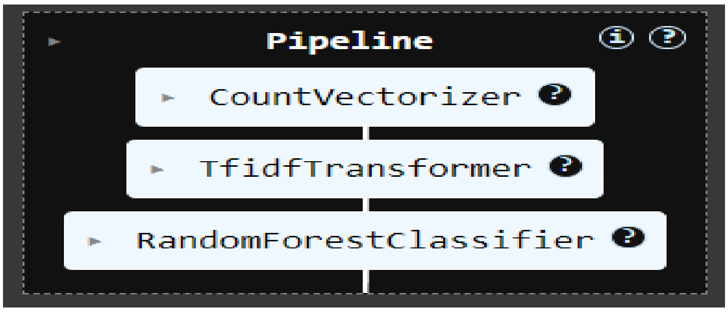

Figure 3.7: Pipeline Trained for RandomForest Classifier

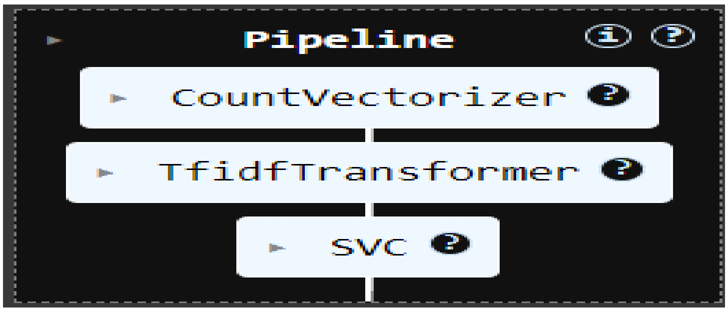

Figure 3.8: Pipeline Trained for Support Vector Classifier

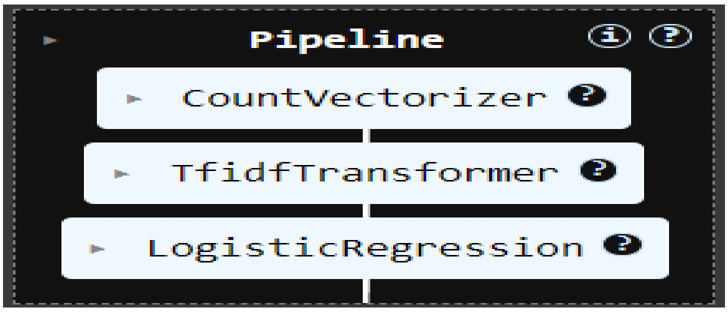

Figure 3.9: Pipeline Trained for Logistic Regression

Chapter 4

Implementation Methodology

4.1 Methodology of Evaluation and Metrics

In this work, evaluation metrics such as **Accuracy**, **Precision**, **Recall**, and **F1-Score** are employed to measure the performance of the three classifiers: Random Forest, Support Vector Classifier (SVC), and Logistic Regression.

- **Accuracy** refers to the number of correct classifications in terms of correctly classified both fake and actual reviews over the total reviews. This is a rough estimate for assessing the goodness of a model, but it comes very handy especially when class distribution is approximately balanced. For unbalanced datasets, though, it won't do very well on its own.

- **Precision** quantifies the proportion of correctly identified positive cases out of all predicted positive cases. This will ensure that actual reviews do not get classified as fake ones, and by this reasoning, decrease false positives.

- **Recall** gauges how well the model is to classify all the actual positive cases so that false reviews are captured and no false negatives occur.

- **F1 Score** is the harmonic mean of Precision and Recall that shows a balanced evaluation metric. It is important for datasets such as Home Appliances Review dataset, where class imbalances may exist, as seen in some similar tasks.

Using these metrics, the study gives a comprehensive performance analysis of the classifiers, so that the best pipeline for detecting fake reviews can be identified.

4.2 Experimental Setup

This section describes the dataset preprocessing, feature extraction, and model training workflows to build and evaluate the three machine learning pipelines.

1. **Dataset Preprocessing:** Raw text data is processed through a number of preprocessing steps, which may include tokenization, stopword removal, punctuation, digit removal, lower casing, stemming/lemmatization of the words. All these preparations are to ensure the quality in feature extraction. Figure 3.2 shows the preprocessing workflow.

2. **Feature Extraction:** Using CountVectorizer and TF-IDF converts the preprocessed dataset to numerical features. CountVectorizer outputs a sparse matrix of word counts, while TF-IDF transforms these counts into weighted scores based on term frequency and inverse document frequency. Features derived from these enable the models to distinguish between important and relatively unimportant terms. Figure 3.5, Figure 3.6 present outputs from CountVectorizer and TF-IDF, respectively.

3. **Model Training Using Three Pipelines:** Three machine learning pipelines were designed.

   - **Random Forest Pipeline:** These comprise the CountVectorizer, TF-IDF, and a Random Forest Classifier for strong ensemble training. Figures 3.7 will help depict this pipeline.
   - **SVC Pipeline:** These constitute the CountVectorizer, TF-IDF, and a Support Vector Classifier for detailed boundary separation in high-dimensional areas. Figures 3.8 will describe this pipeline.
   - **Logistic Regression Pipeline:** Utilizes CountVectorizer, TF-IDF, and LogisticRegression that can be combined to be run as a very simple binary-class classification. Figures 3.9 depicts this pipeline.

   Each pipeline was trained with the dataset, and after training, precision, recall, and F1-scores were calculated to ascertain which model is the better one.

4.2.1 Hardware & Software Environment

- **Software Specifications**
  - Python 3.x Version; Google Colab or Jupyter Notebook.
  - **Main Libraries used:** Scikit learn, numpy, pandas, seaborn, matplotlib, nltk.
- **Hardware Requirement**
  - Intel i5 processor, 8th generation; 8 GB random-access memory.

This environment is capable enough to preprocess data, extract its features, and train the models as well as evaluate their performance.

4.3 Test Application

The system is intended to detect fake reviews on the Home Appliances Review dataset, assisting in fraudulent content detection on e-commerce platforms. It tests pipelines of machine learning trained over labeled data in classifying a review as fake or actual.

1. **Dataset Handling:** Fetched dataset from Kaggle and split it as follows 75% training and 25% testing, so it has enough data trained on and a part left untouched for unbiased testing.

2. **Input Testing:** After the raw dataset has been preprocessed, it is cleaned and fed into the Random Forest, SVC, and Logistic Regression pipelines.

3. **Pipeline Training:** Every pipeline reads the input features which have been extracted by the use of CountVectorizer and TF-IDF, trains a corresponding model, and provides a prediction.

4. **Evaluation:** The application evaluates models on metrics such as Accuracy, Precision, Recall, and F1-Score to establish which is the most precise classifier of false reviews.

Chapter 5

Performance Results and Analysis

5.1 Performance Results

The performance report of three classifiers (Random Forest, Support Vector Classifier, and Logistic Regression) is depicted in the following table:

Model	Accuracy	Precision (Weighted Avg)	Recall (Weighted Avg)	F1-Score (Weighted Avg)
Random Forest Classifier	84.35%	0.85	0.84	0.84
Support Vector Classifier	88.26%	0.88	0.88	0.88
Logistic Regression	86.66%	0.87	0.87	0.87

Table 5.1: Model Performance Comparison

Comparison among the three classifiers—Random Forest, Support Vector Classifier, and Logistic Regression—would present that SVC offers the maximum prediction accuracy in terms of precision, as high as 88.26%. This is primarily because the SVC can look for a best hyperplane for classification in high dimensional feature spaces to distinguish between various classes with very high accuracy for complex data like text.

Logistic Regression closely follows with an accuracy of 86.66%, showing its power in binary classification tasks, especially when features are linearly related. Although Logistic Regression is computationally efficient and interpretable, it does not compete with SVC when the review patterns are not linear and require more complex decision boundaries.

The Random Forest Classifier gets an accuracy of 84.35%, which is the lowest out of the three models. Though this is a flexible and strong ensemble method and it is capable of detecting interactions in features, still it performs slightly less good on high-dimensional data sets and the sparse distribution of features on the tokenized text. Also, the overfitting nature of Random Forest that generally occurs with the text data may have added to its relatively less accurate results in this task.

5.2 Analysis

The results show that all three classifiers have excellent performance in detecting fake reviews, but SVC is the most reliable model because of its margin-based approach, which guarantees a robust separation of classes in high-dimensional spaces. Logistic Regression, though less accurate, is still a good baseline model for datasets with linear feature relationships. Although quite flexible, Random Forest does not handle sparsity and text data complexity well.

Indeed, the results show the further room for improvement of these results with more advanced preprocessing, better feature selection, and fine-tuning hyperparameters. For instance, alternative methods of dimensionality reduction or more advanced NLP features could improve all the three models' performance in more complex cases of datasets.

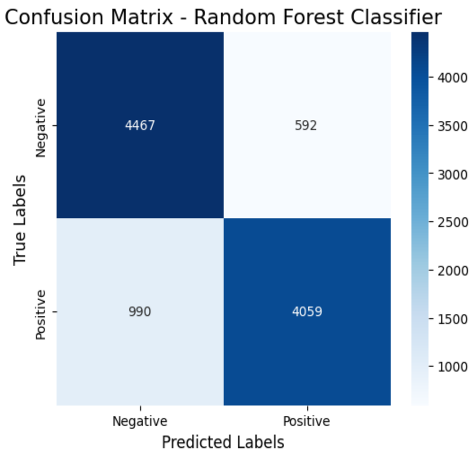

Figure 5.1: Confusion Matrix of Random Forest Classifier

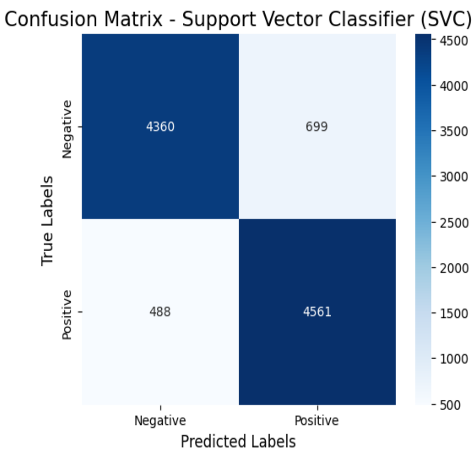

Figure 5.2: Confusion Matrix of Support Vector Classifier

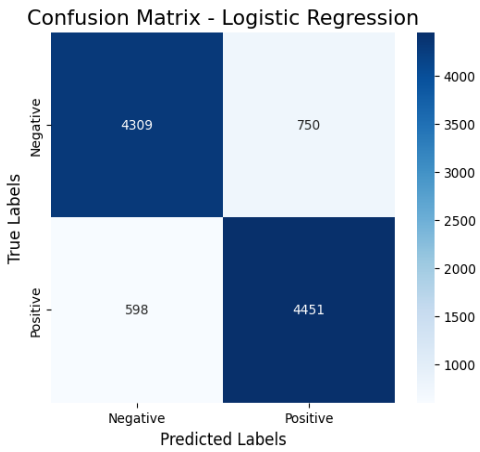

Figure 5.3: Confusion Matrix of Logistic Regressor

Chapter 6

Conclusion and Future Work

6.1 Conclusion

The work presents the performance of machine learning pipelines—Random Forest, Support Vector Classifier (SVC), and Logistic Regression—over the task of fake review detection. Among them, the SVC pipeline has a superior performance, with a precision of 88.26%, due to its capability to identify optimal hyperplanes in high-dimensional feature spaces. Logistic Regression is a strong baseline for data that can be linearly separated. Random Forest tends to be a bit worse than this baseline, as overfitting happens with sparse, high-dimensional text data. The application of robust preprocessing methods and TF-IDF greatly improved the overall classification performance. Therefore, the results prove SVC to be a good solution for high-accuracy fake review detection applications.

6.2 Future Work

Further enhancement of the system in proposed directions are:

1. **Advanced Neural Networks:** This will include deep learning models such as Long Short-Term Memory (LSTM) networks or transformers that can better capture the nuances of contextual and sequential information in text, hence improving detection accuracy.

2. **Multisourced Data Curation:** It would enhance the dependability of the model by augmenting the dataset through multiple sources including blog, social media, and e-commerce platforms from review and not through the biased data set coming from one source.

3. **Deep Embedding Techniques:** BERT and more deep embedding techniques can advance the contextual understanding in words inside word embeddings and better discriminate between actual reviews versus forged ones. Semantic understanding through

BERT's pretraining in language is important while making the distinction between fakes and actual reviews.

4. **Hyperparameter Optimization:** Extra Tools to Systematically Experiment with a High Number of Possibilities of Hyperparameter Configurations and determining the best Possible configurations for a Machine Learning Model's GridSearchCV allows for testing lots of parameter configurations of interest, like what kernel to use in an SVC or how many trees should be fit into a Random Forest model. This improves generalization. Other methods that improve these techniques for better tuning include the AutoML framework and Bayesian optimization, among others.

These will improve the performance of model-based reviewers in discovering fake reviews with greater accuracy and scalability and open doors toward more reliable and robust applications against misinformation online.

Bibliography

[1] N. Balakrishnan, S. Chandru, T. Priyadharshini *et al.*, "Decoding authenticity in online review using machine learning," in *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)*. IEEE, 2024, pp. 421–424.

[2] Z. Shunxiang, Z. Aoqiang, Z. Guangli, W. Zhongliang, and L. KuanChing, "Building fake review detection model based on sentiment intensity and pu learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 6926–6939, 2023.

[3] P. Devika, A. Veena, E. Srilakshmi, A. R. Reddy, and E. Praveen, "Detection of fake reviews using nlp &sentiment analysis," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2021, pp. 1534–1537.

[4] A. Thilagavathy, P. Therasa, J. J. Jasmine, M. Sneha, R. S. Lakshmi, and S. Yuvanthika, "Fake product review detection and elimination using opinion mining," in *2023 World Conference on Communication & Computing (WCONF)*. IEEE, 2023, pp. 1–5.

[5] L. Dinesh, G. Kaviya, N. Kavya, A. Kousalya, and N. Lavanya, "Product reviews on opinion mining using nlp techniques," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2022, pp. 952–957.

[6] O. Dhamdhere, A. Dhanwate, A. Kumbhar, and M. Singh, "Amazon's fake review detection using support vector machine."

[7] S. Biruntha, B. Ashwin, K. Padmasankar *et al.*, "Distinguishing reviews through sentiment analysis using machine learning techniques," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1. IEEE, 2022, pp. 1713–1719.

[8] A. M. Elmogy, U. Tariq, M. Ammar, and A. Ibrahim, "Fake reviews detection using supervised machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[9] D. Zhang, W. Li, B. Niu, and C. Wu, "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information," *Decision Support Systems*, vol. 166, p. 113911, 2023.

[10] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Physica A: statistical mechanics and its applications*, vol. 540, p. 123174, 2020.

[11] C. Singh and S. Tanwar, "Fake review identification using machine learning," in *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*. IEEE, 2024, pp. 582–587.