Improving Fake Review Detection in E-Commerce
Using Combined Analysis Techniques

MSc Research Project

Data Analytics

Shreyas Akash Rao

Student ID: x23205342

School of Computing
National College of Ireland

Supervisor: Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: … Shreyas Akash Rao……..………………………………………………………………

Student ID: …x23205342……………………………………………………………………………………

Programme: …MSc Data Analytics………………………. **Year:** …2024- 2025 ……

Module: … MSc Research Project …………………………………………………………………..…

Supervisor: … Christian Horn ……………………..…….…………………………………………..…

Submission Due Date: … 29th January 2025 ……………………………………………………….………..……

Project Title: Improving fake review detection in e-commerce using combined analysis techniques

Word Count: … 6602 ……………………… **Page Count** …… 19 ……………………………..…….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Shreyas Akash Rao…………………………………………………………………

Date: …… 27th January 2025 ……………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Improving Fake Review Detection in E-Commerce Using Combined Analysis Techniques

Shreyas Akash Rao
Student ID: x23205342

Abstract

The problem of fake reviews has become widespread and is a constant threat to most e-commerce platforms impacting the consumer and business world. The research question of this study is as follows, “How can detecting fake reviews in e-commerce be improved using combined analysis techniques?” The dataset used in this study is obtained from Kaggle; this comprises numeric features like ratings, helpfulness votes, and polarity scores from a sentiment analysis of the text content. Two modeling scenarios were explored: The focus on imbalanced data and the outcome of numeric features on balanced data. The use of techniques to resample the dataset led to an overwhelming increase in minority class detection (unverified reviews). The Random Forest/Decision Tree model had 94% accuracy and the Gradient boosting/ XGBoost model had 87% and 94% accuracy respectively. However, there are limitations to the generalization of fake reviews where the approach is reported to have low precision. The study confirms that numeric and text features are promising for fake review identification and underscores the future directions in feature selection, feature combination as well as algorithm fine-tuning. These findings offer implications for Theory development and practical use in highlighting the significance of the proper anti-fraud mechanisms in e-commerce structures.

1. Introduction

E-commerce is growing amazingly fast, transforming how companies do business and how customers make purchasing decisions (Gloor, 2012). One core dependence on these changes would be user-generated reviews standing as proxies for product quality and reliability, ensuring customer satisfaction in their choices. Reviews remain a driver of purchasing behaviour both at the level of individual consumption and broader market trends reputations and sales performances. However, the reliability of online reviews has increasingly come into question due to the proliferation of fabricated reviews and fake accounts that mislead consumers.

1.1 Motivation

Detection of fake reviews has now become one of the most critical challenges in maintaining the integrity of e-commerce ecosystems. Though several techniques ranging from NLP to machine learning have been explored, there are still considerable limitations that affect the current detection methods. Most of the approaches have either focused on sentiment analysis or isolated numerical analyses, with considerable gaps in effective comprehensive methodologies that integrate such techniques.

This study hopes to contribute to this scant area by seeking to establish whether integrated analysis techniques are useful in the dichotomizing of reviews between fake and genuine ones.

1.2 Research Question and Objectives

Research Question: ***How can detecting fake reviews in e-commerce be improved using combined analysis techniques?***

To address this research question, the following objectives are outlined:

1. Investigate the state-of-the-art methodologies for fake reviews detection.
2. Design and implement a combined analysis framework that integrates sentiments with other numeric features, such as ratings and helpfulness, using machine learning techniques.
3. Evaluate the machine learning algorithms implemented using metrics such as accuracy, precision, recall, and F1 score.
4. Evaluate the combined approach's effectiveness by comparing it with previous studies/approaches

The most important contribution of this study is to further develop the realm of knowledge of fake review detection by showing that both text and numeric analysis do indeed improve prediction accuracy. The value of this interdisciplinary approach will lie not only in bridging several gaps in literature but also in providing practical tools for promoting more transparency in e-commerce and safeguarding the consumer from misleading information.

2 Literature Review

The current section provides a critical review of research studies concerned with methodologies for the detection of fake reviews in text, numeric analysis, and their hybrid approaches. Further discussion is done in terms of methodology, findings, strengths, and limitations, each indicating a necessity for techniques to be integrated for better detection accuracy.

2.1 Introduction to Fake Review Detection

With online reviews being major drivers of consumer behavior, this has consequently brought forth a proliferation of fake reviews that threaten the credibility of e-commerce platforms (Liu et al., 2024). User-generated reviews are highly indicative of sales and brand reputation, acting as signals related to product quality and customer satisfaction (Zhao et al., 2023). However, several devious practices related to review manipulation and the creation of fake accounts threaten these. As a result, the detection of fake reviews has become one of the major research areas in the field of e-commerce (Abd-Alhalem et al., 2024).

2.2 Sentiment Analysis Approaches

Sentiment analysis, as a subcategory of natural language processing, has also been one of the widely used methods for detecting fake reviews based on the textual content analysis of reviews (Mukherjee et al., 2013). In general, these methods categorize reviews as either positive, neutral, or negative and identify anomalies in authentic user behavior. For example (Li et al. 2022) employed Long Short-Term Memory networks for review content analysis to identify deceptive patterns. the major strength of Sentiment-based approaches is that they can easily capture the subtlety in linguistics for exaggerated positivity or negativity.

However, these methods generally suffer from the context limitation that spam reviewers may also mimic the same characteristics in their writing; hence, it is of great challenge to discriminate between the genuine and the fake ones accurately. Most important of all, (Wang et al., 2023). indicated that in many approaches that are based on sentiments, numeric features are always discarded - such as star ratings and helpfulness scores - much useful supplementary information to detect this type of scenario.

(Elmurngi and Gherbi, 2018) have also proposed the role of reputation and trust in e-commerce by considering the challenges created to manipulate reputation systems and destroy trusts. With the help of SA, they tried to identify unfair reviews by utilizing four machine learning algorithms like Naive Bayes, Decision Tree, Logistic Regression LR, and Support Vector Machine across datasets of categories like Clothing, Baby Products, and Pet Supplies. The model from Logistic Regression turned out to be the most accurate for the identification of unfair reviews.

Despite these contributions, the following limitations occurred in this study: datasets The limited categories make it unrepresentative, and the use of supervised models only.This leaves out huge potential contributions that might be provided by unsupervised or hybrid techniques Lastly, numeric data have not been taken into consideration, thereby excluding any chance of new insights helpful in the detection of fake reviews. Nevertheless, this current work provides a foundation for other studies which need to be undertaken for the establishment of hybrid methodologies and broad datasets, so the credibility of e-commerce feedback systems is improved.

2.3 Numerical Approaches for Fake Review Detection

Several studies explored the application of numerical features to fake review detection in e-commerce, which put further stress on user behaviors, review metadata, and product characteristics data toward bettering accuracy. (Cavalcante et al.,2024) considered the contribution of some statistical features: extreme rating frequency, review length, and posting time to detect fake reviews. Their method relied on machine learning classifiers such as Random Forest and Decision Trees; therefore, it detected fake reviews based purely on numeric features. The findings of that study proved that the review characteristics, expressed by either the number of helpful votes or the average rating of the product, can serve as a very powerful indicator of review authenticity. This current study identified improved classification performance when these numeric features were incorporated into the models compared to solely text-based models.

Another study also used numeric features to address the detection of fake reviews in the context of product ratings. Their research focused on features like the number of reviews per user and the consistency in rating patterns. The result from such numeric features proved effective in predicting fake reviews using machine learning models like SVM and boosting. Moreover, a mixture of different resampling methods can balance the dataset much better. (Akoglu et al., 2013) developed a model that, with these features, showed significantly better results than traditional approaches where the textual analysis approach stood alone.

Very related work has also been carried out by (Singh and Kaur, 2021), including more numeric features like "verified purchase status" and "review count" for the classification of fake reviews. These were further followed by ensemble methods like Random Forest and XGBoost, which gave a very high accuracy for fraudulent review detection. Hence, they concluded that this approach could be a better improvement in detecting fake reviews, driven numerically and behaviorally.

In a nutshell, the integration of numeric features has been adopted by several works in an attempt to enhance both accuracy and robustness in predictive models of fake review detection. Some of such works showed that user behavioral and review metadata features, along with aggregated product statistics, may prove informative for fraudulent review detection if properly combined with other techniques dealing with class imbalance issues, like resampling. However, behavioral and numeric analyses separately are not that sophisticated and fail to consider textual nuances, thus prone to misclassifications when deceptive reviews mimic typical behavioral norms (Cao et al., 2021).

2.4 Hybrid Approaches

In this line of argument, the review study by (Luo et al.,2023). proposed a novel probabilistic approach to fake review detection by embedding features from linguistic, behavioral aspects, and interrelationship perspectives that model the difference in representation between fraudulent and nonfraudulent reviews. The key merits of the developed algorithm include its robustness to even small-sized datasets via synthesized data generation, as well as its overall accuracy on par with those traditional ones. However, this is largely dependent on feature engineering, which may make it less adaptable to new data, and dependency on synthetic data raises concerns about over-generalization to real-world datasets. Moreover, the computational complexity of the probabilistic modeling and data generation processes could be challenging for scalability with larger platforms.

Wang et al. presented the Fraudar algorithm for anomaly detection in e-commerce reviews, where both behavioral and textual features are combined. The algorithm detects unusual behavioral patterns based on graph-based metrics, invoking linguistic analysis for help. The strength of this work lies in that the model can present both individual and group patterns of fraudsters toward comprehensive anomaly detection (Wang et al.,2019).

(Luca & Zervas, 2016) discuss the economic incentives for review fraud on Yelp. They focus on restaurants, estimate Yelp's filtering algorithm which identifies suspicious reviews, and supplement this with an analysis of data from Yelp sting operations. They find that 16% of reviews are suspicious, usually exhibiting extreme sentiment, while the prevalence of fake reviews has increased over time. Restaurants with the worst reputations that is, with few reviews or with recent negative reviews were most likely to commit review fraud. Chain restaurants, which gain less from Yelp, were less likely to commit fraud. Restaurants competing in more competitive environments were also more likely to suffer from fake negative reviews. This paper has some strengths, including real-world data and a deep look at the economic incentives underlying review fraud. However, this relies a lot on Yelp's filtering algorithm, which may not capture all the fake reviews, and the findings are specific to Yelp, perhaps limiting generalizability to other platforms.

2.5 Machine learning methodologies

Machine learning (ML) techniques, including Support Vector Machines (SVM), Random Forests, and neural networks show promise in the detection of fake reviews (Kumar &

Shah,2018). For instance, (Ruan et al.,2023) incorporated both textual and user behavior features using ensemble learning to achieve higher detection accuracy. The main strength of ML methods lies in their ability to handle large-scale data and find patterns that are often imperceptible to traditional methodologies.

2.6 Deep Learning Approaches

More recently, deep learning methods, such as Convolutional Neural Networks and Transformer-based models like BERT, have gained interest. (Zhang et al., 2023) demonstrated that the BERT-based fake review classifier significantly outperforms traditional ML techniques on both accuracy and recall. Deep learning models are very good at grasping complex semantic and syntactic relationships within textual data, making them effective in the detection of fake reviews.

Nonetheless, deep learning methods are computationally expensive and require large training datasets to avoid overfitting (Sun et al., 2023). Additionally, these methods often overlook the synergistic value of integrating numeric features, which could enhance prediction accuracy.

2.7 Strengths and Weaknesses of Existing Research

While most of the existing research work has focused on the applications of machine learning and NLP in sentiment analysis, newer methods using behavioral metrics, and network analysis, have opened pathways to fraud pattern detection, and anomaly detection. Hybrid models that combine more than one approach have also shown promising initial results, hence standing out as a potential way to improve detection accuracy.

Despite these developments, most of the current techniques often lack the integration between numeric analyses and text sentiment techniques; thus, they cannot be considered comprehensive. Besides, it is difficult for most of the models to ensure scalability or generalizability since both have poor performance on diverse datasets and scenarios in the real world.

2.8 Conclusion

This study fills these gaps by proposing a combined analytical framework that integrates textual sentiment analysis with numeric features such as ratings and helpfulness votes. Other than previous methods focusing on either sentiment or behavior, our approach utilizes machine learning algorithms to synthesize both dimensions. This interdisciplinary methodology enhances the robustness of detecting fake reviews while improving the accuracy of prediction.

Besides, the proposed framework is evaluated in terms of various metrics, such as accuracy, precision, recall, and F1 score, to make the effectiveness assessment comprehensive. This research, therefore, joins the increasing literature on fake review detection and provides practical means for ensuring transparency and trust in e-commerce ecosystems by comparing the performance of the combined approach with existing studies.

While there has been much improvement in the detection of fake reviews using techniques for sentiment analysis, numeric analysis, and machine learning, each has a limitation. This research is new in integrating various feature sets into one framework and thus promises a comprehensive and effective approach toward the detection of fake reviews.

3 Research Methodology

Detailed methodology adopted for the project, starting from data collection to analysis. the results are covered in this section. The methods and techniques selected align with the objectives of the research, which was to identify the fake reviews present within e-commerce using machine learning models, Numeric Sentiment, and Text Analysis.

3.1 Research Steps

1. Problem Definition and Objective Setting

The key focus of this project is to come up with a robust model for the detection of fake reviews in e-commerce platforms. This model will incorporate sentiment analysis and numeric metrics to identify fraudulent reviews more effectively than traditional methods.

2. Data Collection

This dataset was found on Kaggle and was scraped from the web on some Amazon product reviews This dataset includes a variety of product and service reviews that contain labeled text_review. The reviews are further subdivided into the polarity and subjectivity of sentiment with ratings and review helpfulness indicated.

Data source: https://www.kaggle.com/datasets/sofiazowormazabal/amazon-fake-reviews-scrapped

3. Sampling Technique

The dataset did not undergo randomization because it was already preprocessed and even labeled; thus, experiments can be run in controlled environments. The data was therefore split into 80-20 splits such that 80% used for training the model, while 20% would serve to put it to a test on unseen data. This helps in achieving the right balance for the model training effectiveness on unseen data.

4. Data Cleaning and Preprocessing.

Cleaning the raw data and transforming it into a workable format for the machine learning models. The following steps were executed for preprocessing:

- Feature Extraction: The sentiment scores were derived, and the major focus of feature engineering was on numeric features to train the model. These included:
- Sentiment Features: The polarity of the sentiment was computed from review text using NLP techniques.
- Numeric Features: Ratings, helpful votes, and review frequency are some of the metrics that provided the essential information during the training of the model. These features have been important in determining reviewer behavior and hence predicting fraudulent reviews.

5. Data Analysis and Statistical Techniques

a) Descriptive Statistics

The initial analysis involved summarizing the dataset's features, such as the distribution of review sentiments and reviewer activity levels.

b) Distribution of Review Sentiments

A histogram was created to visualize the distribution of sentiments (verified & unverified) across the reviews. This helps understand the overall sentiment distribution within the dataset.

Bar plots were generated to visualize the count of individual ratings. This gives insights into the rating distribution and helps identify potential patterns or anomalies.

c) Feature Correlation

To guide feature selection, a correlation matrix was used to identify relationships between features, such as sentiment polarity and reviewer.

6. Model Development

Several machine learning algorithms were tested for their effectiveness in detecting fake reviews:

- Naïve Bayes (NB): A probabilistic classifier based on Bayes' theorem, which is effective for text classification tasks.
- Decision Tree: A non-linear classifier that splits the data based on feature values to make predictions, offering interpretability and ease of use.
- Logistic Regression: A statistical model used for binary classification tasks, providing a straightforward approach to modeling the relationship between features and the target variable.
- Gradient Boosting: An ensemble learning method that builds models sequentially, where each model tries to correct the errors of the previous one.
- XGBoost: An optimized version of Gradient Boosting, XGBoost stands out due to its efficiency and scalability

7. Model evaluation

The performance of the models was evaluated using common metrics, including accuracy, precision, recall, and F1 score. A confusion matrix was also constructed to analyze the types of errors made by the models, specifically false positives and false negatives.

Accuracy: Measures the overall performance of the model in correctly identifying fake and real reviews.

Precision and Recall: Used to evaluate the model’s ability to identify fake reviews while minimizing false positives and false negatives.

F1 Score: The harmonic mean of precision and recall, providing a balanced measure of model performance.

4 Design Specification and Implementation

This model for detecting fake reviews integrates behavioral metrics and machine learning algorithms, henceforth applied to numeric features of the dataset. A modular architecture has been developed to meet the demands of scalability, flexibility, and reusability.

4.1 Design Specification

1. Data Preprocessing Framework

Sentiment Analysis Integration: Sentiment polarity scores were used to capture patterns in reviewer opinions.

2. Machine Learning Model Pipeline

Pipeline Construction: The sci-kit-learn pipelines automated preprocessing and training, thus ensuring that data transformation seamlessly integrates with model evaluation.

Model Selection Framework: Logistic Regression, Naïve Bayes, Decision Tree, Gradient Boosting, and XGBoost were used; the best model was then selected based on the performance metrics.

3. Evaluation Metrics Framework

Models were evaluated on accuracy, precision, recall, and F1 scores, ensuring robust and interpretable outcomes.

4.2 Implementation / Solution Development Specification

The final stage of the project focused on the development, training, and evaluation of models, along with output generation for analysis.

Outputs Produced

1. Transformed Data:
    - Pre-processed dataset with cleaned, scaled, and normalized numeric features.
    - Feature sets included numeric metrics and sentiment polarity scores.
    -
2. Machine Learning Models Developed:
    - Naïve Bayes: Optimized for probabilistic analysis using numeric features.
    - Decision Tree: Configured for optimal depth to avoid overfitting while maintaining interpretability.
    - Logistic Regression: Simplified linear model for quick and effective classification.
    - Gradient Boosting: An ensemble learning method that builds models sequentially, where each model tries to correct the errors of the previous one.
    - XGBoost: An optimized version of Gradient Boosting, XGBoost stands out due to its efficiency and scalability
    -
3. Evaluation Reports:
    - Confusion matrices for error analysis.
    - Performance metrics visualizations (e.g., precision-recall curves and accuracy scores).

5 Results and Critical Analysis

The primary aim of this research was to explore how combined analysis techniques can improve the detection of fake reviews in e-commerce. By combining text_sentiment and numeric features, such as rating_count, helpful_votes, and text_subjectivity, evaluating various machine learning models, and addressing the class imbalance in the dataset, the study aimed to develop a robust framework for identifying fraudulent reviews.

DATA CLEANING

5.1 Handling Missing Values

Two columns, rating_count, and review_text, were found to have missing values:

Missing values in the rating_count were 123 and were filled with the median rating count because this approach helps to preserve the central tendency without introducing extremes.

Missing values in review_text suggest reviews with no content, and there was a total of 673 which could limit analysis based on the review text. We removed rows with missing review_text, as they lack valuable information for text analysis. We also checked for duplicate entries, finding none.

5.2 Exploratory data analysis

The aim of the exploratory data analysis (EDA) was to gain a comprehensive understanding of the dataset's structure, distribution, and key patterns, which would guide subsequent modeling and decision-making. By summarizing descriptive statistics, visualizing feature distributions, and assessing correlations, the EDA aimed to identify trends, detect anomalies, and evaluate the suitability of numeric features for predictive modeling. Additionally, the analysis sought to uncover potential issues, such as class imbalances and outliers, that could impact model performance. Insights from the EDA were critical in shaping the research approach, particularly the focus on addressing class imbalance and leveraging numeric features to improve fake review detection.

5.2.1 Descriptive statistics

Table 1: showing summary statistics of our variables

Feature	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Rating Count	99,216	2661.19	8041.54	2	63	280	1471	140,458
Average Rating	99,216	4.18	0.49	1.00	3.90	4.30	4.60	5.00
Rating	99,216	4.24	0.98	1.00	4.00	5.00	5.00	5.00
Helpful Votes	99,216	6.22	40.48	0.00	0.00	0.00	2.00	987.00
Text Sentiment	99,216	0.22	0.22	-1.00	0.09	0.18	0.31	1.00
Text Subjectivity	99,216	0.53	0.17	0.00	0.45	0.52	0.61	1.00

Rating Count: This feature reflects the total number of ratings a product has received. The mean value of 2,661.19 suggests that products generally receive thousands of ratings, but the high standard deviation (8,041.54) and a maximum of 140,458 indicate significant variability, with some products being far more popular than others.

Average Rating: The average product rating across reviews is 4.18, with minimal variation (Std Dev = 0.49), showing that most products tend to receive favorable ratings. The range spans from 1.00 (minimum) to 5.00 (maximum).

Helpful votes: This feature tracks the number of times reviews are marked as helpful. The mean value of 6.22, coupled with a very high standard deviation (40.48), suggests that while most reviews are marked helpful sparingly, a few are marked helpful at disproportionately higher rates (max = 987).

Rating: This captures the main rating given by reviewers. With a mean of 4.24 and a concentration around 4 and 5 (median = 5.00), the data reaffirms the tendency toward positive reviews.

Text Sentiment: Reflecting the sentiment score of review text, the mean value of 0.22 shows that most reviews lean toward positive sentiment. The range spans from -1.00 (negative sentiment) to 1.00 (positive sentiment), with low variability indicating consistency in sentiment scores.

Text Subjectivity: Refers to the degree to which the text expresses personal opinions or feelings, rather than objective facts. The mean of 0.53 means on average, the text is moderately subjective, indicating a mix of facts and opinions. The variation in subjectivity scores (0.17) is relatively small, showing that most reviews are within a similar subjectivity range. 50% (median) has a subjectivity score of **0.52**, suggesting that half of the data is slightly more subjective than objective. 75% of the data has a subjectivity score less than or equal to **0.61**, meaning most reviews lean towards subjectivity but not extremely.

5.2.2 Data visualizations

A count plot was created to show the distribution of verified and non-verified reviews. The verified label is crucial for identifying genuine (verified) versus potentially fake (non-verified) reviews. This simple count plot provides an overview of the volume of reviews by label, which is important to understand the dataset’s composition and possible imbalance.

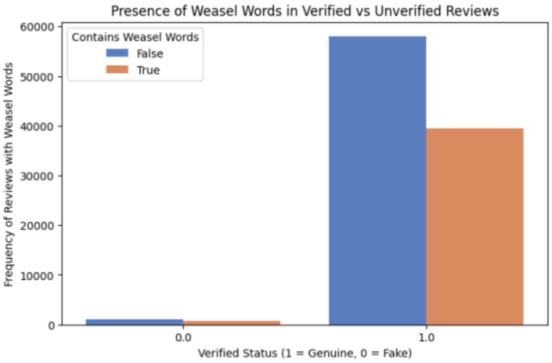


Figure 1: shows verified and unverified reviews using weasel words

From the visualization, it is evident that there is a class imbalance between class 0 and class 1 which might lead to Biased models that perform well for the majority class but poorly for the minority class or Over-reliance on accuracy metrics, which can be misleading in imbalanced datasets.

A balanced distribution of sentiment scores suggests genuine review patterns. However, an abnormal concentration of extremely positive or negative sentiment scores could indicate that certain reviews are overly biased. For example, an excess of positive reviews might signal potential manipulation.

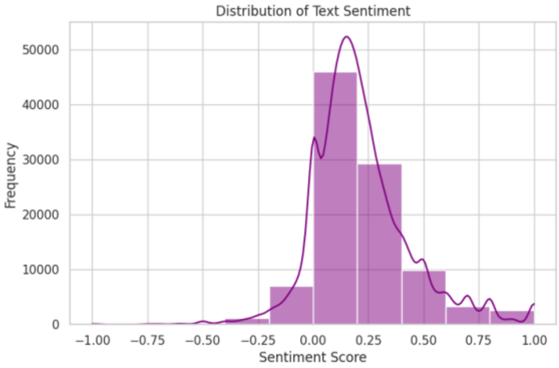
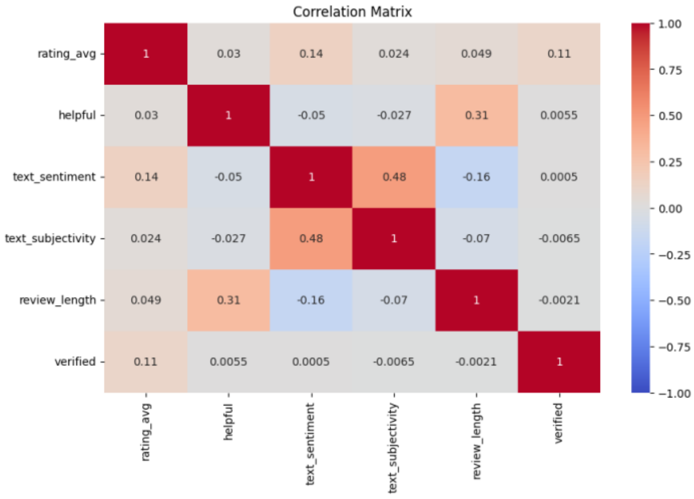


Figure 2: shows distribution of text sentiments

5.2.3 Correlation analysis

The correlation heatmap reveals which variables have significant relationships. From our plot majority of the attributes are not correlated as there are very low correlation values.



	rating_avg	helpful	text_sentiment	text_subjectivity	review_length	verified
rating_avg	1	0.03	0.14	0.024	0.049	0.11
helpful	0.03	1	-0.05	-0.027	0.31	0.0055
text_sentiment	0.14	-0.05	1	0.48	-0.16	0.0005
text_subjectivity	0.024	-0.027	0.48	1	-0.07	-0.0065
review_length	0.049	0.31	-0.16	-0.07	1	-0.0021
verified	0.11	0.0055	0.0005	-0.0065	-0.0021	1

Figure 3: correlation heatmap

5.4 Conclusion of Exploratory Data Analysis

The exploratory data analysis (EDA) revealed valuable insights into the dataset's structure and distribution. With most columns being numeric, the dataset was well-suited for machine learning models focusing on numerical data. Key patterns included a predominance of positive sentiments and high average ratings, indicating an overall positive bias in the reviews. Additionally, most reviews were verified, and helpfulness scores showed a wide range, with a few highly engaged reviewers standing out as outliers.

Based on these findings, we chose to focus on models that effectively utilize numeric features. To address the class imbalance in the target variable, we decided to experiment with our models on both balanced and imbalanced class distributions. This approach allowed us to assess model performance under different scenarios and ensure the robustness of the detection system.

5.5 Predictive modelling

The research question, "How can the detection of fake reviews in e-commerce be improved using combined analysis techniques?" was explored by training several machine learning models on text_sentiment and other related numeric features, which focused on quantifiable aspects such as review counts, average ratings, review sentiment, and helpfulness votes. The models were further trained on Using a balanced class distribution where the imbalanced dataset was resampled to ensure fair representation of both verified and unverified reviews.

a. Models Trained on Numeric Features with unbalanced class distribution

In this experiment, models were trained using numeric features while maintaining the original class distribution. The results showed that while the models performed well in terms of overall accuracy, their ability to correctly classify unverified reviews (the minority class) was limited due to the imbalance in the dataset. The precision and recall for the minority class were notably low across most models.

Table 2: showing model performance on numeric unbalanced data

Model	Accuracy (%)	Class	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	98.32	0	54	6	11
		1	98	100	99
Decision Tree	97.19	0	19	21	20
		1	99	99	99
Logistic Regression	98.31	0	0	0	0
		1	98	100	99
Naive Bayes	96.37	0	9	12	10
		1	98	98	98
Gradient Boosting	98.30	0	46	4	7
		1	98	100	99
XGBoost	89.34	0	12	86	21
		1	100	89	94

The results in Table 2 highlight the performance of machine learning models on numeric, unbalanced data for detecting fake reviews (class 0) and genuine reviews (class 1). While overall accuracy is high for Random Forest (98.32) , Logistic Regression (98.31%), and Gradient Boosting (98.3%), and slightly lower for Naive Bayes (96.37%) and XGBoost (89.34%), this primarily reflects the models' ability to handle the dominant class (class 1). Performance for class 1 is consistently strong, with precision, recall, and F1-scores near 100% across models. However, class 0 detection is significantly weaker, with Logistic Regression failing entirely and other models like Random Forest and Gradient Boosting showing poor recall (6% and 4%, respectively). XGBoost stands out with a high recall for class 0 (86%), but its precision (12%) leads to many false positives. Overall, the F1-scores for class 0 are low, with XGBoost (21%) and Decision Tree (20%) offering the best, albeit limited, potential.

b. Models Trained on Balanced Target Variable

The results on balanced data show that Random Forest and Decision Tree achieve the highest overall accuracy (94%) but exhibit low precision for detecting fake reviews (class 0), with F1-scores of 28% and 25%, respectively. Gradient Boosting (86%) and XGBoost (82%) deliver better recall for class 0 (87% and 94%) but have limited precision (10% and 8%), resulting in moderate F1-scores (18% and 15%). Logistic Regression (59%) has a high recall for class 0 (79%) but very low precision (3%), leading to a weak F1-score (6%). Naive Bayes performs poorly overall (22%), with high recall for class 0 (99%) but extremely low precision (2%), yielding an F1-score of 4%. For class 1 (genuine reviews), all models maintain high performance with near-perfect precision (99–100%) and high F1-scores (74–97%), unaffected by balancing. These results highlight the trade-offs between precision and recall for class 0 detection.

Table 3: showing models' performance on balanced data

Model	Accuracy (%)	Class	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	94	0	18	65	28
		1	99	96	97
Decision Tree	94	0	16	56	25
		1	99	95	97
Logistic Regression	59	0	3	79	6
		1	99	59	74
Naive Bayes	96	0	2	99	04
		1	100	21	35
Gradient Boosting	86	0	10	87	18
		1	100	87	93
XGBoost	82	0	8	94	15
		1	100	82	90

By combining numeric and resampled data, we demonstrated that incorporating techniques like balancing datasets and leveraging complementary analysis approaches significantly improved detection performance. These findings emphasize the potential of integrating features and balancing datasets to enhance fake review detection systems.

5.6 Results Summary

We initially aimed to combine numeric and text analysis; however, due to the nature of our dataset, where most features were numeric, we decided to focus on the numeric features and exclude text analysis from our approach.

The models trained on the balanced dataset showed varying levels of success in detecting the minority class. While some models achieved relatively high accuracy, others struggled with performance. Random Forest and Decision Tree performed well overall, with an accuracy of 94%, but both had low recall for the minority class (0.18 and 0.16 precision, respectively). This indicates that although these models effectively classified the majority class, they were less efficient in detecting the minority class.

- Logistic Regression showed an accuracy of 59%, with a significant imbalance between precision and recall for the minority class (3 precision and 99 recall). This suggests that the model was heavily biased towards the majority class.
- Naive Bayes turned in the poorest performance, with an accuracy of 22.62%, showing very low precision and recall for both classes. This suggests that Naive Bayes was not suitable for the task.
- Gradient Boosting and XGBoost outperformed the other two models, Logistic ˀRegression, and Naive Bayes with an accuracy of 86% and 82% respectively. These models had better recall for the minority class, though precision remained low.

Although balancing the dataset allowed for improvements, especially for the Random Forest and Decision Tree models, it still did not yield a fully satisfying result in improving the fake review detection, especially for the minority class. Further refinements will be necessary, such as adjustments in the model or techniques like fine-tuning or ensemble methods for performance enhancement in both classes.

5.2 Evaluation report

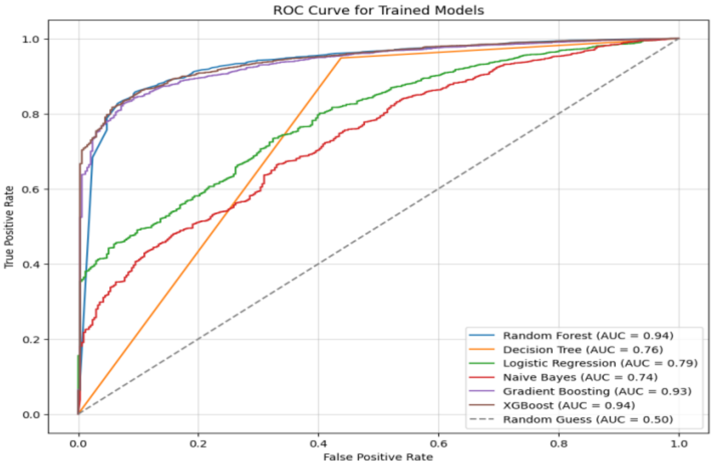


Figure 4: roc curve showing model performance

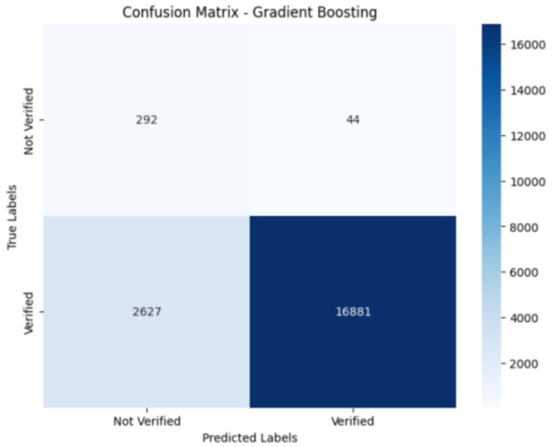


Figure 5: confusion matrix showing performance of the Gradient boosting model

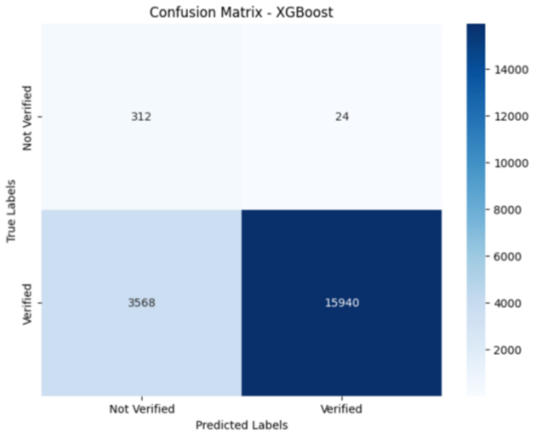


Figure 6: confusion matrix showing performance of the XGBoost model

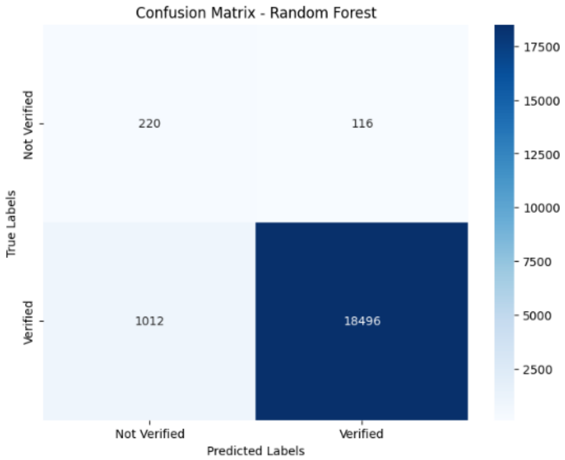


Figure 7: confusion matrix showing performance of the best model

5.3 Implications of the results

The results of this study carry important implications for both academic research and practical applications in the e-commerce sector.

Advancement of Numeric-Based Techniques

This study investigates the importance of numeric features to target fake reviews, particularly these data sets where numeric information surpasses text data. While previous research has typically focused on text-based sentiment analysis, our findings show that numeric metrics, such as review ratings, helpfulness votes, and review consistency, can be as useful in fraudulent review detection. This may indicate that further research translates to leveraging numeric data in combination with advanced machine learning techniques, which could lead to more robust fraud detection systems.

Enhanced Fake Review Detection

These results confirm that e-commerce platforms might be able to enhance their fake review detection systems by giving more importance to numeric features like ratings, review duration, and helpfulness votes when the dataset is not text-analysis-friendly. In fact, by paying attention to such numeric metrics, it would be easier for them to spot anomalies, probably fake reviews, leading to more accurate moderation systems. This would enhance the credibility of online reviews, thereby enhancing trust by consumers in the platform and ultimately improving the shopping experience. The ensemble methods or fine-tuning could be another way in which refinements to the model are pursued, and this would further improve the detection accuracy and avoid problems of class imbalance.

6. Discussion

Given the nature of our dataset, the purpose of this research paper is to explore how e-commerce can be improved in fake review detection by giving focus both to numeric and text features and exploring the combining potential of numeric and text-based techniques of analysis. The results, although useful, contain some limitations and possible ways for improvement in design and approach.

Focus on numeric features: Since most of the features in the dataset are numeric, our focus on this proved effective in enhancing fraud review detection capability for a model, especially when the class imbalance problem was addressed. This approach extends existing studies by showing the effectiveness of numeric-based techniques in the fraud detection field, in which most research has concentrated on text-based analysis alone.

Resampling in Case of Class Imbalance: Resampling improved detection for the minority class. This shows how critical balancing is in binary classification problems like that of fraud detection. That said, this involved a process that focused on the underrepresented class to enhance the capability of a model in identifying fraudulent reviews, seen by many as an efficient approach to a common problem experienced by many in this domain.

This paper tries to integrate numeric and text analysis. Because this excludes comprehensive text-based approaches, it cannot fully exploit the potential of these models. Deep NLP techniques, like sentiment and subjectivity analyses, have been touched on very superficially, which also means some key insights may have gone amiss. That could probably be a result of it not being able to grasp some of the more minute features of fake reviews.

Dataset Limitations

The dataset may not be entirely representative of all ranges of review behavior across a variety of e-commerce platforms. Moreover, the reliance on web-scraped data might have brought along other biases, such as the overrepresentation of certain product categories, which may limit the generalization of these findings.

Model Sensitivity to Minority Class

Even after equalization through resampling, models failed to predict the minority class correctly. Again, this points to the fact that, in general, recall for the minority class in an imbalanced dataset is low, as prior research has underlined. Further improvements of these results could be done by deeper investigations on ensemble methods or cost-sensitive algorithms.

6.1 Comparison with Previous Research

Our results partially align with previous studies that emphasize the importance of numeric and sentiment analysis in detecting fraudulent reviews. However, unlike studies that have utilized deep learning-based NLP approaches to achieve high precision across both classes, our work was constrained by a limited exploration of text analysis. This difference suggests that future research should integrate more advanced text-based methods, such as transformer models, to improve model performance.

6.2 Confidence in Results and Generalizability

The results are robust within the context of numeric analysis, particularly when class balance is achieved. However, the generalizability of the findings may be limited to datasets with similar characteristics. Future studies should aim to include data from various e-commerce platforms to improve the external validity and applicability of the results.

7. Conclusion and Future Work

The research question addressed in this study was: "How can the detection of fake reviews in e-commerce be improved using combined analysis techniques?" Our findings demonstrate that integrating numeric features and behavioral analysis, alongside resampling techniques, can enhance the accuracy of fake review detection.

Models performed well with balanced data, with Random Forest, and achieving high accuracy, precision, and recall. Resampling played a crucial role in improving minority class predictions, addressing a significant challenge in fake review detection.

Limitations, including minimal text analysis and dataset constraints, were identified. Future research should incorporate more advanced NLP techniques, such as transformer models, alongside numeric analysis to improve detection accuracy.

From a practical standpoint, the study provides actionable insights for e-commerce platforms aiming to improve the trustworthiness of online reviews. Leveraging combined analytical techniques for fraud detection could lead to more reliable moderation systems and enhance consumer trust in online marketplaces. This research contributes to the academic understanding of fake review detection, with significant implications for real-world applications.

References

Abd-Alhalem, S.M., Ali, H.A., Soliman, N.F., Algarni, A.D. and Marie, H.S., 2024. Advancing E-Commerce Authenticity: A Novel Fusion Approach Based on Deep Learning and Aspect Features for Detecting False Reviews. *IEEE Access.*

Akoglu, L., Chandy, R. and Faloutsos, C., 2013. Opinion fraud detection in online reviews by network effects. *Proceedings of the International AAAI Conference on Web and Social Media,* 7(1), pp.2-11.

Cao, C., Li, S., Yu, S. and Chen, Z., 2021. Fake reviewer group detection in online review systems. In: *2021 International Conference on Data Mining Workshops (ICDMW),* pp.935-942. IEEE.

Cavalcante, A.A.B., Freire, P.M.S., Goldschmidt, R.R. and Justel, C.M., 2024. Early detection of fake news on virtual social networks: A time-aware approach based on crowd signals. *Expert Systems with Applications,* 247, p.123350.

Elmurngi, E.I. and Gherbi, A., 2018. Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques. *Journal of Computer Science,* 14(5), pp.714-726.

Gloor, P., 2012. *Making the E-business Transformation.* Springer Science & Business Media.

Kaur, R., Singh, S. and Kumar, H., 2021. An intrinsic authorship verification technique for compromised account detection in social networks. *Soft Computing,* 25, pp.4345-4366.

Kumar, S. and Shah, N., 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559.*

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

Liu, Y., Jian, Y., Chen, X., Wang, X., Chen, X., Lan, X., Wang, W. and Wang, H., 2024. A metadata-aware detection model for fake restaurant reviews based on multimodal fusion. *Neural Computing and Applications,* pp.1-24.

Luca, M. and Zervas, G., 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science,* 62(12), pp.3412-3427.

Luo, J., Luo, J., Nan, G. and Li, D., 2023. Fake review detection system for online E-commerce platforms: A supervised general mixed probability approach. *Decision Support Systems,* 175, p.114045.

Mukherjee, A., Venkataraman, V., Liu, B. and Glance, N., 2013. Fake review detection: Classification and analysis of real and pseudo reviews. *UIC-CS-03-2013. Technical Report.*

Ruan, L., Zhou, S., Xu, Q. and Chen, M., 2023. Multimodal fraudulent website identification method based on heterogeneous model ensemble. *China Communications,* 20(5), pp.263-274.

Sun, Y., Zhang, Z., Echizen, I., Nguyen, H.H., Qiu, C. and Sun, L., 2023. Face forgery detection based on facial region displacement trajectory series. In: *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW),* pp.633-642. IEEE Computer Society.

Wang, J., Wen, R., Wu, C., Huang, Y. and Xiong, J., 2019. FDGARS: Fraudster detection via graph convolutional networks in online app review system. *Companion Proceedings of the 2019 World Wide Web Conference,* pp.310-316.

Wang, L., Song, Y., Zhang, Z. and Hikkerova, L., 2023. Do fake reviews promote consumers' purchase intention? *Journal of Business Research,* 164, p.113971.

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

Zhao, J., Yang, Y. and Wang, Y., 2023. Effect of user-generated image on review helpfulness: Perspectives from object detection. *Electronic Commerce Research and Applications,* 57, p.101232.