Pakistan Journal of Scientific Research
Multidisciplinary & Peer Reviewed
Volume: 5, Number: 1, Pages: 40-53, Year:2025

Article

Ensemble-based Machine Learning Model for Online Fake Reviews Detection

Rabia Ashraf [1], Mirza Adnan Baig [1], M. Saqib Rehan [1], Shoaib Nawaz [1] and M. Shahzad Abbas [2]

[1] Islamia University Bahawalpur, Rahimyar Khan Campus, Pakistan
[2] National College of Business Administration and Economics, Pakistan
* Corresponding email address: Shoaib.03339770257@gmail.com

Submitted: 01-03-2025**, Revised:** 15-05-2025, **Accepted:** 20-06-2025

Abstract
Online shopping/e-commerce sites are usually unknown to most customers, and they do not know the seller or the goods and services. They will purchase and make any decisions without consulting the reviews done by customers. It does not matter whether they are true or not, but product reviews can play a huge role in the bottom line of a company. Some e-commerce sites have a section where one can confirm the authenticity of the seller, but the majority of buyers would rather read reviews done by real people who have bought the target product and used it. Due to the possible number of reviews regarding a particular product being hundreds or even thousands, it is dubious as to which ones are authentic. Machine learning (ML) has, in recent years, enabled machines to perform tricky tasks with close human levels of expertise. It is possible to find different ways. Conventional means of fake reviews are time-consuming and usually unproductive due to the vast number of reviews produced. Moreover, there is no accuracy or robustness. So, we require a powerful ML-based solution that will be able to automatically evaluate the reviews, distinguish between the authentic and fake ones and then, in a very small time period, choose the most valuable comments of others. To achieve this. In that regard, we propose a modern fake reviews detecting model using ML. The evolution of this study assumes the combination of baseline learning, deep learning and ensemble learning algorithms for fake reviews detection. Therefore, Naive Bayes, Random Forest, Decision Tree, SVM, and K-N Neighbour have been paired together to train and test our proposed model. The proposed model of voting consists of a strict pre-processing procedure and feature extraction. The functions that were carried out before preprocessing are tokenization, removal of stop words, punctuation, and even deletion of rare words. We availed the step of feature engineering, which enhances data prior to entering the next stage, which is the advanced bi-grams, whose name is the N-gram and TFIDF. We have done several experiments and compared the future model and the state-of-the-art models with reference to one another. The obtained data yields that our proposed model is superior to the received data regarding the Uni-Bi-Gram TFIDF-features and effectively classifies the reviews into two classes, real and fake, with 93Percent success precision.

Keywords: Natural language processing, fake and real reviews, machine learning, Textual Information.

1 Introduction

With the rise of online shopping in recent years, consumer behaviour and buying patterns have changed massively. Another important part of such transformation is the dependence on customer reviews, which are very instrumental in determining the quality and efficiency of products and services. Yet, the validity of these reviews is now questioned, and in this regard, fake reviews present a serious problem. Too many good or bad reviews that are fake would misinform customers and disadvantage the names and volumes of the products. The fact that fake reviews are on the rise requires the creation of powerful systems that will identify and disregard such inconsistent pieces of criticism. An increasing desire to utilize machine learning-based methods to automate the process of detection[1]. Conventional techniques of spotting fake reviews require a lot of labour,

https://doi.org/10.57041/1k2hwy61

and they are of relatively low efficiency as there has been an immense number of reviews being produced. Moreover, inaccuracy and insufficient robustness are still issues. In such a way, our research suggests a more developed machine learning-driven method of identifying fake online reviews. We employ various classification methods that are known as Random Forest, K-Nearest Neighbour, Support Vector Machine, Decision Tree and Naive Bayes. In comparison, our proposed model casts a judgment over the act of these models when it comes to separating false and authentic reviews. TF-IDF features and n-gram analysis are used to enhance the accuracy in the evaluation of the effectiveness.

1.1 Motivation

The climatic growth in the e-commerce platforms has transformed consumer expendables by completely transforming customer reviews as one of the critical sources of information concerning product quality and satisfaction. Nevertheless, fake reviews can easily invalidate the validity of such reviews and mislead the consumer and even deteriorate the reliability of online platforms. The desire to build the confidence of the consumers, allow fair players in the business world and resolve the unrealistic verification of reviews using manual reviews. The number of reviews online is due to the popularity of the topic, which is why we study it. Also, the opportunities of machine learning and natural language processing get better on a daily basis and, therefore, automated fake review systems can be designed with increased powers and efficiencies.

1.2 Objectives

The key main purpose of this research is to develop an AI model that should be able to differentiate between real and fraudulent online reviews, rooted in the processes of machine-based learning algorithms. The study will entail a comparison of the efficiency of a few algorithms, such as machine learning algorithms, Naive Bayes (NB), SVM (Support Vector Machine), Decision Tree (DT), Random Forest (RF) and K-Nearest Neighbour (KNN). So, to achieve higher detection accuracy, we will implement textual and linguistic features to be acquired by TF-IDF and n-gram analysis. Moreover, the research is expected to determine an effective assessment framework, which will evaluate the system based on the synthetic and real data.

1.3 Scope Of Our Research

The data analytics and information retrieval services include our research and seek to compete with online marketplaces and companies such as Agoda and Amazon. The online service platforms have gained a lot of popularity since there are effective marketing strategies that engage users who are in need of goods and services. Such sites engage major actors, such as clients, industries, and the management of the sites. The management gives the businesses a platform where they can sell their products or services, and they work together on the marketing plan to ensure they get customers. Business also depends so much on the word-of-mouth response of past customers to increase the number of sales they make to the client, and this makes the business most efficient and effective.

1.4 Key Contribution

1. It proposes a multilayer preprocessing framework that employs two text datasets, where one set contains fake news, and another set is based on real news.
2. NLP is integrated into the preprocessing stage of the data, where it is prepared to be utilized during word embedding.
3. We combined such Machine Learning base-algorithms as Naïve-Bayes (NB), K-Nearest-Neighbour (KNN), Decision Tree (DT), Support-Vector Machine (SVM), and Random Forest to test and train our proposed model. The suggested sophisticated voting model has

sophisticated preprocessing and attribute extraction. The preprocessing functions include the tokenization process, elimination of stop words, punctuations, stemming and elimination of rare words. The feature engineering step is the correction of the data concerning the advanced bi-gram features titled TFIDF and N-Gram.

4. In order to illustrate our results more easily, we may, just after we noticed the highest accuracy in classification of our results through 5 Machine Learning algorithms and a combination of multiple deep learning models, state a superior rate in the classification efficiency of our findings.

1.5 Related Works

Fake review detection is one of the issues that have already been extensively studied in the context of sentiment analysis and natural language understanding. The studies on this topic have revealed that different methods and strategies have been implemented to resolve this problem and make progress in this field.

1.6 Traditional Fake Reviews Detection Techniques

Conventional methods in the detection of fake reviews were mostly dependent on methods such as heuristic approaches and rule-based systems. Such techniques were based on predetermined rules and patterns, such as the indication of suspicious reviews. For instance, [1] discussed how fake reviews can be determined by recognizing duplicates and also by analyzing the distribution of the reviews. Although these techniques worked as a source of basic knowledge, they neither had the sophistication nor the flexibility to deal with the dynamic nature of fake reviews. Due to machine learning, more advanced metrics have started to develop. They applied the supervised learning techniques to categorize fake ones, and used linguistic features and n-grams to increase the accuracy of those. Similarly, [2] proposed the usage of graph-based models in the discovery of fake reviews based on the behaviour of reviewers and the structure of social networks. Feature engineering is a crucial component of the machine learning model. [3] Showed that incorporation of several different features, including sentiment scores, synthetic patterns and metadata features, can be helpful in enhancing the level of classification. Also, the application of TFIDF Term-Frequency and Inverse-Document-Frequency and n-gram analysis to get contextual and linguistic elements in reviews has been widely researched.

Fake review detection has also been investigated using ensemble learning techniques that aggregate two or more classifiers to enhance the prediction performance. [4] Introduced a stacking ensemble model where a number of different base classifiers were combined, leading to a greater level of detection. The paper pointed out the possibilities of ensemble methods with regard to overcoming the weaknesses of single classifiers.

2 Materials & Methods

Figure 1 illustrates the process of the suggested framework. The proposed model consists of two sections, and one of them is preprocessing, and the other is the splitting of the train and test. The process is discussed as follows.

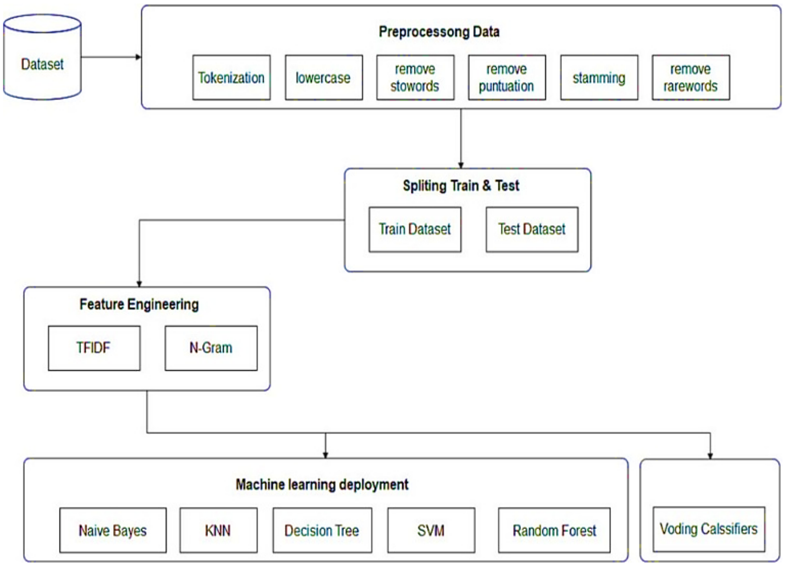

Figure 1. Proposed Approach

2.1 Fake Reviews Detection Using Advanced Machine Learning

First, in our suggested model, it translates an input sentence into a set of vectors. Further, it creates a vocabulary graph using the co-occurrence dependencies of words within a particular document. It indicates the global information of the text with the help of the output of the vocabulary graph. Meanwhile, the local information in the input text can be learned with a pre-trained language model. It has multi-layer attention schemes that help in combining the local and global input information of the text in question [10]. When we were coming up with the outline of our study, one aspect which was very important to us was to determine which particular points of the proposed model are to be made clear. The rationale behind this process was driven by three significant aspects that involved: (1) coming up with a manageable number of groups so that the success of the model could be achieved; (2) picking groups that were as common as possible with the data set so that they are significant.

2.2 Preprocessing Of Data

The critical step in the proposed way is the preparation of the data. The data of the world cannot be used, so pre-processing of data is needed. This research brought into play several preprocessing techniques to ensure that the data contained in the raw contents of data were transformed to a format appropriate to empower data. In the proposed framework, the preprocessing techniques are clarified below.

a. Text Cleaning

The first part was to clean the text data of any irrelevant information, HTML Tags, and URLs: All the HTML tags and URLs were removed in the reviews in order to concentrate on the textual part of the reviews.

- Special Characters: The punctuations and other special characters irrelevant to the flow of meaning of words and language were deleted.

b. Tokenization

This is the unity of the most applied methods in NLP (natural language processing), i.e. tokenization. It is a precursor step to the use of some other forms of preprocessing. Tokens are the single words of an article. As an example, the statement, I love this book would be divided into tokens, and these tokens are I, love, this, and book as a result of the tokenization process. [6].

c. Convert Upper Case Lowercase

In the process of trying to create consistency and minimize disparities within the text, the case of all words in the reviews was converted to lowercase. This is done in order to ensure that such words as" Good" and" Good" are processed in the same way.

d. Removing

Eliminating Stop words is words that are widely utilized in text, like the words an, is, to and this, which do not have a lot of weight when they are active in the text. These words are taken from the data in this study, and then the process of detection of fake reviews is further taken [7]. In the English language, there are many forms in one sentence. Such differences with regard to a given text result from comparable information as we construct the model of NLP or even the model of machine learning. There is a possibility that some of these models do not perform as one might have anticipated. To develop a strong model, language needs to reduce the redundancy of words and eliminate repetitive words to their counter minimal methodologies [8-12].

e. Feature Extraction

The mission of feature extraction is to complement the assigned task of a design appreciation structure or a machine learning system. The feature extraction process includes simplification of the input to the key aspects of it, in a way that will give both the perfection of ML (machine learning) and the model of DL (deep learning) some more sensible data. It is compulsory to make sure that all the extraneous or irrelevant feature in context is removed as far as the accuracy of the model is concerned [2].

f. N-Grams

The capturing of the contextual information was done by using both unigrams and bigrams. N-grams take into account word strings, and this could become especially helpful in understanding patterns and anomalies in the fake reviews. Its main aim will be to collect information about the presence of n-grams in the presented information and use it to predict the next words. There is a philological approach, which is called a unigram and is based on the previous work predicting the following one. The presented structure can be considered a bi-gram language model, where words from the previous two words are used to predict the third one [8, 13-16].

g. TF-IDF

The advanced capability determines the weight of the word in the document as compared to the frequency in the total setup, allowing highlighting of essential words between the fake and the authentic reviews.

$$Wx,y = tf(x,y) \times \log \frac{N}{dfx} \qquad (1)$$

The dataset is gathered on Zenodo [9] and should be available to any researcher. There are five columns in the dataset, and the characterization of the product is the product Categories and the ranking/rating of the products given by the users or customers. Tagging the reviews generated by the computer (CG), also known as Fake, and the human-generated (HG), also known as Real reviews.

Table 1. Dataset Features and Description

Feature	Feature Description
Product-Category	It is the products-type where the customer input reviews will be received.
Rating/Ranking	The post- purchase customer rating of the quality of the product ranged between 0 and 5.
Label	It is the false and true comments class.
Text	This is the review text message that the customers paste on the reviews section.

Text is the most important feature of the data because they are the reviews of the products and the labels because each of these reviews is referred to the category is part of the label.

2.3 Creation Of Dataset

There are two classes that the classes have been partitioned into as illustrated in Figure 2, which also indicates the spread of the classes along with the ratings of the product. This is good in the fact that classes are balanced and the values are already narrowed and as such, there is no need to scale favor value of the classes.

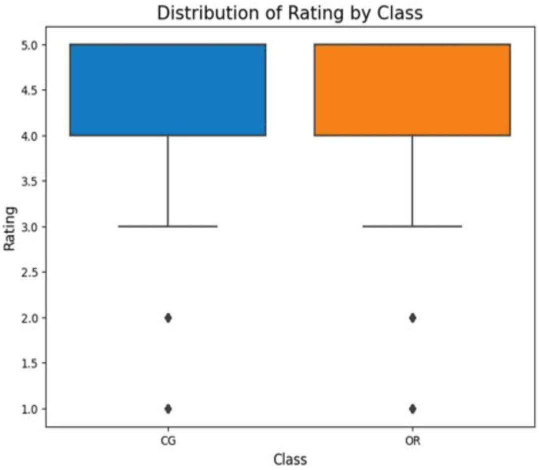

Figure 2. Rating in Classes Count Plot

2.4 Classes Balance

The well-balanced dataset and the classes are by far the most important part in the text classification to reach a scenic model ruling, as far as our dataset has well-balanced classes, as portrayed in Figure 3. Well-balanced classes assist us in not receiving a biased trend and take care of precision, recall and F1 proper classifications. The correction of class imbalance can be varied, and depending on the problem at hand and the approach of achieving the proposed study objective, it can be corrected by correcting the imbalance of classes by approaching the mean value, by balancing the classes, or by balancing the classes using the weight of each class.

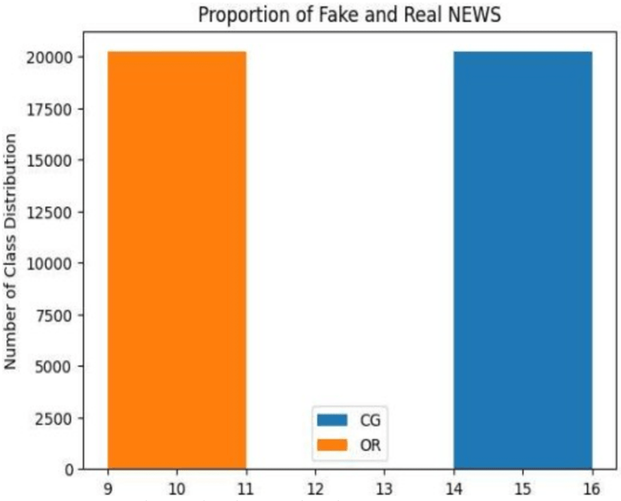

Figure 3. Balance in the Both Classes

2.5 Balance In Both Classes

In the data set, the product has 10 of the product categories in terms of reviews given, and their comments are available in the set as well. The products are read back in various categories; thus, the reviews are as unique as possible. The products have 4,000 reviews category is 10. Figure 4 gives the type of product encompassed by the coverage of reviews done under the dataset.

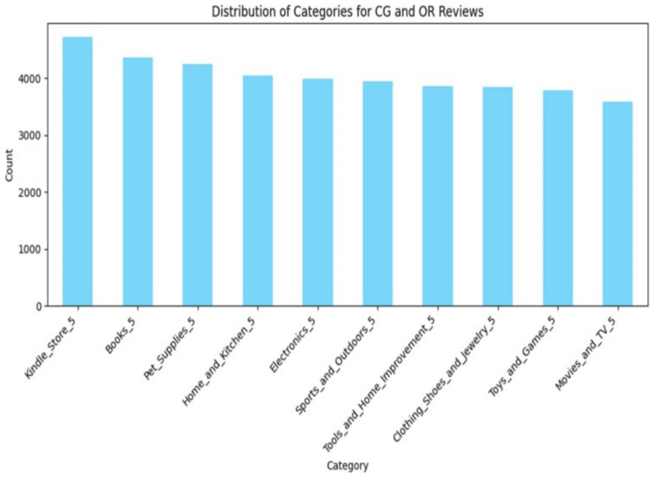

Figure 4. The Product category included how reviews are included in the dataset.

2.6 Rating Of Products

The other variable in the set of data is the rating of the rating. The rating is the value that discusses the quality of the product. Once the product gets into the hands of the customer, the customer consumes the product and gives comments about the product, such as " the product quality is good or “the product quality is bad, the safe delivery of the product, etc. The rating, therefore, is used to determine the excellence of the services as well as the quality of the products. The customer giving his rate only has the option to choose between 1 and 5 stars; a star of low quality and rate, and so with 5 stars, hence good quality and rate. This is also another value addition that can prove to be of attraction, so as to affect the tone of the reviews of the products as well. The distribution of the rating is described based on the score or the rating between 1 star and 5 stars, as shown in Figure 4, indicates that the maximum of the ratings which are being observed are the 5-star rating and the minimum rating observed is the 1-star rating, and even the rated 5 products are the maximum being found within the whole data set.

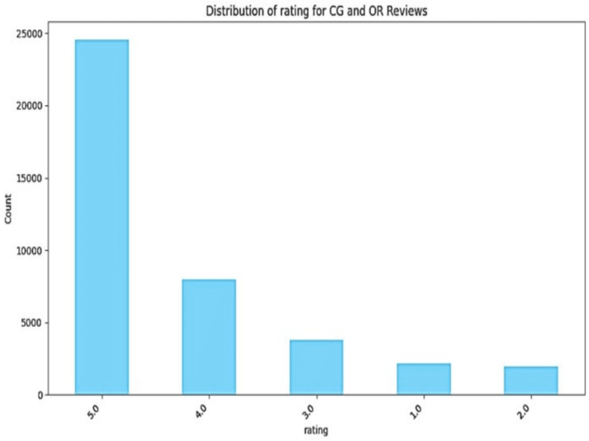

Figure 5. Product rating classification and rate of rating according to score

2.7 Generation Of Proposed Classification Model

Once we know the psychology of the dataset, we apply the principal component analysis to reduce the dimension of the structures (TF-IDF) dataset, and then we plot the information on the graph to have the visual pattern of the dataset. The shapes allow us to understand whether the data is linear, distributed, complex or overlapped, and thus let us know more about the workflow.

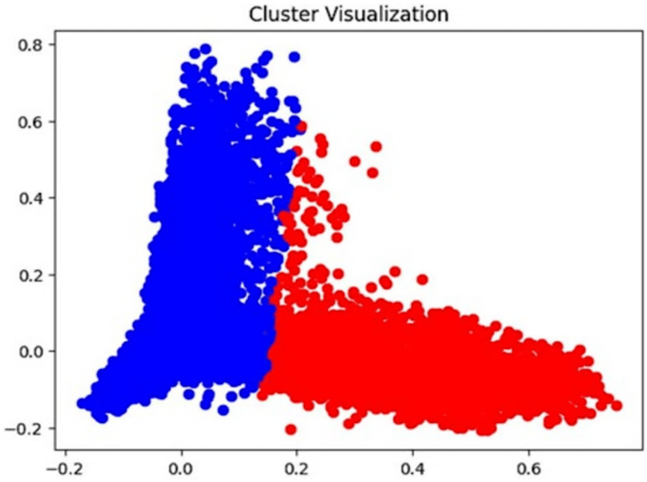

Figure 5: The plot of the data points using PCA to know the patterns of the Data points

Figure 5 shows a plot that depicts how the data is broken into the two partitions, where one is blue, representing the class 0, and the other is red, indicating the class 1. The throw is in linear form, and we can draw the line I among the data of blue and red.

2.8 Simulation, Results, And Discussion

The experiment has been undertaken, and a simulation has been done. We evaluated five distinct base machine learning algorithms (Support Vector-Machine, K-Nearest-Neighbour, Decision-Tree, Naive Bayes, and Random-Forest) in terms of unigram feature, bi-gram feature and uni-bigram feature. These methods showed different results for the accuracy of different algorithms and feature sets. The voting algorithm that included all three kinds of features was implemented, and its performance with respect to classification improved. This comparative analysis underlined the applicability of the uni-bigram features, which boosted the accuracy that was essential to identify deceptive reviews.

2.9 Performance Metrics

In the next subsection, we have feature-based performance analysis of each algorithm.

A. Uni-Gram Base Algorithms

The Uni-Gram: The Uni-gram refers to one word selected at a time, and it is one of the characteristics that is drawn in the term persistence and also in the inverse document frequency [10]. The Naive Bayes is the one that provides us a testing conception of 85% for the accurate classification of data, and the precision gets a score of 85%. KNN is the procedure that gives the lowest percentage of classification prediction, which is 59%, and this is only because of the structured, unstable information set. The second model is the Decision tree, and the Decision tree has the same precision, recall and F1-Score, which is 75 percent. What is key about the decision tree is the algorithms which provide us with identical outcomes in the accuracy and precision, recall and F1 Score. Figure 6 presents the results.

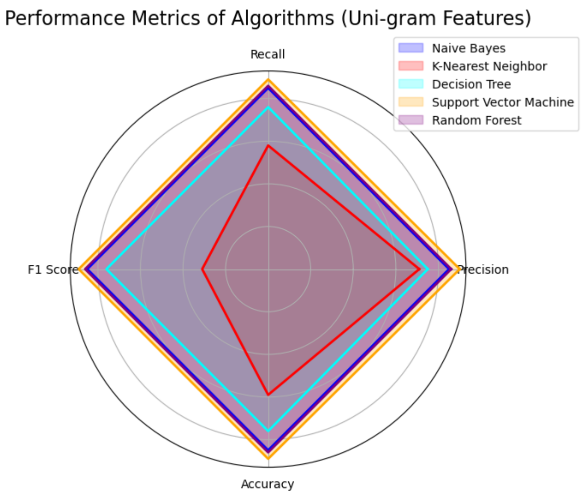

Figure 6. NB, KNN, DT, support vector machine, RF with uni-gram

In Figure 7, the Support Vector Machine is the most efficient model, providing 90% correct classification, and the factors are Precision 89, Recall 89, and F1-Score 89% which is the most outperforming algorithm among the five, and the 2nd efficient algorithm is the random forest, with the correct classification. They use the overall results of the Uni-gram features, and the most advanced algorithm embraces the support of the vector machine.

B. Base Algorithms

During this step, the features that were used based on bi-grams were provided to test basic machine learning algorithms. Findings (Figure 8) indicate that the ratios of performances did not differ significantly with the variability of ratings found in unigram features. An example is that K-Nearest Neighbour (KNN) has increased accuracy rate by 1moment to 60 % and Support Vector Machine (SVM) remained very high at 90% and rose to 91% in later processes. Table IV summarizes the performance of all underlying algorithms, giving special consideration to the small effects of bi-gram attributes on general accuracy. Nevertheless, this does not take away the importance of variation in the feature engineering, as is evidenced by Figure 7 in relation to the performance of Naive Bayes on bigram features.

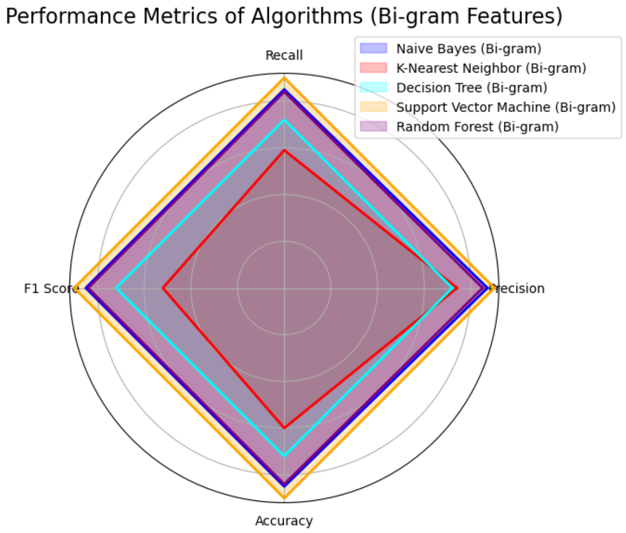

Figure 7. Base algorithms with Uni-gram feature classification

C. Uni-Bi-Gram Base Algorithms

It is the TF-IDF features used currently with the base algorithms, and here is named in Figure 7 the precision of the model in accurately identifying the classes, recall, F1 score, and the model accuracy. The support vector machine provides more degraded results, and KNN, the support vector machine with Bi-gram it was correctly classified 90 percent but in case of Tri Gram we observe the model performance of the machine offers lower results, that is 1 percent and this 1 percent results increase in the overall algorithms results which gives out that the features are impacting on the Models performance. Figure 9 contains the Base algorithm results and uses Tri-Gram Features. Figure 8 shows the Base algorithms' result with Tri-Gram Features.

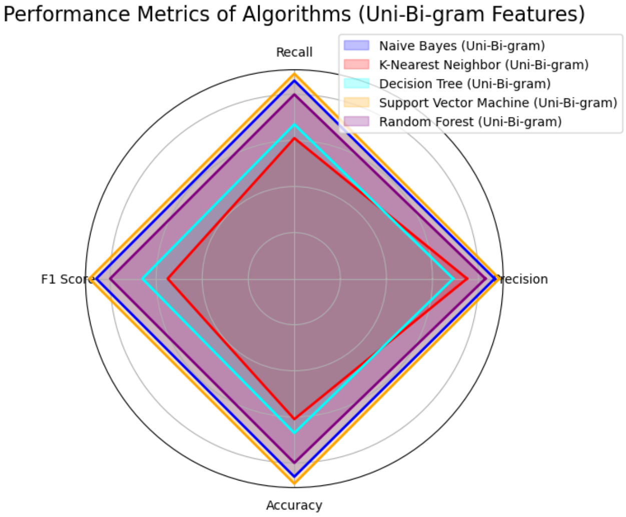

Figure 8. Base Algorithms Result Using Uni-Bigram TFIDF Features

D. Voting Classifiers

To run the Voting Classifier, the proposal will generate improved classification based on the same algorithms as their base models, namely, K-Nearest-Neighbour, Random-Forest, Decision-Tree, Naïve-Bayes, and Support-Vector-Machine. Such algorithms will be based on the use of Uni-Gram, Bi-Gram and Uni-Bi-Gram features of TF-IDF. These three sets of features will also be used in consistency with their use in the base algorithms to determine the precision of the voting classifier.

The result of Figure 9 indicates that their performance is decreasing relative to the Base Algorithms, such as Support Vector Machine, 91% ninety-one percent, on the unigram features. The presentation of the Voting Classifiers is not that bad because it is above 80 percent like Uni-Gram at 84 percent and Bi Gram at 83 percent and the Tri- Gram is 81 percent and one entity is that the psychology gets to see the training and test dataset and the performance is nearly the same ratio-wise and not a big change can be expected in the Voting Classifier.

The voting-label outcomes with Precision, Recall, and F1-Score and results are presented in Figure 10. We can mention that by choosing the SVM with Uni-Gram features, we would obtain the best in the base model classification. The support vector machine has a 91% accuracy rate, and with an elective classifier, the maximum correct classification rate remains 86 percent, which is more than>5 percent above the base algorithms. We will consider the Voting models as a product to be used in sorting since it leads to variation in the results and the accuracy of the model is lower in voting algorithms.

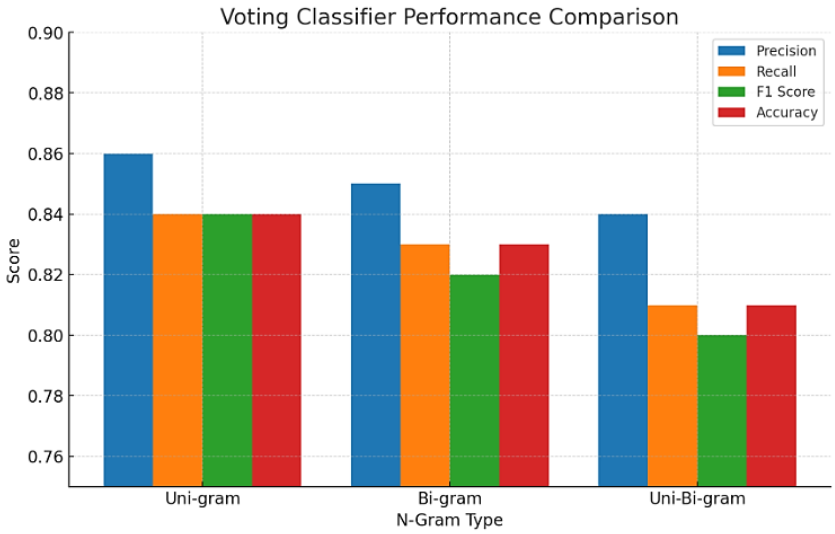

Figure 9. The 5-voting classifiers with TF-IDF N-gram Features

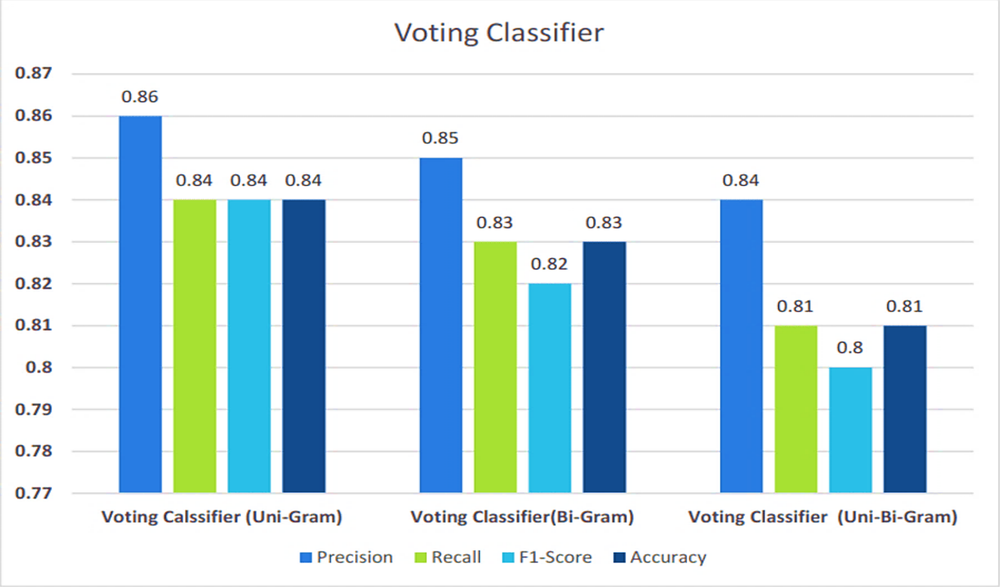

Figure 10. Voting Classifier with All Features

3 Experimental Results

A. Compression Of Traditional and Proposed Study Results

Table 2 provides the results of the models and their truth in regard to the previous results of the study and the suggested studies' results. The data, too, does not run parallel in both studies, and the model is not similar either, and the matter, too, is similar to the fake and the real reviews classification. The information provides a list of previous research performance of the models and future study results of fake reviews. The articles refer to the various years, authors, the nature of the datasets/ sources, the type of classification and the nature of the features, along with the level of accuracy. Looking at the outcomes of the past works, it is possible to mention that several researchers have examined the issue of detecting fake reviews on the basis of different strategies. They have deployed classifiers such as Random Forests (RF), SVM, NB, DT, and LR; they have used the contextual features and behavioural features. The accuracy level, as indicated in the studies, ranges between 77 and 87 marks. The use of Voting-Classifier (KNN), Decision Tree, Naive Bayes, Support Vector Machine, and Random Forest will be considered in the results of the proposed study. This kind of characteristic is the TFIDF (depends on n-gram (bi-gram, unigram). The suggested research will be able to classify CG GH reviews and provide the accuracy rates between 59 percent and 90 percent. It needs to be noted that the accuracy rates, regardless of the sources that were taken into account by the authors of the data, are not alike in regard to the different studies and classifiers. There are some reports that describe a very high percentage of accuracy (up to 93) and others that state the accuracy levels to be low (down to 55).

Table 2. Compression of the Proposed Study Results

Previous Study Results					
Year	**Author**	**Dataset Type/Source**	**Classifier**	**Feature Type**	**Accuracy**
2016	Dongsong Zhang et. al	Real life/Yelp	SVM, DT, RF and NB	Contextual, Behavioral	**81,82,86,87,81**
2022	Saleh Nagi et, al	CG HG, Amazon Review	SVM, DT, RF, NB, ADaBoost	TF-IDF	**90,92,90,90**
2023	Sami Ben jabeur, et,al	Manchester Online Reviews	Bibliometric	TF-IDF, Textual Feature	**92.1**
2023	Joni salminen	CG HG Review	GPT-2, ULM Fit	TF-IDF	**55-91%**
2023	Rene´ Theuerkauf, et,al	CG HG Review	FC-Combination	TF-IDF	**79-91%**
Proposed Study Results					
2025 Proposed Study		CG GH Reviews	Naive Bayes	TFIDF-Uni-Gram	**85%**
			K-Nearest Neighbor	TFIDF-Uni-Gram	**59%**
			Decision Tree	TFIDF-Uni-Gram	**76%**
			Support Vector Machine	TFIDF-Uni-Gram	**89%**
			Random Forest	TFIDF-Uni-Gram	**86%**
			Naive Bayes	TF-IDF-Bi-Gram	**86%**
			K-Nearest Neighbor	TF-IDF-Bi-Gram	**85%**
			Decision Tree	TF-IDF-Bi-Gram	**60%**
			Support Vector Machine	TF-IDF-Bi-Gram	**72%**
			Random Forest	TF-IDF-Bi-Gram	**90%**
			Naive Bayes	TF-IDF-Uni-Bi-Gram	**86%**
			K-Nearest Neighbor	TF-IDF-Uni-Bi-Gram	**61%**
			Decision Tree	TF-IDF-Uni-Bi-Gram	**67%**
			Support Vector Machine	TF-IDF-Uni-Bi-Gram	**89%**
			Random Forest	TF-IDF-Uni-Bi-Gram	**80%**
			Voting Classifier	Uni-Gram	**84%**
			Voting Classifier	Bi-Gram	**83%**
			Voting Classifier	**Uni-Bi-Gram**	**81%**

Such differences may be explained by the differences in the datasets, feature engineering methods, classifiers, and so on, characteristic of each of the studies. On the whole, the statistics show that the study is still actively studied in fake review research and that a lot of different methods and ways of action are also examined with the aim of making the research more precise. The offered research contributes to the said body of research by introducing novel combinations of classifiers and their performance against reviews on the CG GH data set in terms of TF-IDF features. All the comparison results have been shown together in one figure, as shown in Figure 11, with the model name, accuracy, precision, recall, and F1 score.

4 Conclusion

Our paper has conducted a 360-degree study to resolve the issue of fake reviews. We have approached a road trip on an uneven data set comprising authentic reviews and fake reviews, and managed to tie up with the machine-learning models that are effective. Such methodologies as Support Vector Machines (SVM), k-Nearest-Neighbours (KNN), Naive Bayes, Decision Trees and Random Forests helped us to identify deceptive patterns in reviews.

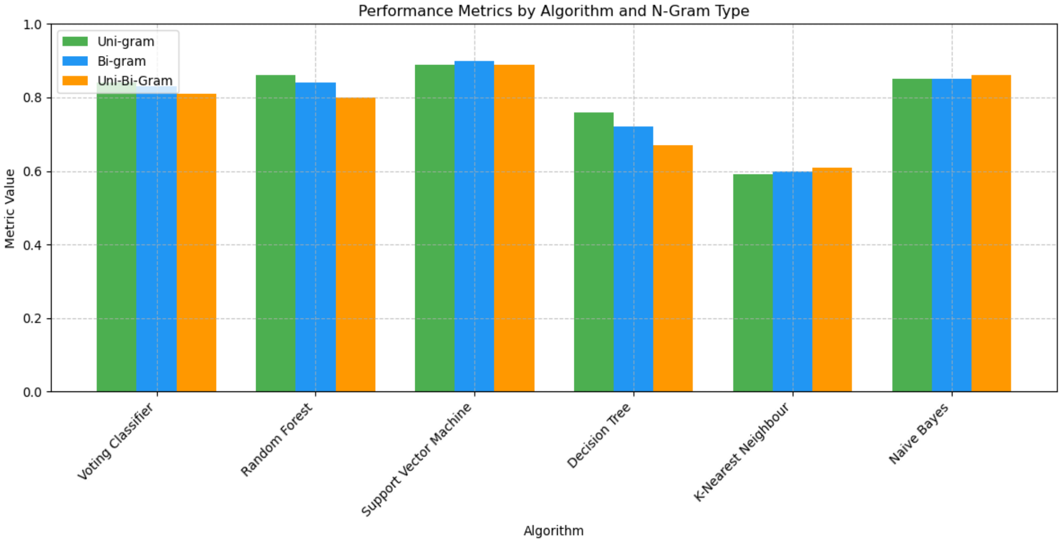

Figure 11. Model Performance with All Features.

The core aspect of our implementation was the utilization of TFIDF (Term Frequency Inverse Document Frequency) as an appropriate option that helped to enable one to use textual feature extraction so as to accommodate both unigrams and bigrams in order to capture the nuances of language. It reached such an excellent 93 percent accuracy rate that it can be demonstrated that SVMs are useful models for the identification of genuine and bogus reviews. This kind of outcome demonstrates the potential of our suggested approach with respect to making the user-generated information on different websites, whether these are e-commerce stores or local business reviews, more reliable. To continue our work, we can say that the results of our research should guide us to investigate other feature extraction techniques, to use rich machine learning techniques and encompass contextual information to increase the detection rates even further. They will enable mitigating the negative impact of the problem of fake reviews and, thus, produce a more trustworthy digital space. In conclusion, we think that the state-of-the-art machine learning techniques are of critical importance in combating fake reviews by validating what is said in the fake reviews. By using such tools, we are going to empower consumers with reliable information and, thereby, work on the notion of digital platform transparency and integrity. Investigation into the usage of more effective structural forms of lexical graphs, pursuant to such an example as WordNet or the graph already made on the basis of the word co-occurrence data, may further enhance the effectiveness of detection. The use of higher vocabularies or graphs, which have the capacity to capture the additional, deeper information, is something which can result in improved detection. We did not use the PCA (Principal Component Analysis); however, we could use it in the future.

Funding Statement

The authors declare they have no conflicts of interest to report regarding the present study.

Conflict of Interest

The Authors declare that they have no conflicts of interest to report regarding the present study.

Author contributions

Conceptualization and review design, R.A., M.A.B. and S.N., methodology, R.A. and M.S.A.; reference-management tooling, R.A; validation of screening and inclusion/exclusion criteria, M.S.R.,

M.A.B, and M.S.A.; formal analysis and synthesis, R.A. and M.A.B.; investigation (literature search and data extraction), R.A.; resources, M.A.B. and M.S.A; writing original draft preparation, R.A. and S.N.; writing review and editing, M.S.A.; visualization (figures and summary tables), M.S.R.; supervision, S.N.; project administration, M.S.A. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data is available on reasonable request.

References

1. N. Jindal and B. Liu, "Analyzing and detecting opinion spam," in Proceedings of the International Conference on Web Search and Data Mining (WSDM), 2008, pp. 219–230.
2. A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "Identifying fake reviews: A study on real and fabricated review classification," University of Illinois at Chicago, Chicago, IL, USA, Technical Report UIC-CS-03-2013, 2013.
3. L. Li, W. Ren, B. Qin, and T. Liu, "Document representation learning for detecting deceptive opinion spam," in Chinese Computational Linguistics and Natural Language Processing with Naturally Annotated Big Data, Nanjing, China: Springer, 2015, pp. 393–404.
4. Y.-R. Chen and H.-H. Chen, "Detecting opinion spam in web forums: A case study," in Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 173–183.
5. H. Deng, L. Zhao, N. Luo, Y. Liu, G. Guo, X. Wang, Z. Tan, S. Wang, and F. Zhou, "Detecting fake reviews using semi-supervised learning," in Proceedings of the IEEE International Symposium on Parallel and Distributed Processing Applications and IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), December 2017, pp. 1278–1280.
6. S. Meftah and N. Semmar, "A neural network approach for part-of-speech tagging in social media texts," in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.
7. K. Sparck Jones, "Stop words," in Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR '88), 1988.
8. Khan, Muhammad Nasir, Ali Altalbe, Fawad Naseer, and Qasim Awais. "Telehealth-Enabled In-Home Elbow Rehabilitation for Brachial Plexus Injuries Using Deep-Reinforcement-Learning-Assisted Telepresence Robots" *Sensors* 24, no. 4: 1273, **Feb. 2024**.
9. Zenodo, [Online]: https://zenodo.org/records/13444249.
10. Kaggle dataset: Fake and Real Reviews, available at: https://www.kaggle.com/datasets/shoaib03339770257/fake-real-reviews/data.
11. Ali Altalbe, Muhammad Nasir Khan, Muhammad Tahir and Aamir Shahzad, "Orientation Control Design of a Telepresence Robot: An Experimental Verification in Healthcare System," Applied Sciences, 13, no. 11: 6827, June **2023**.
12. M. Costa Filho, D. N. Rafael, L. S. G. Barros, and E. Mesquita, "Combating fake reviews: Safeguarding consumers through persuasion knowledge," Journal of Business Research, vol. 156, p. 113538, 2023.
13. Naseer F, Khan MN, Addas A, Awais Q, Ayub N. Game Mechanics and Artificial Intelligence Personalization: A Framework for Adaptive Learning Systems. *Education Sciences*. 2025; 15(3):301. https://doi.org/10.3390/educsci15030301
14. Naseer F, Addas A, M. Tahir, Khan MN, Sattar N. Integrating generative adversarial networks with IoT for adaptive AI-powered personalized elderly care in smart homes. *Frontiers in Artificial Intelligence*. 2025; 8:1520592. https://doi.org/10.3389/frai.2025.1520592
15. A. Addas, F. Naseer, M. Tahir and Muhammad Nasir Khan, "Enhancing Higher Education Governance through Telepresence Robots and Gamification: Strategies for Sustainable Practices in the AI-Driven Digital," *Educ. Sci.* 2024, *14*(12), 1324; https://doi.org/10.3390/educsci14121324
16. Shazia Rehman, Abdullah Addas, Erum Rehman, Muhammad Nasir Khan, "The Mediating Roles of Self-Compassion and Emotion Regulation in the Relationship Between Psychological Resilience and Mental Health Among College Teachers, Psychol Res Behav Manag. 2024;17:4119-4133