Unified Detection: Enhancing Information Reliability through Machine Learning Classification of Fake, Spam, and Legitimate Content

Saharsh Gupta, Ayush Goyal, Mehul Kumar, Saurav Kumar, Nishi Jain

Vivekananda Institute of Professional Studies-Technical Campus, GGSIPU, India

Corresponding author: Saharsh Gupta, Email: saharshg895@gmail.com

The continued increase of spam and fake information in the digital age has raised serious concerns about the veracity and authenticity of digital content. To tackle this issue, we have developed a unified machine learning-based classification system that distinguishes between spam, fake, and legitimate information, addressing a gap in existing solutions which typically focuses either on spam or on fake information only. In this research, we collected a diverse dataset which includes both data from YouTube spam collection, email and SMS spam databases, as well as fake news from the some of the largest fake news datasets available—WELFake fake news dataset and GossipCop fake news dataset. Our approach underwent preprocessing and feature extraction steps and thereafter the implementation of different machine learning models like logistic regression, k-Nearest Neighbors, XGBoost, Extra Trees Classifier, and Random Forest. The performance of these models, were examined using various performance metrics like accuracy, precision, recall and F1 Score. Result: the best-performing model, with the most optimal results was Extra Trees Classifier among the baseline models followed by XGBoost and Random Forest, and the Voting Classification significantly improved the accuracy of these baseline models. This unified approach offers a comprehensive solution for automating the filtering of digital content, substantially enhancing the reliability of information by simultaneously addressing multiple types of misinformation. This study contributes to the development of scalable tools that can be deployed across various platforms to ensure the integrity of digital information.

Keywords: Spam and fake information, Classification system, Machine learning, Voting Classification, Unified approach.

in: *Advancements in Communication and Systems*. Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 541–556. DOI: https://doi.org/10.56155/978-81-955020-7-3-48

1 Introduction

The spread and consumption of information have utterly changed with the development of digital communication. Information sharing has been increased in its speed and availability through the newly upgraded venues—social media, online news sites, and instant messaging services. However, this digital transformation also boomed the spread of false information and spam, causing serious threats to the authenticity and trustworthiness of information [1], [2]. In so many economies of the world, the major sources that are online platforms followed by the social media are now the major sources of news for the most population, thereby almost replacing the use of the printed newspapers by most [3], [4]. Utility integration available on the platforms is at a big risk and can at most shift the course of the events. In particular, the problem of fake news spread through the internet received dramatic resonance after the US presidential election of 2016 [5], [6]. Fake and spam have a sharp growth, not just deceiving the population but undermining the credibility of the genuine information sources.

Existing detection systems normally target either fake news or spam. Thus, there should be a single system developed that can classify information into fake, spam, and legit. The spreading of false political speech is known to shift the course of the events in the election field, since undesirable consequences for society are possible [7]. Besides, it was proved that false news disseminates much faster than true information, and, thus, its impact is multiplied [8]. In this paper, we introduce a new approach by developing a comprehensive machine learning-based classification system that accurately infers among these three categories. The identification of the exact information, whether it is fake, spam, or legit, is crucial for keeping the integrity of digital platforms. In that view, this work contributes to the development of filtration that is automated to offer a scalable solution to enhancing the reliability of information. The focus of this study will be to better the overall quality of information available to the public by curbing fake and spam information.

2 Literature Review

The detection of fake news has attracted a lot of research interest, and several methods have been proposed to address this challenge. In a study by Allcott and Gentzkow [1], the role of social media in the diffusion of fake news around the 2016 US elections was considered by analyzing posts that contained URLs to news articles. The authors showed how fake news stories engaged more users than real news stories; however, it is actually very hard to find the real origin and the real spreading path for the sake of the analysis.

Vosoughi et al. [2] investigated the diffusion of true and false news in Twitter, with the help of a significantly large dataset. They find false news always diffused farther, faster, deeper, and more broadly than true news for all categories of information. This research was based on machine learning to classify the truth and falsity of news, with psychological reasons behind the rapid spread of false news. The dataset collected may be biased, as it was taken from within the platform of Twitter only.

In a study by Zhang and Ghorbani [5], it was shown that voters were indeed susceptible to deception with respect to political statements. They subsequently built a fake news detection systembased on the content and social-context. This study was based on news articles and user interactions.They utilized a large news-based dataset and used a variety of machine learning techniques to test the performance of their approaches. From their analysis, they found out that even with the best approaches, there remained a significant room for better results, especially for dealing with nuanced cases. Several approaches have been used for proposing the problem of fake news detection using both conventional machine models and deep learning models.

Pennycook and Rand [4] conducted an analysis about the reasons behind psychometric cases in spreading fake news. They developed warning mechanisms by experimental data to test their developed

hypothesis and found that simple warnings indeed do help in significantly lowering the perceptions of fake news accuracy. The only issue with their research was that it completely relies on controlled experiments, and this cannot capture the real-world situation.

Basak et al. [6] developed a deep ensemble method for fake news detection. They used both convolutional neural networks (CNN) and bi-directional long short-term memory (Bi-LSTM) to ensure better accuracy. They used the FakeNewsNet dataset, which consists of news datasets from popular fact-checking websites. Their model performed very well with high precision and recall, but the main issue with their model was its complexity and computational requirement.

Karimi et al. [7] developed and proposed a multi-class fake news detection system. They combined automated extraction of features and multi-source fusion to handle news articles of different levels of fakeness. Their dataset comprises the news articles of multiple sources, social media interaction, and user comments. Their proposed framework was highly accurate in news articles classification, but faced scalability and real-time detection problems because it required a large amountof computational resources. All these strategies need the availability of varied and rich datasets that can then be used for training and evaluating the detection systems of fake news and spam. Two of the most popular datasets in this research area are the WELFake dataset and the GossipCop dataset. These two datasets provide diverse examples of fake and real news that can be utilized in coming up with sound classification models.

Zhou and Zafarani [9] conducted an overview of fake news research, discussing basic theories, methods of detection, and opportunities. They noted that the interest of the moment is in the production of high-quality datasets and also noted that the current datasets are not diversified and possibly have biases in them. Spam filtering machine learning approaches have also been reviewed adequately to point out the various methods and their performance. A state of the art review was done by Guzella and Caminhas [10] of recurrent patterns and support vector machines for spam filtering. This review is essential to understand how spam filtering technologies are changing. They have touched on other datasets as well, with specific mention of the Enron spam dataset and the TREC 2007 spam corpus, giving details of their strengths and weaknesses. More so, Almeida et al. [11] have also touched on recurrent patterns and support vector machine (SVM) approaches to filtering out spam, reflective of advanced methods able to cope with the identification and filtering out of unwanted content. In this work, emphasis was placed on feature selection and the difficulty of coping with the evolution of spam tactics.

Madani et al. [12] proposed a more general fake news detection solution based on feature extraction, natural language processing, curriculum learning, and deep learning. They used a heterogeneous dataset in their study and achieved good results and improved detection accuracy. However, the intricacy of their model has constrained some computational requirements and its implementation.

Maqsoodet al. [13] designed an intelligent framework in deep learning for SMS and email spam detection. They applied advanced machine learning techniques to develop a robust detection system. In all truth, it was highly accurate, but however, presented a real challenge concerning computational load and scalability.Sumathi and Raja [14] used machine learning algorithms to develop spam detection in social networks. In their analysis of different models, the authors pin down the most workable for real-time detection. However, they note that their study is generalized, needing a lot of other datasets. Table 1 provides a brief overview of all the researches discussed in this section

Table 1. Overview of the all the research paper discussed in this section

Study	Dataset Used	Objective	Method Employed	Results	Limitations
Allcott and Gentzkow [1]	News articles shared on social media	Fake news detection	Analysis of social media's role in spreading fake news	Fake news stories were widely shared and had significant reach	Difficulty in accurately identifying the true origin and propagation path of fake news
Vosoughi et al. [2]	Twitter posts	Fake news detection	Machine learning classification	False news spreads primarily due to human behaviour	Potential bias as the data primarily focused on Twitter
Zhang and Ghorbani [5]	News articles and user interaction data	Fake news detection	Hybrid approach using contentand context	High accuracy in detecting fake news	Challenges in distinguishing between varying degrees of fakeness
Lazer et al. [3]	Comprehensive dataset of news articles	Fake news detection	Range of machine learning techniques	Some models performed well, significant margin for improvement	Handling nuanced cases of misinformation
Pennycook and Rand [4]	Controlled experimental data	Fake news detection	Psychological factors and warning mechanisms	Simple warnings significantly reduced the perceived accuracy of fake news	Reliance on controlled experiments which may not fully capture real-world dynamics
Basak et al. [6]	FakeNewsNet dataset	Fake news detection	Deep ensemble framework integrating CNN and Bi-LSTM	High precision and recall	Model's complexity and computational requirements
Karimi et al. [7]	Articles from multiple news sources and social media interactions	Fake news detection	Multi-class fake news detection method	High accuracy in classifying news articles	Struggles with scalability and real-time detection
Zhou and Zafarani [9]	Multiple datasets including WELFake and GossipCop	Fake news detection	Comprehensive survey of fake news detection methods	Emphasized the need for high-quality datasets	Lack of diversity and biases in existing datasets
Guzella and Caminhas [10]	Enron spam dataset, TREC 2007 spam corpus	Spam detection	Review of machine learning approaches to spam filtering	Detailed review of spam filtering technologies	Varies depending on the dataset and spam tactics
Almeida et al. [11]	Enron spam dataset, TREC 2007 spam corpus	Spam detection	Use of recurrent patterns and SVM for spam filtering	Effective in identifying and filtering out unwanted content	Challenges in adapting to evolving spam tactics

Shu et al. [8]	Social media posts	Fake news detection	Data mining perspective using machine learning techniques	Comprehensive overview of detection methods and models	Challenges of detection due to the dynamic nature of social media and evolving tactics of fake news creators
Madani et al. [12]	Diverse datasets for fake news detection	Fake news detection	Feature extraction, natural language processing, curriculum learning, and deep learning	Significant improvements in detection accuracy	Model complexity and computational requirements
Maqsood et al. [13]	SMS and email datasets	Spam detection	Deep learning-based intelligent framework	High accuracy in spam detection	Managing computational load and scalability
Sumathi and Raja [14]	Social network datasets	Spam detection	Machine learning algorithms	Effective real-time spam detection	Need for more diverse datasets to improve generalizability

This indicates that while some steps in the right direction have been made on how to classify fake news and spam, the need for a unified classification system that would contain a variety of categories of information is crucial. This research attempts to develop a comprehensive machine learning system that will accurately categorize information as either fake, spam, or legit.

3 Methodology

The proposed solution will be targeted at building a strong Machine Learning-based classification system to properly distinguish between fakes, spam, and legitimate information. We used a diverse data set from concatenations on YouTube Spam Collection [15], Email Spam [16], SMS Spam [17], WELFake [18], and GossipCop Fake NewsDatasets[19]. We collected these datasets from Kaggle and UCI Machine Learning Repository.

3.1 Data Summary

Here is a summary of the dataset: Figure 1. shows which datasets were used to create “Final Dataset” on which different machine learning models were applied later. The final dataset contains two columns, the text and the target. Figure 2. shows the distribution of individual datasets in final dataset. Figure 3. shows distribution of fake, spam and legit information in final dataset. Figure 4. showing the distribution of spam, fake, and legit/ham entries across different datasets.

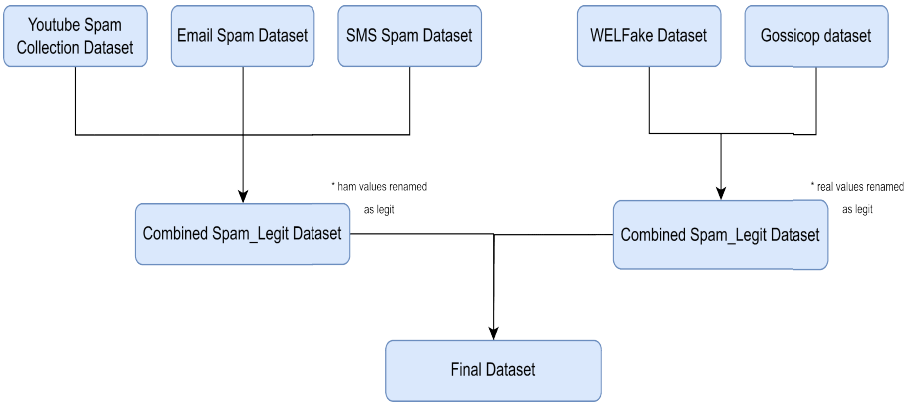

Figure 1. Data Collection Process

3.2 Data Preprocessing

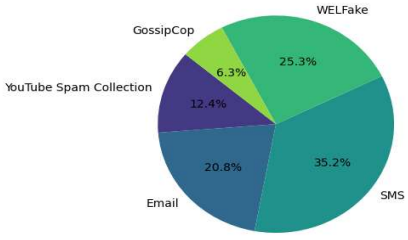

Figure 2. Distribution of Data Sources

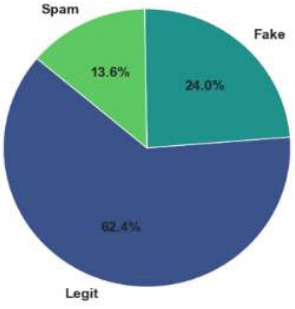

Figure 3. Distribution of Final Dataset

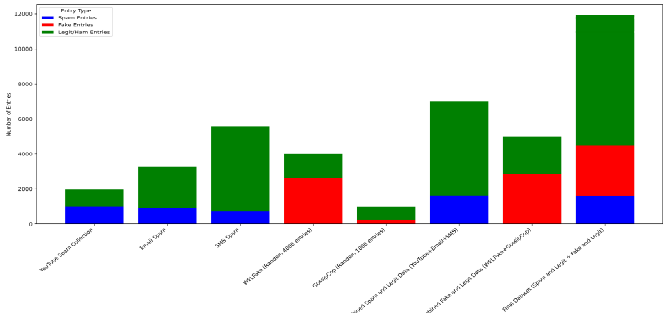

Figure 4. Distribution of spam, fake, legit /ham entries across different datasets

In this research, we utilized the Natural Language Toolkit (NLTK) for preprocessing the text data. To ensure the quality and consistency of the data across different types, a comprehensive preprocessing pipeline was implemented. Figure 5 shows various steps used in our data preprocessing stage. The following steps were performed using specific components of the NLTK library:

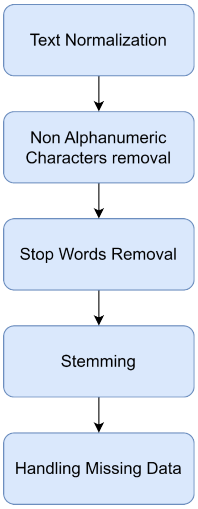

Figure 5. Step by step flow chart of data preprocessing stages

1. Text Normalization
   - Lowercasing: All of the text data was converted to lowercase through standard string methods for consistency. This was done with the lower method to equate words such as "Machine" and "machine."
   - Tokenization: The text was converted into words or tokens using nltk.word_tokenize. What tokenization does is that it makes the text broken up into manageable pieces needed for further processing.

2. Removing Non-Alphanumeric Characters
   - Removed non-alphanumeric characters like punctuation and special symbols, among others, from the text to clean the text of any noises. This ensures that only essential words are left out for further investigation.

3. Stop Words Removal
   - Commonly found stop words have been removed using the list of stop words in NLTK(nltk.corpus.stopwords). Think of stop words as noise, as it doesn't mean a lot.

4. Stemming
   - Utilized the Porter Stemmer to bring words down to their base form, such as bringing "talking" to "talk." Stemming can help to reduce dimensionality by treating different forms of a word as the same. In this way, the model generalization may be increased.

5. Handling Missing Data
   - Filling missing values with logical substitutes or removing them depending on the context.

3.3 Feature Extraction

TF-IDF(Term Frequency-Inverse Document Frequency) was used for vectorizing text data so that it can be consumed by the machine learning models. TF- IDF is the product of the following two:
Term Frequency (TF): It indicates how frequently a word is used in a document.
Inverse Document Frequency (IDF): This is what describes the importance of the word. Words that tend to appear in most documents will receive a lower IDF score.
The word's TF-IDF score combines the word's TF and IDF scores. This score then describes, in the context of the whole corpus, the importance of the word to this document.

3.4 Model Description

We experimented with various ML models for classifying the data into fake, spam, and legit categories. Below is a detailed description of each model and their respective parameters used:

A. Logistic Regression (LR): Logistic Regression is typically used for classification problems. The model estimates the probability of a class label by fitting some logistic function to the input features. It performs exceptionally well with linearly separable data, making it a straightforward and efficient choice for many classification tasks. The model uses the solver='liblinear' for efficiency with small datasets, and the L1 penalty (penalty='l1') encourages sparsity in the feature weights, which is beneficial for high-dimensional data and feature selection.
B. Support Vector Machines (SVM): Support Vector Machines make very good classifiers, more especially in high-dimensional spaces. The basic aim of the SVMs is to find the most suitable hyperplane that maximizes the margin between classes. SVMs handle well the problem of sparsity in very high-dimensional data. For instance, in text categorization, by having in place a sigmoid kernel (kernel='sigmoid') applied in this implementation, we can admit non-linear decision functions. We set gamma=1.0 to control the reach of the influence of each training sample and also to balance out complexity against performance.
C. XGBoost: It is one of the best gradient-boosting algorithms, optimized for speed and performance. It builds an ensemble of weak learners, mainly built with decision trees, that are responsible for correcting the mistakes of their predecessors. Since XGBoost is robust, scalable, and effective at handling missing data, many practitioners apply it to structured data tasks. The model contains 50 configured estimators(n_estimators = 50). This is for balancing between performance and efficiency in computations. The random state(random_state=2)is set for the reproducibility of the results.

D. Random Forest (RF): Random Forest can be explained asanextension of decision trees. It is used for both regression and classification problems. We will be using Random Forest Classifier for our particular problem. It uses multiple decision trees. Random sample of the dataset is given to each decision tree. Each decision tree gives its result and based on majority voting final result is obtained. This approach helps tackle the problem of overfitting often seen in decision trees algorithm and provides good generalization. The model comprises 50 trees (n_estimators=50) to capture diverse patterns in the data without increasing the computational requirements by much. This model has a random state (random_state=2) set, so it is reproducible.
E. Extra Trees Classifier (ETC): The Extra Trees Classifier has some differences from Random Forest Classifier like the way that node splitting is done. Actually, splitting is done on nodes in a completely random way, instead of looking for the most discriminative threshold. This extra randomization reduces variance and overfitting even more. It is this that makes the ETC quite strong in performance during classification tasks. In addition, it offers tremendous ways to handle high-dimensional data and be computationally efficient. The model builds 50 trees (n_estimators=50) and a random_state(random_state=2) to secure the robustness of performance and reproducibility.
F. AdaBoost: It is short for Adaptive Boosting. In general, AdaBoost is an ensemble approach to putting together multiple weak classifiers to build a strong classifier. Classic AdaBoost works by adding models one at a time with increasing weight on the errors made by preceding models. Then all models are pooled with a weighted vote for a final prediction. This makes AdaBoost a practical tool to boost the performance of some simple models, say, decision trees. We set our model to have 50 estimators(n_estimators=50). since, from some empirical point of view, this is a good balance between performance and computational efficiency Besides, we use a random state for reproducibility (random_state=2).
G. k-Nearest Neighbors (k-NN): k-NN Classification is used to determine class of a data point based on the class of its k-nearest data points. The majority class among the k nearest data points is selected as the class of the desired data point. It is intuitive and easy to implement. On the other hand, k-NN might get computationally intensive and lose practicality if the data becomes high-dimensional. K-NN is usually employed as a base of comparison for more sophisticated models. In this research the model uses default settings and serves as one of the simplest types of baseline model to compare with other more sophisticated models.
H. Voting Classifier: Voting Classifier is an ensemble method to make predictions based on the results of multiple models, using majority voting. We used our top three models based on performance metrics to be used as base models for Voting Classifier. Voting Classifier will aggregate the predictions from all the classifiers to arrive at a final decision, using the strengths of all individual models to enhance the overall performance.
I. Stacked Ensemble: The Stacked Ensemble model is an ensemble method which uses predictions made by different base learners to a higher-level meta-learner model. In our case, the first-layer models are Extra Trees Classifier, XGBoost, and Random Forest, with the final estimator for the last step being a Support Vector Classifier. The idea is to put models together, each with its own strengths and weaknesses, to get better overall predictive performance.

3.5 Training and Validation

We divided our dataset into training and testing sets, allocating 80% of the data for training and 20% for testing.Training dataset was used to train model and testing dataset was used to evaluate model's performance.

3.6 Evaluation Metrics

Following evaluation metrics were used to assess model performance:

Accuracy is an objective measure but could be very misleading to use when classes are imbalanced.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Precision is about the accuracy ofpositive predictions.

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

Recall relates to the completeness of the actual positives.

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

F1Score balances precision and recall.

$$\text{F1Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Error matrix is a tool used for evaluating performance of a classification model. Figure 6 shows a structure of an error matrix

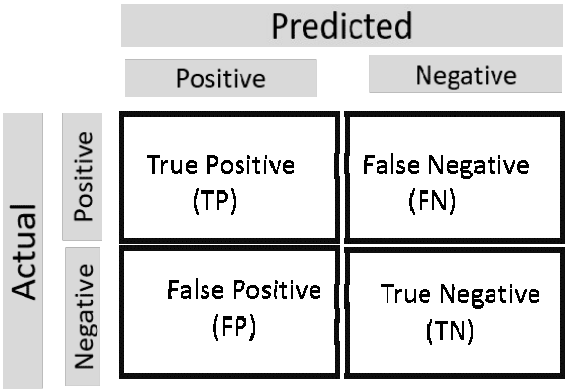

Figure 6. Error Matrix

4 Experimental Work

We assessed several ML models to thoroughly understand each model's capability to accurately classify data as fake, spam, or legitimate. All performance metrics are measured on testing dataset to examine the performance of our models on unseen data.

Table 2 shows the performance achieved by our baseline models. Figure 7. provides a comparison between our model's performances.

Table 2. Comparison of Base Model Performance

ML Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	86.38%	86.66%	86.36%	86.02%
Random Forest	87.64%	87.56%	87.64%	87.51%
Extra Trees Classifier	87.76%	87.80%	87.67%	87.77%
KNN	69.92%	76.46%	69.92%	62.41%
XGBoost	87.22%	87.78%	87.22%	86.76%
Support Vector Classifier	86.92%	86.99%	86.92%	86.58%
AdaBoost	78.96%	78.95%	78.96%	78.70%

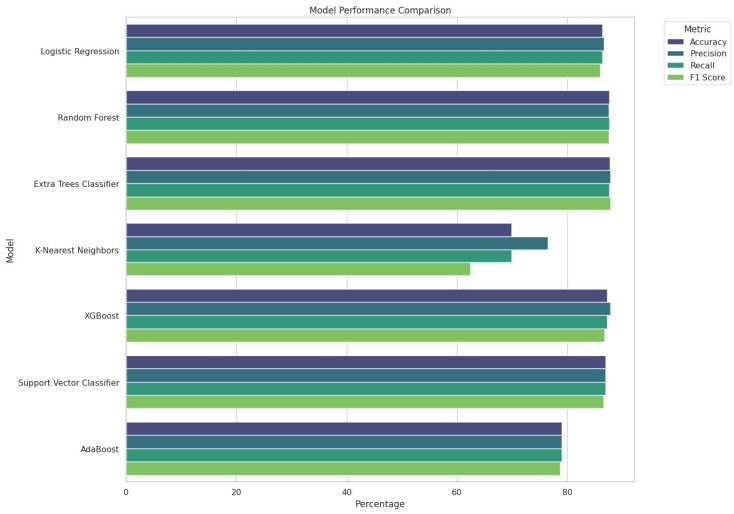

Figure 7. Base Models Performance Comparison

We also applied the Voting Classifier using the top three performing models (ETC, RF, XGBoost) to combine their predictions by majority voting.The Stacked Ensemble model uses the top three models (ETC, RF, XGBooost) as base models and SVC as the final estimator. This approach leverages the strengths of different algorithms to enhance overall performance.

Table 3. Performance metrics of Voting Classifier and Stacked Ensemble method

Metric	Voting Classifier (ETC, XGBoost, RF)	Stacked Ensemble (ETC, XGBoost, RF with SVC as final estimator)
Accuracy	89.41%	85.96%
Precision	89.17%	86.65%
Recall	89.14%	85.96%
F1Score	88.95%	85.95%

Table 3 shows performance metrics of Voting Classifier and Stacked Ensemble model. These ensemble methods are designed to use multiple models to improve overall classification performance. Figure 8. summarizes the key performance metrics for both the Voting Classifier and the Stacked Ensemble, providing a clear comparison of their effectiveness in classifying data into fake, spam, and legitimate categories.

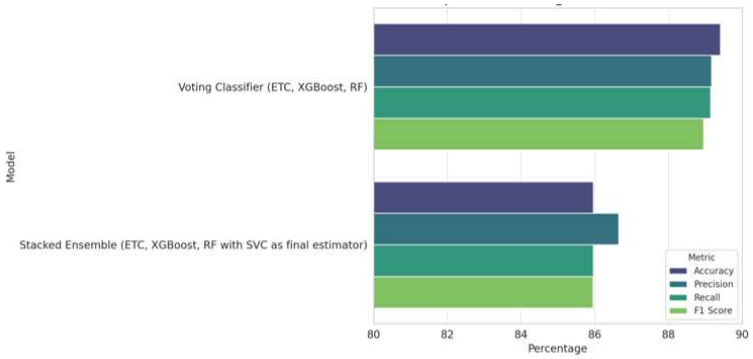

Figure 8. Comparison of Voting Classification and Stacked Ensemble model performance

Figure 9,10 and 11 shows the Classification report of top three baseline model Extra Trees Classifier, Random Forest and XGBoost respectively. Figure 12 and 13 shows the classification report of Voting Classifier and Stacking Ensemble model in classifying the information in dataset as fake, spam or legit respectively.

	precision	recall	F1 Score	support
fake	0.82	0.84	0.83	545
legit	0.90	0.92	0.91	1510
spam	0.88	0.76	0.82	332
accuracy			0.88	2387
macro avg	0.87	0.84	0.85	2387
weighted avg	0.88	0.88	0.88	2387

Figure 9. Classification Report of Extra Trees Classifier

	precision	recall	F1 Score	support
fake	0.84	0.81	0.83	545
legit	0.89	0.93	0.91	1510
spam	0.87	0.74	0.80	332
accuracy			0.88	2387
macro avg	0.87	0.83	0.85	2387
weighted avg	0.88	0.88	0.88	2387

Figure 10. Classification Report of Random Forest Classifier

	precision	recall	F1 Score	support
fake	0.92	0.74	0.82	545
legit	0.85	0.97	0.91	1510
spam	0.92	0.67	0.77	332
accuracy			0.87	2387
macro avg	0.90	0.79	0.83	2387
weighted avg	0.88	0.87	0.87	2387

Figure 11. Classification Report of XGBoost Classifier

	precision	recall	F1 Score	support
fake	0.90	0.81	0.85	545
legit	0.89	0.95	0.92	1510
spam	0.90	0.75	0.81	332
accuracy			0.89	2387
macro avg	0.89	0.84	0.86	2387
weighted avg	0.89	0.89	0.89	2387

Figure 12. Classification Report of Voting Classifier

	precision	recall	F1 Score	support
fake	0.75	0.89	0.81	545
legit	0.90	0.89	0.89	1510
spam	0.92	0.69	0.79	332
accuracy			0.86	2387
macro avg	0.86	0.82	0.83	2387
weighted avg	0.87	0.86	0.86	2387

Figure 13. Classification Report of Stacking Ensemble method

5 Result Analysis

From Figure 14 and 15 we can conclude that Extra Trees Classifier outperformed all other baseline models in terms of accuracy, precision, recall, and F1Score. Random Forest and XGBoost also performed well, with high accuracy and balanced metrics. Logistic Regression, SVC, and AdaBoost showed good performance but were outperformed by the ensemble methods. k-NN had the lowest performance, indicating it may not be suitable for this particular classification task.

These results demonstrate the effectiveness of ensemble methods like Extra Trees and Random Forest in handling complex classification tasks, leveraging multiple decision trees to improve predictive performance. The use of advanced models like XGBoost provided a similar level of accuracy and balance across key metrics.Extra Trees Classifier achieved an accuracy of 87.76%, a precision of

87.80%, a recall of 87.67%, and an F1Score of 87.77%.Random Forest achieved an accuracy of 87.64%, a precision of 87.56%, a recall of 87.64%, and an F1Score of 87.51%. XGBoost model achieved an accuracy of 87.22%, a precision of 87.78%, a recall of 87.22%, and an F1Score of 86.76%.

The Voting Classifier (ETC, XGBoost, RF) achieved an accuracy of 89.41%, a precision of 89.17%, a recall of 89.14%, and an F1Score of 88.95%.TheStacked Ensemble (ETC, XGBoost, RF with SVC as final estimator) achieved an accuracy of 85.96%, a precision of 86.65%, a recall of 85.96%, and an F1Score of 85.95%.

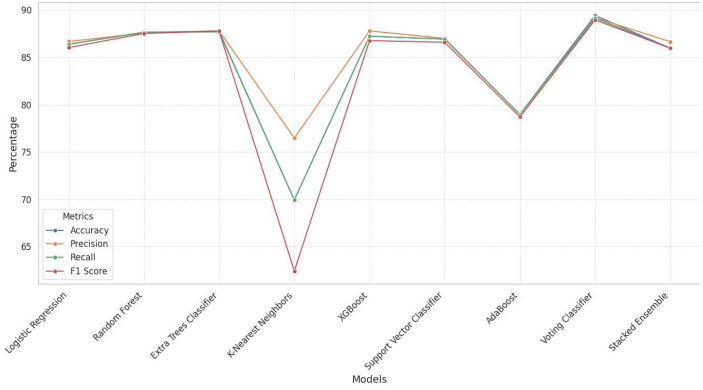

Figure 14. Overall Performance Metrics Comparison of Different Models

It is observed from the results that the best overall results were when The Voting Classifier was used to combine the three top models together, whereby ETC made the best overall performance in the individual models, followed by Random Forest and then XGBoost, which actually indicates that using ensemble methods, specifically Voting Classifier in our case, can improve overall performance.

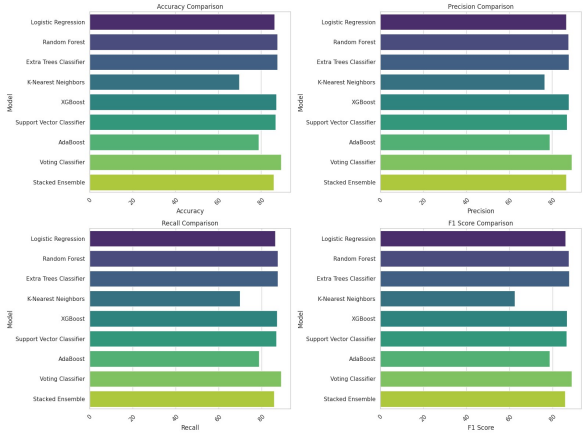

Figure 15. Detailed Performance Comparison of Machine Learning Models

6 Conclusion and Future Work

In this work, an advanced classification system that uses machine learning techniques was developed to appropriately identify fake, spam, and legitimate content. In this regard, the Voting Classifier proved even more effective in comparison to the other models. However, models like Extra Trees Classifier also gave decent performance followed by Random Forest Classifier and XGBoost classifier in classifying information under complex data patterns. Logistic Regression was somewhat basic in its effectiveness, though along with the KNN model, it proved less effective compared to the models that were superior.
This research work has advanced prior works by using a wide range of data sources and a comprehensive machine learning model to categorize information into distinct categories. However, a consolidated approach has been more effective compared to those previously modelled, either for fake news or for spam, respectively, as it has enhanced the ability to conduct multicategory classification in an effective manner. Nevertheless, the study is not without limitations. The diversity of data in the dataset and coverage across a plethora of dimensions subsuming all the variants of fake news and spam is much influential to the effectiveness of the model and its generalization.

Future research should integrate diversified data sets and exploration on reducing the model bias.In future timeframe, this research project will look forward to incorporating more data sources to enhance the strength and validation synchronously, as well as exploring deep learning mechanisms associated with transfer learning with the aim of increasing performance. Moreover, this system, which is taken into consideration in the future, must be tested in real-world scenarios so as to establish its real-time effectiveness and reliability.

The application value of the study is practical. The classification system developed in this project can be used to make digital information more reliable. This can be done because fake news and spam can be automatically sorted out. Such things could be used in social media platforms, email services, and news aggregators that will ensure the veracity of information and protect the user from misinformation.

References

[1] Allcott, H., & Gentzkow, M. (2017). The role of social media in spreading fake news during the 2016 election. Journal of Economic Perspectives, 31(2), 211-236. https://doi.org/10.1257/jep.31.2.211

[2] Vosoughi, S., Roy, D., & Aral, S. (2018). Dissemination of true and false news on the internet. Science, 359(6380), 1146-1151. Retrieved from https://science.sciencemag.org/content/359/6380/1146 on January 15, 2024.

[3] Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., &Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094-1096. https://doi.org/10.1126/science.aao2998

[4] Pennycook, G., & Rand, D. G. (2018). The implied truth effect: Adding warnings to some fake news stories raises perceived accuracy of unwarned stories. Management Science, 66(11), 4944-4957. https://dx.doi.org/10.2139/ssrn.3035384

[5] Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management, 57(3), 102025. https://doi.org/10.1016/j.ipm.2019.03.004

[6] Basak, K., Ekbal, A., & Bhattacharyya, P. (2019). A deep ensemble framework for fake news detection and multi-class classification of short political statements. In Proceedings of the 16th International Conference on Natural Language Processing (pp. 9-17). Retrieved from https://aclanthology.org/2019.icon-1.2/ on January 15, 2024.

[7] Karimi, H., Roy, P., Saba-Sadiya, S., & Tang, J. (2018). Multi-source multi-class fake news detection. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 1546-1557). Retrieved from https://aclanthology.org/C18-1131 on January 15, 2024.

[8] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Detecting fake news on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36. https://doi.org/10.1145/3137597.3137600.

[9] Zhou, X., &Zafarani, R. (2020). Survey of fake news: Core theories, detection techniques, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40. https://doi.org/10.1145/3395046.

[10] Guzella, T. S., &Caminhas, W. M. (2009). A review of machine learning methods for spam filtering. Expert Systems with Applications, 36(7), 10206-10222. https://doi.org/10.1016/j.eswa.2009.02.037

[11] Almeida, T. A., &Yamakami, A. (2012). Advances in spam filtering techniques. In D. Elizondo, A. Solanas, & A. Martinez-Balleste (Eds.), Computational Intelligence for Privacy and Security. Studies in Computational Intelligence, vol 394 (pp. 99-104). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-25237-2_12

[12] Madani, M., Motameni, H., & Roshani, R. (2023). Fake news detection using feature extraction, natural language processing, curriculum learning, and deep learning. International Journal of Neural Systems. https://doi.org/10.1142/S0219622023500347

[13] Maqsood, U., Ur Rehman, S., Ali, T., & others. (2023). An intelligent framework based on deep learning for SMS and e-mail spam detection. Computational Intelligence and Soft Computing, 12(1), 6648970. https://doi.org/10.1155/2023/6648970

[14] Sumathi, M., & Raja, S. P. (2023). Machine learning algorithm-based spam detection in social networks. Social Network Analysis and Mining, 13(2), 1086-1098. https://doi.org/10.1007/s13278-023-01108-6

[15] Alberto, T. C., &Lochter, J. V. (2017). YouTube Spam Collection. UCI Machine Learning Repository. https://doi.org/10.24432/C58885

[16] https://www.kaggle.com/datasets/venky73/spam-mails-dataset

[17] https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset

[18] https://www.kaggle.com/datasets/vcclab/welfake-dataset

[19] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). FakeNewsNet: A data repository with news content, social context and spatiotemporal information for fake news research. Companion Proceedings of the The Web Conference 2018, 100-104. Retrieved from https://github.com/KaiDMML/FakeNewsNet