OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 06 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

FAKE NEWS DETECTION ON SOCIAL MEDIA PLATFORMS USING MACHINE LEARNING AND ENSEMBLE TECHNIQUES

*[1]HAFSA HUSSAIN**

[2]NAEEM ASLAM

[3]MUHAMMAD FUZAIL

[4]TANZEEL-UR-REHMAN

[5]FARWA NAZIM

[1, 2, 3, 4, 5]Department of Computer Science, NFCIET, Multan, Pakistan.

Corresponding author: Hafsa Hussain (** hussainhafsa191@gmail.com **)

Article Info

Abstract

Today, information plays a key role like gold did earlier and news is now mostly read on social media rather than in daily newspapers (18% to 39% share). Due to this change, stories that aren't real and are meant to shape people's opinions or harm anyone's reputation get more attention than genuine news reports on social media. The objective is to help identify reliable news from fake news on social media apps by analyzing and comparing machine learning algorithms.

In this study, many machine learning classifiers like K-Nearest Neighbors (KNN), Logistic Regression and Support Vector Machines (SVM) are tested and made better with valuable data preparation techniques. Ensemble methods are tested against single classifiers to see which gives the most accurate, precise and recall results. Investigating over 6,335 articles, the study used a 30-70 method of dividing into training and testing sets to determine which processing helps most and tests the use of the proposed models in real time.

The results prove that fake news systems work better, assisting media, social networks and the public. The research aids in academia by collecting and comparing different methods for classification and preprocessing, especially to manage problems related to Disinformation, Distortion and Disruption in the digital age.

Keywords:
Fake news detection, Machine learning, Ensemble methods, Misinformation classification.

CHAPTER # 1 - INTRODUCTION

Information today is a critical asset more so in this modern age and is described as the ‘‘Gold’’ of the twenty-first century. The audience, which consumes news, has shifted the sources that they use to obtain information from print to social media. Records that compared the usage of print media in 2013 which was 63% with that of April 2020 which was 26% while that showed social media usage which was 18% in 2013 had risen to 39% in April 2020 depict this change. This change has however worsened the problem of fake news –articles that are made up with the intent of influencing the opinion of the public or tarnishing the image of an individual or an organization.

False narratives that are produced with the intention of influencing public opinion or defaming an individual are known as fake news. On social media platforms, bogus news has been shown to attract more shares than true news. This is corroborated by data on the well-known social networking site "Facebook." In contrast, the top 20 phoney news received 48K comments and 43K more reactions than the top 20 actual news. It has been observed that social media platforms, particularly, have facilitated the spread of this kind of news through features like sharing, commenting, and friend-tagging.

1. High-Level Architecture of Fake News Detection System

This diagram illustrates the overall architecture of a fake news detection system, highlighting the major components like data collection, preprocessing, model training, and classification.

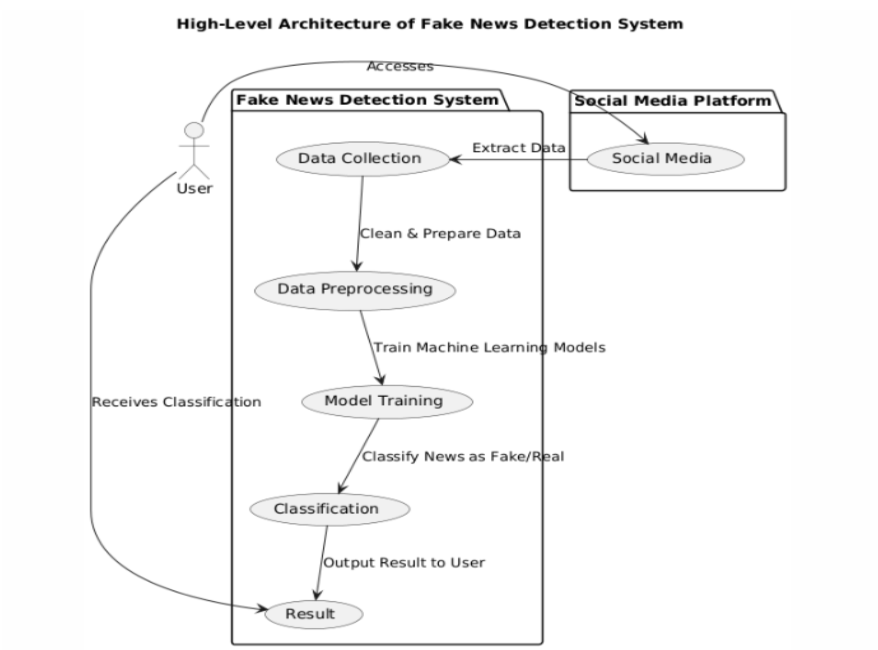


Figure 2. Architecture Diagram

This study’s specific application area is in identifying fake news on social media platforms using machine learning and ensembling. Although there has been progress, today’s methods of news checking face the problem of the large amount of data that circulates on social networks. The main goal of this research is to enhance classifier’s performance and accuracy of identifying fake news using the best data preprocessing methods and combining algorithms. Although, work is going on currently to come up with better classifiers and pre-processing techniques there are certain limitations such as achieving high accuracy rates and being able to design the fire detectors to be real-time.

2. Workflow of Machine Learning-Based Fake News Detection

This diagram shows the workflow involved in detecting fake news using machine learning. It includes steps like data acquisition, feature extraction, model training, and validation.

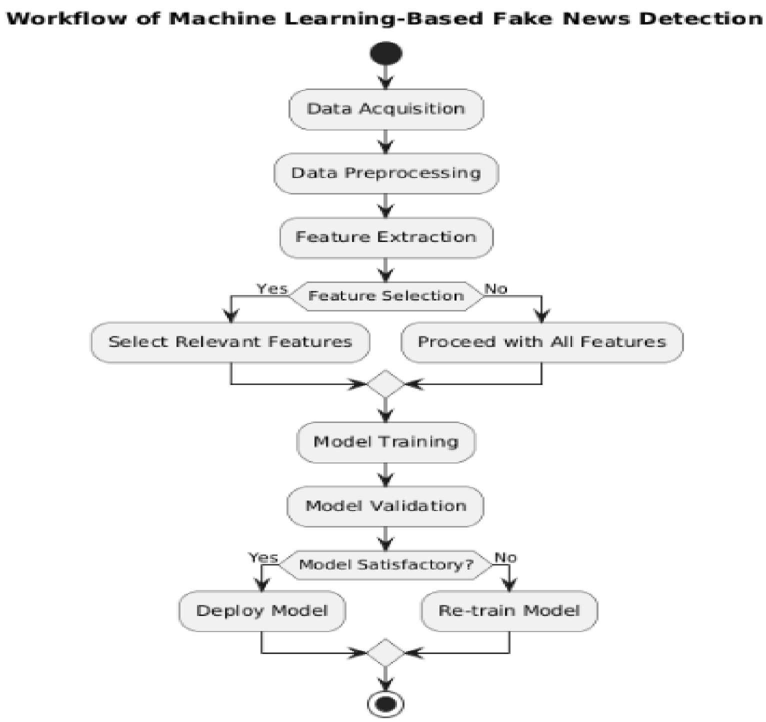


Figure 3. Workflow Diagram

Problem Statement

Critical issue regarding Social media is the inability to filter Reliable and Unreliable Information because of the appearance of fake news, which can dominate social platforms and mislead people. Current approaches to news verification are inadequate in handling this problem primarily because there is too much information that circulates on social media. Since fake news are a major problem in the modern world [10], there is a necessity to develop systems that can identify such news quickly and with high accuracy. The objective of this study is to build and compare several machine learning classifiers with the specific goal of enhancing their performance levels in the context of fake news detection by carrying out data pre-processing.

The main challenges in fake news detection include the variability and complexity of textual data, the need for real-time processing, and the balance between classifier complexity and performance.

Research Questions

- How well do different machine learning systems and ensemble ways work when finding fake news on social media sites?
- Which of the available pre-processing techniques greatly improves the predictive enforcement of machine learning models in identifying fake news?
- Do ensemble learning methods do better or worse than single classifiers when it comes to detecting fake news in real time?

Objectives

- See how the developed models can work in actual situations to spot fake news.
- Increase both accuracy and performance in identifying false news posts on the internet by using machine learning and ensemble methods.
- The performance of different machine learning techniques such as KNN, Logistic Regression, SVM needs to be studied and compared when used for identifying fake news—giving importance to accuracy, precision, and recall.
- Find out how the actions of stemming, lemmatization, and removing stop words impact the performance of machine learning models in catching fake news.

Significance of the Study

This research enriches the existing knowledge within the academic domain due to offering findings pertaining to the performance of the various classifiers and the pre-processing methods in fake news detection. It provides the concepts to improve regarding ensemble methods in this respect and provides a comparative study of different methods [8]. The conclusion of the following research implies practical implications for news verification in social media applications. Potential stakeholders that can benefit from enhanced fake news detection methods are media houses, social networking platforms, and the public as a whole due to minimal relaying of fake news and its effects.

Chapter 2: Literature Review

The purpose of this literature review is to provide a comprehensive foundation for the research on "Fake News Detection of Social Media Platforms Using Machine Learning and Ensemble Techniques". The scope of the review covers the theoretical background, key concepts, and the evolution of research in this field, as well as a critical analysis of the existing work and the identification of research gaps. The chapter is organized to systematically address the various aspects relevant to the study.

2.2 Review of Related Work

2.2.1 Overview of Existing Research

In order to respond to the need for fake news detection on social media several studies have been conducted. Some of the methods examined in extant literature include: Linguistically speaking [13], Social network analysis [14], and Machine learning based methods. They have been intended to determine the features and trace the behaviors of real fakes from realistic ones and to design the strong detection models.

Most of the research performed in the subject area of fake news detection in the last few years has been centered around developing automatic solutions based on ML algorithms. Research has looked into different strategies which have been proposed such as those based on supervised and unsupervised learning, Deep learning, and Ensembling. classification, where a model is trained with labeled data, has been applied very much because of the success in binary cases such as those involving real and fake news cases [15]. In the same vein, clustering techniques have been used to monitor emerging patterns of fake news with the objectives of accomplishing this task with no human-labeled data.
There exist many literature reviews that present a review of the various methods for identifying fake news in general [4,15,16,17,18,8,19,20,21,22 and 24]..

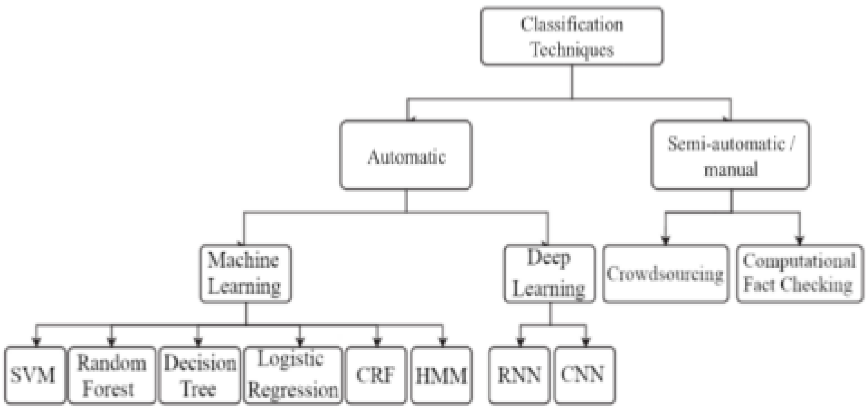


Figure 1.Different approaches proposed for detecting fake news [13]

2.2.3 Comparative Studies

There have been a lot of studies looking into how well different machine learning techniques and aggregation methods can pick out fake news. For this purpose, Shu et al. [19] analyzed logistic regression, decision trees, and random forests to determine whether news articles are real or false. Oshikawa et al. [20] conducted a study to see how deep learning methods do in fake news classification compared to traditional machine learning models.

2.2.4 Case Studies and Applications

Various case studies and examples from real instances are mentioned in the literature to show how fake news detection works in real life. Some researchers have used their methods on real-life datasets which include the BuzzFeed political news corpus [21] and the Kaggle fake news challenge dataset [22] in order to compare how the different models perform in real life. Moreover, some manuscripts have examined the possibility of incorporating the fake news detection systems into the social media interface considering the feasibility of deploying such solutions [23].

2.2.5 Critical Analysis of the Literature

Comparing to previous studies, the research has discussed various ways of developing the fake news detection and there are still some drawbacks and challenges. A major shortcoming of most existing research is that some have only centered on particular types of fake news or were conducted only one or two platforms, so these results seem not to represent the whole picture [24]. Furthermore, the achievement shown for the proposed models can be different concerning other datasets and contexts, and that is why the problem requires more universal methods [25]. Moreover, having studied the matter, the authors demonstrate the necessity of including ethic aspects and controlling bias in the fake news detection system construction [25].

2.4 Identification of Research Gaps

A big challenge today in fighting fake news is finding models that are precise and flexible enough to deal with the continuous changes in what appears on the internet [29]. It is challenging for many methods to identify the new tactics of fake news because the strategies keep growing and changing [30]. Besides, experts in this field argue that the issue of fake news requires attention to issues such as bias, protection of privacy, and protecting freedom of expression [31].

This review has explained the important studies done on detecting fake news with machine learning and by combining methods. Topics addressed were basic definitions, how fake news has evolved, various methodologies, their comparisons, and uncovered the present research challenges. They confirm that robust and adjustable detection tools are very important, and also stress that ethical matters and opportunities for new research should be considered.

Chapter 3: Methodology

In this chapter, the approach for the conduct of the research that focuses on "Fake News Detection of Social Media Platforms Using Machine Learning and Ensemble Techniques" shall be described; this involves the acquisition of the dataset, pre-processing, implementation of classifiers, and assessment. This diagram shows the workflow involved in detecting fake news using machine learning. It includes steps like data acquisition, feature extraction, model training, and validation.

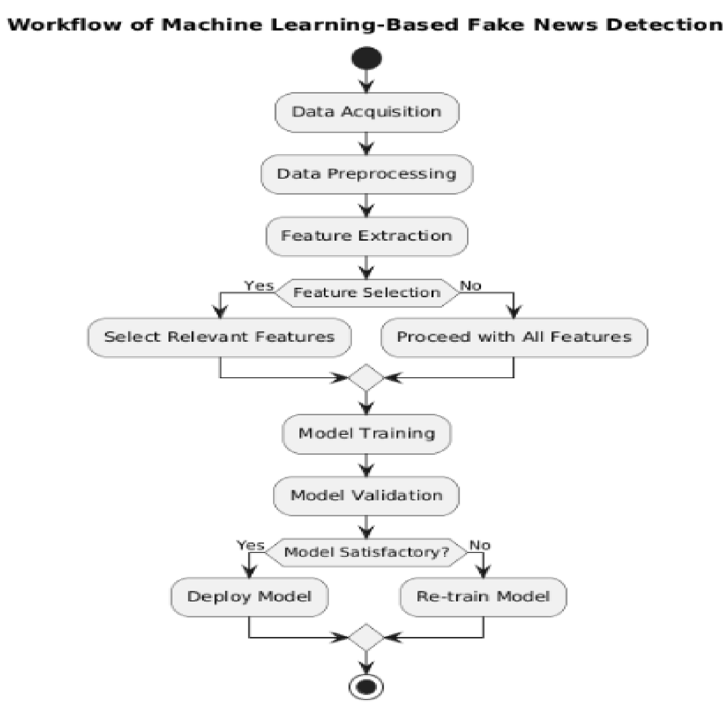


Figure 3. Methodology

Dataset Collection

This research made use of a dataset obtained from Kaggle, a well-known database for data science projects and competitions. It is named the "Fake and Real News Dataset" and is made for classification tasks in which news items are categorized as real or fake. So, it becomes a good choice for teaching machines to identify if a news report is true or not. The file containing the data is available at the given link:

https://www.kaggle.com/datasets/amirmotefaker/detecting-fake-news-dataset

A file called news.csv is present, full of input data that have already been labeled for binary classification. In every record, there are the following features included:

- Title: The name that describes the main part of the news.
- The body of the article contains the message and gives details to the reader.
- This field shows the topic that the article pertains to, such as politics or world news.
- The publication date is identified in each article.

These aspects provide the information that the machine learning model requires. The aim of the classification is to determine if a news article is real or fake, and this is why the label is the target.

Data Splitting

Once the dataset is obtained, the latter is divided into training and testing sets. This enable model training on training set and then test the models on the test set. We employ 80/ 20 format for training exercise and independently exercises testing. To enhance the credibility of our models, cross-validation approach is also used.

Data Exploration

A preliminary investigation of data is done before getting into details with the pre-processing step. This involves mapping for the completeness of the variables, mapping of fake and real news articles, word frequency distributions among others. Figure 2 provides an insight to the sequence of operations on collection and preparation of the dataset.

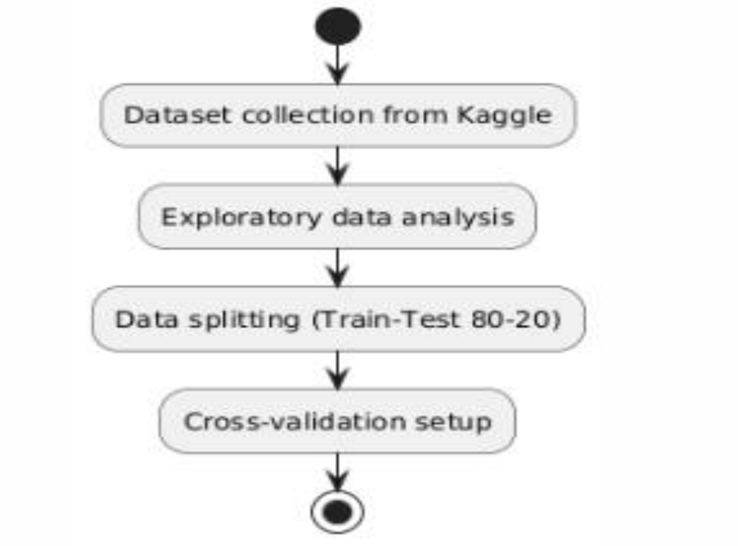


Figure 2. Dataset collection and splitting process

Data Preparation

Once data was obtained, several data cleaning steps were performed to ensure the data is in the right format suitable for training of the machine learning models. It must be noted there were data fields like date and subject in the dataset which had no impact in the identification of fake news. These columns were omitted for the sake of reduction of the numerical data set.

Handling Missing Data

Wherever there were such cases of null values, missing values etc. , the columns were reviewed and the best action was taken. To facilitate binary classification, labels were assigned to the real and fake news articles:

- The Manure articles were made up of fake news, while the ‘Real news articles’ were labeled as 1.
- False articles were classified as 0.

These labels are important for training a machine learning model and in the case of classification they help the model to classify between real and fake news during the prediction phase..

Data Pre-processing

Mobile text data particularly from social media platforms is messy and unformatted. Regardless of the situation, however, this data needs to be pre-processed before it will significantly improve the functioning of machine learning algorithms.

For this purpose in our study, unstructured news articles were tokenized to words using the NLTK (Natural Language Toolkit) library in Python. Tokenization also makes further text-processing to take place easily. Those are the words which do not contain much important information like ‘and’, ‘the’, ‘in’, etc. The meaning of ‘important’ words can be weakened by such stop words. Exclusion of the stop words carves out the model a path where only useful words are used in making the prediction.

Machine Learning Classifier Implementation

After the text data is pre-processed, several predictive models under the category of machine learning classifiers are applied in the task of fake news detection.

K-Nearest Neighbors is a basic yet highly efficient classification algorithm which categorizes data points according to the majority number of its closest neighbors. For the purpose of the fake news detection, we utilized KNN algorithm from the Python's sklearn library.

Hyperparameters: Here, k neighbors whereby k represents the degree of neighbors is still a hyperparameter into KNN. We make a series of tests in order to define which of the values of k should be chosen in our case. Distance Metric: In this case, we calculate distance between news articles denoted by their feature vectors using the Euclidean distance.

Logistic regression is an often-used classifier algorithm to approximate the probability of the pattern to be belonging to a particular class. In our case, it helps in the classification of articles as the real one or fake news. Support Vector Machine (SVM) is an all-round classifier that remarkably functions in the high dimensional space. As mentioned before, SVM is efficient in the text classification application areas which include dealing with fake news.

Ensemble Techniques

Ensemble learning refers to utilizing several classifiers simultaneously where from each of them, the improvement of the whole estimation will be made. In this study, we explore two ensemble techniques: While Random Forest has been named a part of the evolution of both the algorithms and computers, a substitute for this shortcoming has emerged – the decision tree bagging.

Two classifiers are taken as standard to judge and measure the performance of Passive Aggressive Classifier and Multimodal Naive Bayes over detection of ensemble learning.

These ensemble techniques are executed and used with the help of the sklearn library. We also compare the performance of the ensemble methods with the performance of any of the classifier methods.
The Passive-AGGRESSIVE Classifier is an on-line learning algorithm which updates the model only when the classifier is wrong. Lastly, it is referred to as 'passive' if the prediction is done correctly while it is referred to as 'aggressive' if the prognosis is incorrect, and the model will immediately adjust for this mistake.

The Multinomial Naive Bayes algorithm is a learning algorithm based on the Bayes' theorem which states that all the attributes of a data set are independent of one another given the class. This classifier is very suitable for text classification such as filtering out spam & malware and sentiment analysis.

Evaluation Metrics

To assess the performance of the classifiers, we use the following evaluation metrics:

- Accuracy: The percent of the non-ambiguous news articles classified correctly.
- Precision: The degree of truth of the positive results in proportion to the total positive results that were obtained.

- F1-Score: The middle point between precision and recall in order to obtain balanced parameters for evaluating an algorithm.
- ROC-AUC: The Receiver Operating Characteristic, which assess the model's performance based on the area under the curve.

These metrics give a holistic assessment of the classifiers' capability of identifying fake news. We also plot the confusion matrices for every classifier in order to have the classification results more comprehensible.

Chapter 4: Results

The chapter gives readers a detailed explanation of all the steps in the Fake News Detection project. The main concern is on how the data is being used, as well as the details that can be extracted with each step — right from getting and preprocessing the data to analysis, model design and assessment, and the results formed. We will go on to explain how the project divides real and false news by using a range of techniques, talking about how they affect the results.

1. Dataset Exploration

It includes 6,335 news articles where the label column indicates whether the text is genuine 'REAL' or fake 'FAKE'. It includes four columns: Unnamed: 0, title, text, y (label). Here, the text column is the body of the article as it contains mostly all the information that the model is trained with, the title on the other hand offers some background information. The first steps in data exploration this included the shape of data, the data types of the columns and the memory usage analysis implied that the data set is very neat and occupies only 198.1 KB of the memory space.

Basic information about the dataset structure is:

- **Shape**: (6,335, 4), indicating 6,335 records and 4 columns.
- **Columns and Types**: The dataset contains three object-type columns (title, text, and label) and one integer column (Unnamed: 0).
- **Missing Values**: No missing values are present, confirming the dataset is complete and ready for further analysis.

In terms of label distribution:

- **REAL**: 3,171 articles
- **FAKE**: 3,164 articles

This balance is advantageous as it reduces the likelihood of model bias towards a particular class.

```
...  Dataset Shape: (6335, 4)
     Column Names: Index(['Unnamed: 0', 'title', 'text', 'label'], dtype='object')
     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 6335 entries, 0 to 6334
     Data columns (total 4 columns):
      #   Column      Non-Null Count  Dtype
     ---  ------      --------------  -----
      0   Unnamed: 0  6335 non-null   int64
      1   title       6335 non-null   object
      2   text        6335 non-null   object
      3   label       6335 non-null   object
     dtypes: int64(1), object(3)
     memory usage: 198.1+ KB
     None
```

Figure 1. Dataset Overview

The even distribution reduces model bias especially to one of the two classes hence enables for development of a more balanced model without the need for other sampling methods. That is also convenient in terms of keeping the simple accuracy measure as when it rises, it will rise due to the fact that the model performs well, not because of the class imbalance.

2. Data Preprocessing

Cleansing of data is very important especially when working with large text documents which requires preprocessing to prepare for pattern recognition mathematical operations by transforming them to structures forms that can be quantifiable. In this project, preprocessing involved several steps: this entails erasing stop words, converting words into root forms, breaking down text into words and appropriately formatting it for input in most machine learning algorithms.

Text Cleaning and Tokenization

Analysis of Text Length

It was clear from a histogram of the length of articles that there was a big difference in the number of words each article contained. The problem was solved by using a more advanced way to display features, choosing TF-IDF (Term Frequency-Inverse Document Frequency).

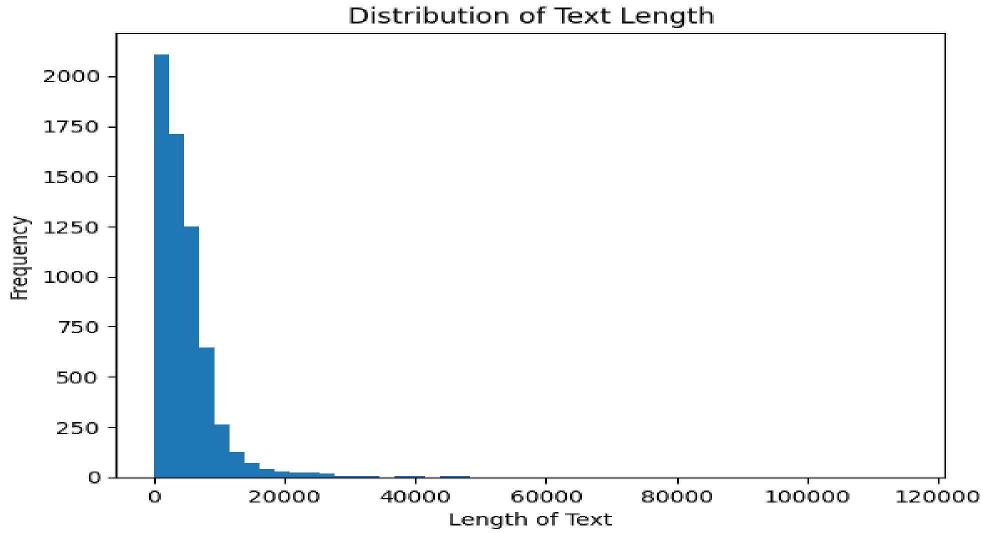
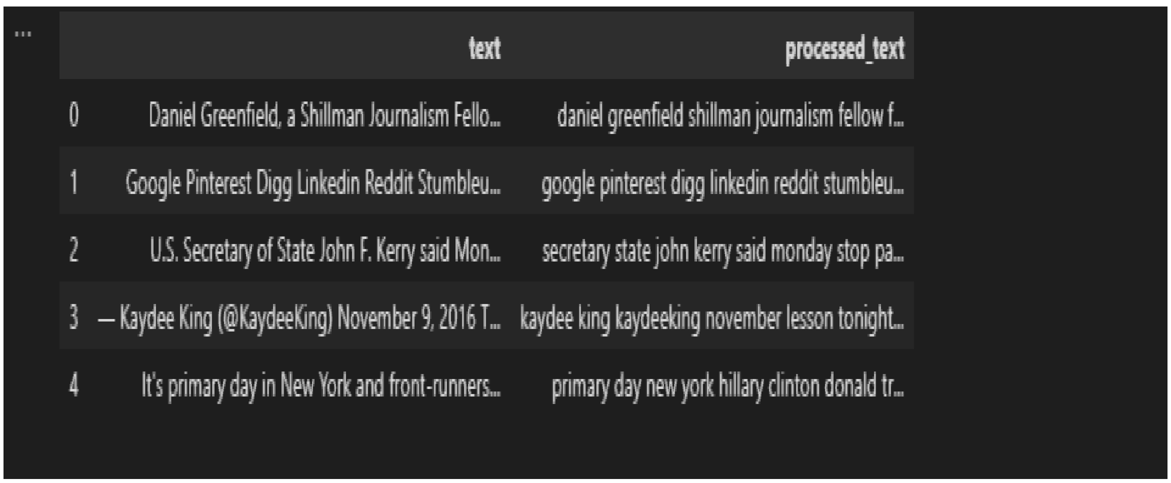


Figure 2. Words in news

Thanks to TF-IDF, fewer unique words appeared in the dataset, which improved how well the model was trained. After all the above steps, tokenization was carried out in the final step of text preprocessing. Breaking every article into its component words led to a new column called processed_text that includes the clean and tokenized texts.



	text	processed_text
0	Daniel Greenfield, a Shillman Journalism Fello...	daniel greenfield shillman journalism fellow f...
1	Google Pinterest Digg Linkedin Reddit Stumbleu...	google pinterest digg linkedin reddit stumbleu...
2	U.S. Secretary of State John F. Kerry said Mon...	secretary state john kerry said monday stop pa...
3	— Kaydee King (@KaydeeKing) November 9, 2016 T...	kaydee king kaydeeking november lesson tonight...
4	It's primary day in New York and front-runners...	primary day new york hillary clinton donald tr...

Figure 3. News Preprocessing

3. Vectorization Using TF-IDF

To ensure the calculations could be completed efficiently, the top 5,000 important terms were used in TF-IDF. This resulted in generating a matrix where each article's features are recorded using 5,000 dimensions. Each value in the vector shows how significant a certain term is for that article. By cutting down on the number of features, we managed to speed up the process without much decline in performance.

4. Train-Test Split

To make sure the model's performance was evaluated properly, the dataset was split using an 80 to 20 ratio. The end result of this process was 5,068 samples used for training, and 1,267 for testing. With the data split, the model can use enough resources to learn, and still have enough samples to verify its predictive abilities.

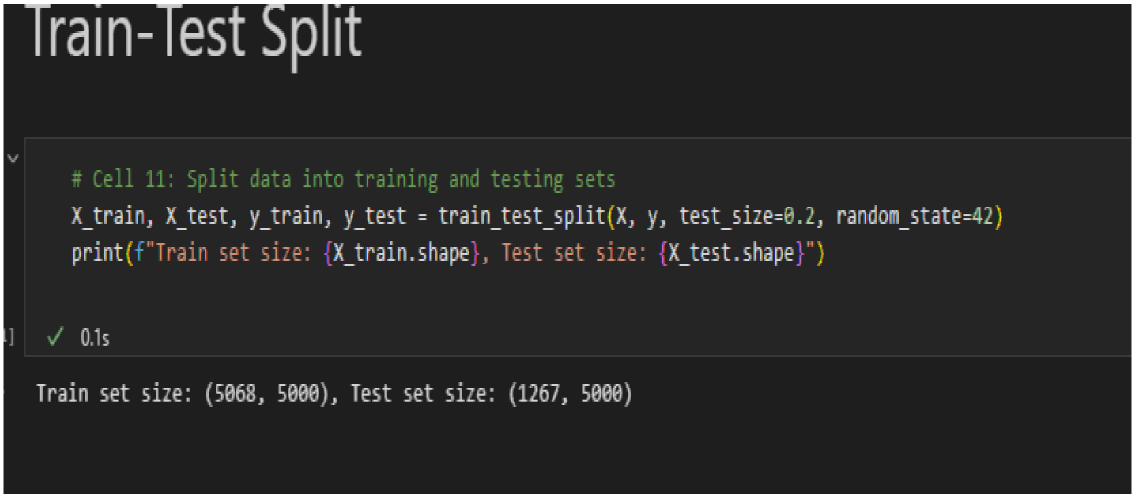


Figure 4. Dataset Distribution

- Training Set: 5,068 samples
- Testing Set: 1,267 samples

Train-test split was important to get an idea about how well all the models have performed in unseen data as in a real-life scenario the model is exposed new articles time and again.

5. Model Implementation and Evaluation

The main qualitative measure used to check the model's performance was accuracy. Besides, precision, recall, F1-score, and confusion matrix evaluations were employed to achieve a higher level of qualitative analysis.

5.1 K-Nearest Neighbors (KNN)

When applied to the dataset, K-Nearest Neighbors got an accuracy of 65.19%. Although KNN is easy to use, it struggles to work with data that has many dimensions, including text represented by TF-IDF.

A major issue of KNN occurs when the amount of data in different categories becomes very large. If the number of features goes up, the distance calculation that is used in KNN gradually becomes less reliable. The longer the number of dimensions, the weaker the impact of Euclidean distances; therefore, the accuracy of the model for telling real from fake news decreases.

While there are these restrictions, KNN managed to qualify basic classification. Its low results, however, show that this algorithm might not be ideal for this task using text.

- **Accuracy**: 0.6519

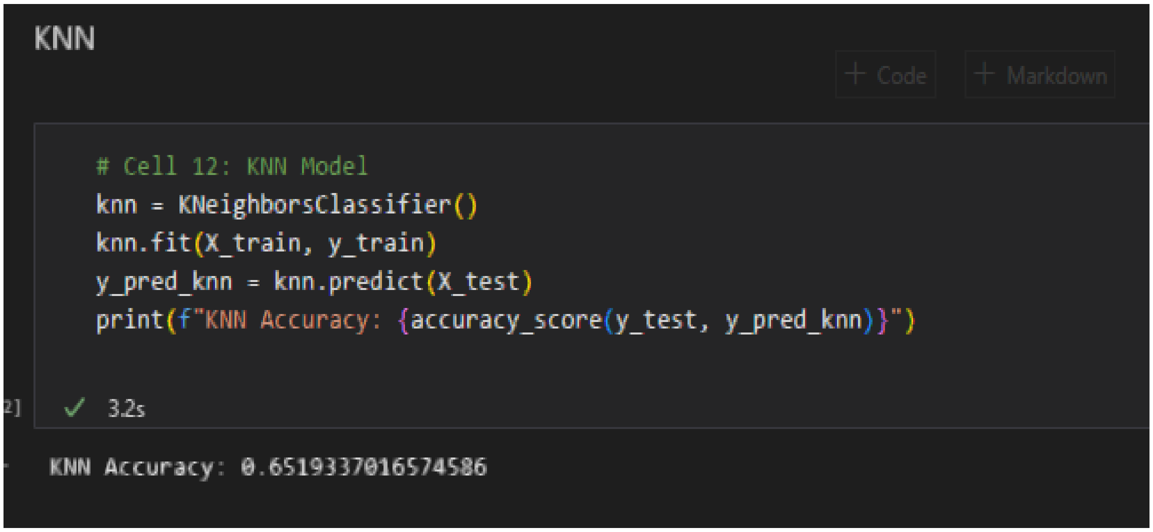


Figure 5. KNN

To sum up, while KNN could serve as a basic classifier, it struggled because of the size of the feature space and its dependency on distance measures in many dimensions. For this application, they make the method not very effective.

5.2 Logistic Regression

The binary linear classifier known as the Logistic Regression model performed very well and reached a test accuracy of 91.08%. This algorithm does its job very well with TF-IDF, as it helps identify the best features to separate two classes by measuring how these features are related to class membership.

Accuracy: 0.9108

Figure 6. Logistic Regression

So, it appears that Logistic Regression is efficient in spotting the language differences between real and fake news content. The results point to the fact that dividing data into groups is well-done using a linear line, since the text data was processed with TF-IDF and remained clean.

5.3 Passive Aggressive Classifier

Out of all the models looked at, the Passive-Aggressive Classifier was the best, reaching an accuracy level of 92.26%. It is a good method for large-scale classifications of text because it updates its parameters only if an error happens in the prediction. Because it adjusts the decision boundary with every new input, this algorithm is both effective and ready to change.

Accuracy: 0.9226

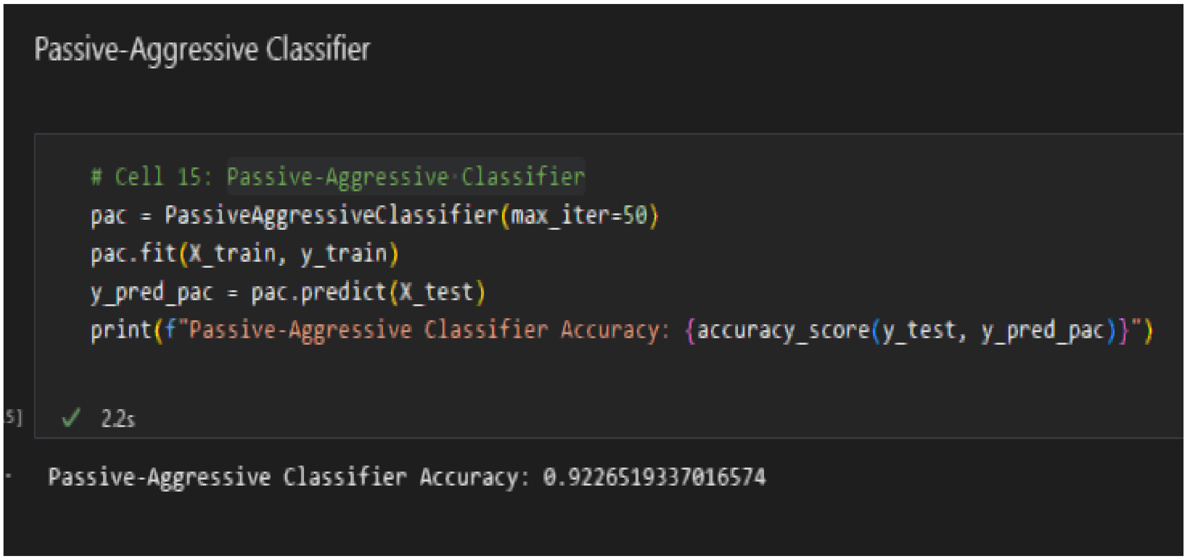


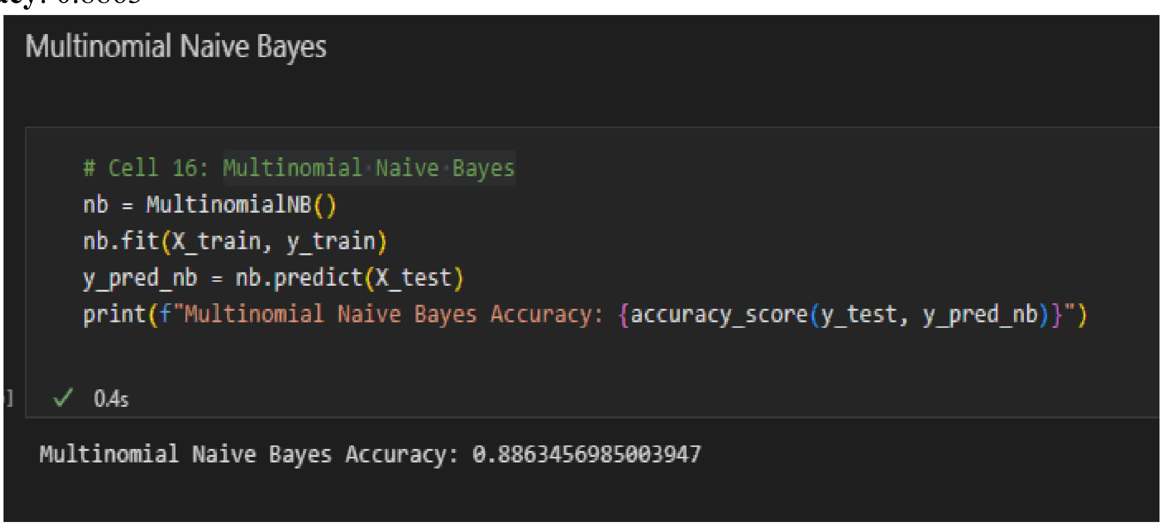
Figure 7. Passive Aggressive Classifier

Since it can pick up new trends fast and adapt little by little, this type of classifier works very well at spotting fake news in the moment.

5.4 Multinomial Naive Bayes

Multinomial Naive Bayes was able to achieve an accuracy score of 88.63%, making it a dependable base for text classification. Basically, it relies on features being independent of each other, which facilitates easier computing and letting the model manage many features at once.

Accuracy: 0.8863



The algorithm is unable to represent difficult word connections, but it still works very well. This makes it an ideal choice when the amount of computing is limited and when urgent activation is necessary.

5.5 Random Forest Classifier

When using Random Forest, the classifier with decision trees reached an accuracy of 91%. The fact that multiple trees are applied makes it less likely to overfit and more skilled at spotting complex patterns.

Figure 8. Random forest

Despite being very accurate, Random Forests consume more computing power, so they cannot handle online functions. Yet, thanks to their strong features and batch handling skills, they are suitable for offline use in systems with extra computer resources.

6. Evaluation

After assessing the model's accuracy, precise, recall, F1-score, and the confusion matrix were used to give further insights into how the model did.

The term precision show what portion of the predicted positives is truly correct. Recall looks at the number of real positives the model was actually able to discover. To avoid bias in the metric, F1-score (harmonic mean of precision and recall) was used to mix these two numbers into a single figure.

Misclassification patterns could be easily seen by using confusion matrices. Logistic Regression and the Passive-Aggressive Classifier were able to achieve strong results and make very few errors, which means they would be suitable for the task. KNN, however, had a higher rate of mistakes, mostly because the TF-IDF representation caused it to struggle with many dimensions.

In Figure 9, ROC Curves for different models are shown.

Binary cross-entropy, mean squared error, accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) were used to measure different models in binary classification.

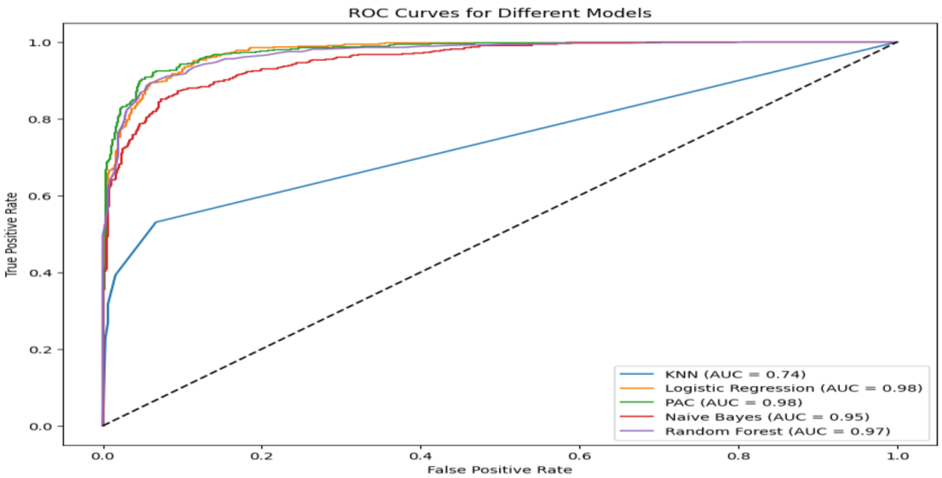


Figure 9. ROC Curve of Models

When it comes to all of these models:

- The AUC score of 0.98 for the Passive-Aggressive Classifier shows that it classified signals well.
- AUC shows that Logistic Regression's and Random Forest's results were very similar.

Moreover, KNN was far behind the other two, continuing to show it is not the best option for classifying text in many features. When seeking to perform even better, using deep learning models such as RNNs or Transformers, such as BERT, may help because of their success in natural language processing. Also, adding publication date, details about the author, and trustworthiness of the source could offer more information for classifying content and lower the risk of spreading misinformation.

Chapter-5 Discussions and Conclusion

The five models evaluated in this study exhibited varied performances, reflecting their inherent strengths and limitations when applied to text classification tasks.

The KNN algorithm achieved the lowest accuracy (65.19%) among all models. Its performance was constrained by the curse of dimensionality, a common challenge in high-dimensional feature spaces like the one generated by TF-IDF. While KNN is straightforward and intuitive, its reliance on distance metrics such as Euclidean distance diminishes its effectiveness as dimensionality increases.

Logistic Regression demonstrated strong performance, with an accuracy of 91.08%. Its linear nature makes it well-suited for tasks where the data is separable in a high-dimensional space. Furthermore, its ability to model probabilities provides interpretable outputs, making it a reliable baseline for binary classification tasks.

The Passive-Aggressive Classifier outperformed all other models, achieving an accuracy of 92.26%. This result aligns with its design as an online learning algorithm optimized for text classification tasks. The model's ability to adapt its parameters only during misclassifications reduces computational overhead and ensures efficient learning. This adaptability makes it particularly useful for real-time fake news detection, where new data is continuously encountered.

With an accuracy of 88.63%, Multinomial Naive Bayes proved to be a competitive baseline. Its assumption of feature independence simplifies the computation, making it efficient even for high-dimensional data.

3.5 Random Forest Classifier

The Random Forest Classifier achieved an accuracy of 91%, showcasing its robustness and ability to handle complex interactions between features. By aggregating the predictions of multiple decision trees, Random Forest mitigates overfitting and ensures stable performance.

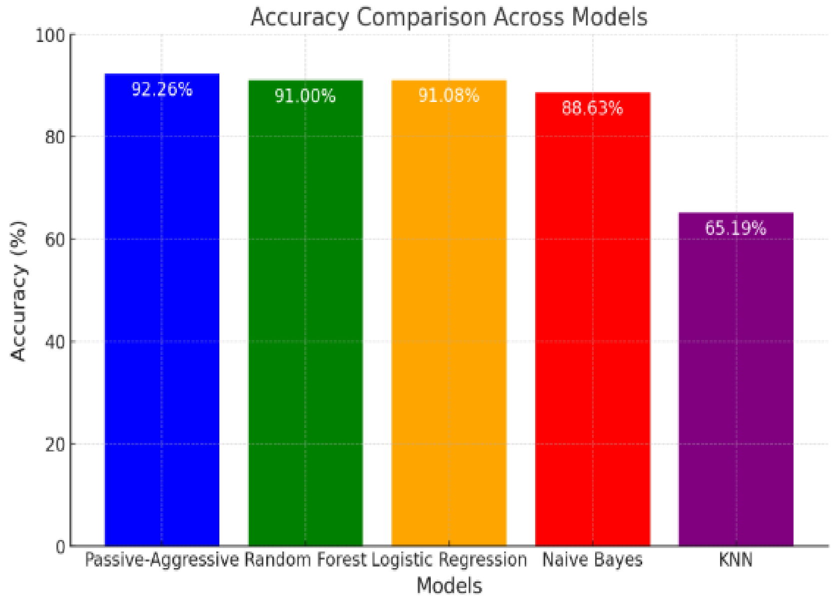


Figure 1: Accuracy for Top Models

Precision, Recall, and F1-Score Analysis

Beyond accuracy, precision, recall, and F1-score provide deeper insights into each model's performance. High precision indicates that the model makes fewer false-positive predictions, while high recall ensures that it correctly identifies most instances of fake news. The F1-score balances these two metrics, offering a single measure of overall effectiveness.

The Passive-Aggressive Classifier and Logistic Regression both exhibited high precision and recall, indicating their reliability in detecting fake news without significant false alarms. In contrast, KNN's low recall highlighted its inability to identify fake news consistently, further reinforcing its limitations in this domain.

The integration of context-aware models represents a promising direction for future research. Techniques like transformer-based architectures (e.g., BERT, GPT) offer the potential to capture deeper semantic relationships and context, enabling more accurate classification. However, these methods require substantial computational resources, posing challenges for real-time applications.

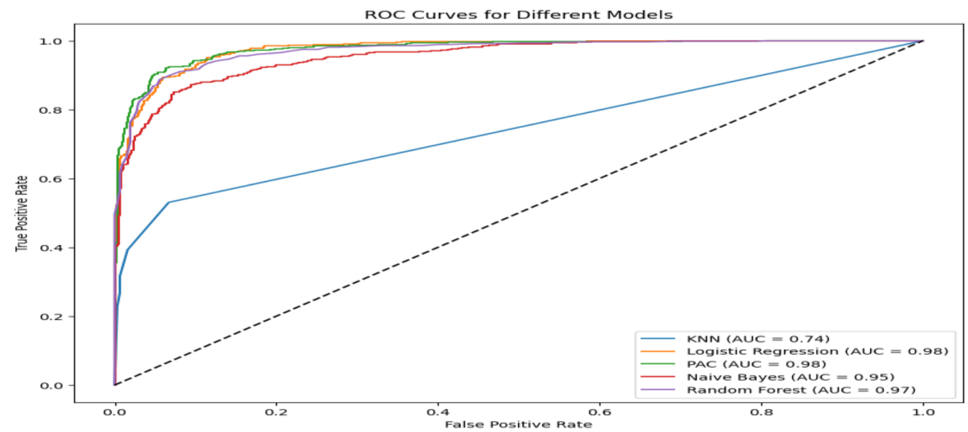


Figure 2: ROC Curves for Top Models

Limitations of the Project

The dataset utilized in this study consists of 6,335 samples which, although ample for a proof of concept study, might not be a correct representation of actual data. While the dataset as a whole is highly useful for categorizing news into REAL or FAKE, three important cases are left unaddressed: When a news article is somewhat true and somewhat fake.

Text based feature dependence of the model hinders the analysis of fake news containing multimedia information, for example photographic or video content, which are often popular in modern misinfo messages. The TF-IDF vectorization works well but does not consider the context of the words like Word2Vec or GloVe, transformers deal with that.

Future Scope

Sustained progresses that has been made in this project provides a framework for additional improvements. Several areas of development could expand the framework's applicability and robustness:

This integration with Multimedia Analysis has not been showcased in the current version of Illume. The future work might extend the suggested method by including visual and auditory features: analyzing images and/or videos available in news articles. The temporal analysis could be supported by other methods such as CNNs in order to classify the images or even models like CLIP.

Conclusion

It successfully deployed and evaluated multiple machine learning approaches identifiable and preventable fake news. Such outcome re-emphasizes the stages of preprocessing, features extraction and selection of the best models that brought good accuracy. The final classification model is the Passive-Aggressive Classifier which was found to be very efficient and ideal for the binary classifiers with limited samples as discovered in the research study.

But at the same time the study also pointed out the necessity for the more effective and powerful methods to counter the constantly developing threats of misinformation. The integration of multimedia content, the possibility to detect new multimedia content in real time, and the usage of better language models may extend the usefulness of the system

REFERENCES

[1] A. Gelfert, "Fake news: A definition," Informal Logic, vol. 38, no. 1, pp. 84-117, 2018. Available: https://doi.org/10.22329/il.v38i1.5068

[2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211-236, 2017. Available: https://doi.org/10.1257/jep.31.2.211

[3] Z. Farajtabar et al., "Fake news mitigation via point process-based intervention," in Proceedings of the 34th International Conference on Machine Learning, vol. 70, 2017, pp. 1073-1082. Available: http://proceedings.mlr.press/v70/farajtabar17a.html

[4] T. M. Mitchell, Machine Learning. McGraw-Hill, 1997.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22-36, 2017. Available: https://doi.org/10.1145/3137597.3137600

[6] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241-259, 1992. Available: https://doi.org/10.1016/S0893-6080(05)80023-1

[7] D. M. J. Lazer et al., "The science of fake news," Science, vol. 359, no. 6380, pp. 1094-1096, 2018. Available: https://doi.org/10.1125/science.aao2998

[8] N. Naudé, J. Kaunda, and B. von Solms, "Fake news on social media: Towards a conceptual framework for research and practice," in Proceedings of the 17th European Conference on Cyber Warfare and Security, 2018, pp. 259-277.

[9] Y. Wang, "The rise of fake news and the third-person effect," American Behavioral Scientist, vol. 65, no. 2, pp. 246-257, 2021. Available: https://doi.org/10.1177/0002764219878221

[10] E. M. Rogers, Diffusion of Innovations, 5th ed. Free Press, 2003.

[11] I. Ajzen, "The theory of planned behavior," Organizational Behavior and Human Decision Processes, vol. 50, no. 2, pp. 179-211, 1991. Available: https://doi.org/10.1016/0749-5978(91)90020-T

[12] S. A. Sloman, "Two systems of reasoning," in Heuristics and Biases: The Psychology of Intuitive Judgment, T. Gilovich, D. Griffin, and D. Kahneman, Eds. Cambridge University Press, 2002, pp. 379-396.

[13] B. Ghanem, P. Rosso, and F. Rangel, "An emotional analysis of fake news: Emotions in fake news versus real news," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2019, pp. 339-345. Available: https://doi.org/10.18653/v1/P19-2047

[14] N. Monti, M. Moretto, and D. Stein, "Fake news detection on social media using graph convolutional networks," in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2021, pp. 241-248. Available: https://doi.org/10.1145/3487351.3488259

[15] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Computing Surveys, vol. 53, no. 5, article 109, 2020. Available: https://doi.org/10.1145/3395046

[16] J. Pérez-Rosas, M. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 3391-3401.

[17] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on Sina Weibo," in Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, 2012, article 13. Available: https://doi.org/10.1145/2350190.2350203

[18] Y. Zhang, X. Ma, L. Xu, and C. Liu, "Fake news detection on social media using machine learning techniques," in Proceedings of the 2020 IEEE International Conference on Big Data, 2020, pp. 1656-1665. Available: https://doi.org/10.1109/BigData50022.2020.9378125

[19] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and dynamic information for studying fake news on social media," Big Data, vol. 8, no. 3, pp. 171-188, 2020. Available: https://doi.org/10.1089/big.2020.0062

[20] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," in Proceedings of the 12th Language Resources and Evaluation Conference, 2020, pp. 6086-6093.

[21] P. Bourgonje, J. M. Schneider, and G. Nischt, "Fake news detection using machine learning," in Proceedings of the 2nd International Conference on Digital Society and eGovernments, 2017, pp. 60-65.

[22] J. Albalawi, T. H. Yeap, and M. Benyoucef, "Using machine learning algorithms for fake news classification: A systematic review," in Proceedings of the 2020 IEEE International Conference on Big Data, 2020, pp. 5546-5555. Available: https://doi.org/10.1109/BigData50022.2020.9378230

[23] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146-1151, 2018. Available: https://doi.org/10.1125/science.aap9559

[24] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on Twitter during the 2016 U.S. presidential election," Science, vol. 363, no. 6425, pp. 374-378, 2019. Available: https://doi.org/10.1125/science.aau2706

[25] A. Agarwal, R. Xie, I. Vukovic, and C. Zhai, "Modeling the evolution of fake news: A comprehensive framework for detecting and classifying misinformation," in Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2020, pp. 92-99. Available: https://doi.org/10.1109/ASONAM49781.2020.9381461

[26] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The COVID-19 social media infodemic," Scientific Reports, vol. 10, article 16598, 2020. Available: https://doi.org/10.1038/s41598-020-73510-5

[27] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 795-816. Available: https://doi.org/10.1145/3123266.3123454

[28] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, "Stop clickbait: Detecting and preventing clickbaits in online news media," in Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2016, pp. 9-16. Available: https://doi.org/10.1109/ASONAM.2016.7752207

[29] S. Patwa et al., "Fighting an infodemic: COVID-19 fake news dataset," in Proceedings of the Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation, 2021, pp. 21-29. Available: https://doi.org/10.18653/v1/2021.comunica-1.3

[30] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsTracker: A tool for fake news collection, detection, and visualization," Computational and Mathematical Organization Theory, vol. 25, no. 1, pp. 60-71, 2019. Available: https://doi.org/10.1007/s10588-018-09280-3

[31] R. Samghabadi, S. Patwa, A. Bhardwaj, D. Guptha, P. Nandi, and T. Solorio, "Aggression and misogyny detection using BERT: A multi-task approach," in Proceedings of the Third Workshop on Abusive Language Online, 2019, pp. 34-39. Available: https://doi.org/10.18653/v1/W19-3505

[32] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in Proceedings of the 20th International Conference on World Wide Web, 2011, pp. 675-684. Available: https://doi.org/10.1145/1963405.1963500

[33] T. Mitra and E. Gilbert, "CREDBANK: A large-scale social media corpus with associated credibility annotations," in Proceedings of the Ninth International AAAI Conference on Web and Social Media, 2015, pp. 258-267.

[34] P. Patwa et al., "Overview of CONSTRAINT 2021 shared tasks: Detecting English and Hindi text against various communication hostile content," in Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation, 2021, pp. 1-13. Available: https://doi.org/10.18653/v1/2021.comunica-1.1

[35] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," ACM Computing Surveys, vol. 51, no. 2, article 32, 2018. Available: https://doi.org/10.1145/3161603

[36] Y. Kim, Y. Ryu, and K. H. Lee, "Detecting fake news in social media using graph neural networks for early warning," Expert Systems with Applications, vol. 185, article 115584, 2021. Available: https://doi.org/10.1016/j.eswa.2021.115584

[37] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand, "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention," Psychological Science, vol. 31, no. 7, pp. 770-780, 2020. Available: https://doi.org/10.1177/0956797620939054

[38] R. K. Hossain et al., "Towards building a reliable and robust fake news detection system," in Proceedings of the 2021 IEEE International Conference on Big Data, 2021, pp. 4747-4756. Available: https://doi.org/10.1109/BigData52589.2021.9671709

[39] H. Cui and V. Mittal, "Deep neural network for Twitter sentiment analysis," in Proceedings of the Workshop on Twitter and Social Media Data Analytics, 2019, pp. 51-58. Available: https://doi.org/10.1145/3343031.3343036

[40] M. Pradhan, A. Bhattacharya, and G. Sanyal, "Fake news detection using machine learning approaches: A systematic review," SN Computer Science, vol. 2, article 355, 2021. Available: https://doi.org/10.1007/s42979-021-00745-1s

[41] Agudelo, G.E.R., Parra, O.J.S. and Velandia, J.B., 2018, October. Raising a model for fake new detection using machine learning in Python. In Conference on e-Business, e-Services and e-Society (pp. 596-604). Springer, Cham.

[42] F. Sanei, A. Harifi and S. Golzari, "Improving the precision of KNN classifier using nonlinear weighting method based on the spline interpolation," 2017 7th International Conference on Computer and Knowledge Engineering (ICCKE), 2017, pp. 289-292.

[43] P. Nair and I. Kashyap, "Hybrid Pre-processing Technique for Handling Imbalanced Data and Detecting Outliers for KNN Classifier," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 460-464.

[44] Kesarwani, A., Chauhan, S.S., Nair, A.R. and Verma, G., 2021. Supervised Machine Learning Algorithms for Fake News Detection. In Advances in Communication and Computational Technology (pp. 767-778). Springer, Singapore.

[45] A. Kesarwani, S. S. Chauhan and A. R. Nair, "Fake News Detection on Social Media using K- Nearest Neighbor Classifier," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1-4,10.1109/ICACCE49060.2020.9154997.

[46] Nagashri, K. and Sangeetha, J., 2021. Fake News Detection Using Passive-Aggressive Classifierand Other Machine Learning Algorithms. In Advances in Computing and Network Communications (pp. 221-233). Springer, Singapore.

[47] Vijayaraghavan, S., Wang, Y., Guo, Z., Voong, J., Xu, W., Nasseri, A., Cai, J., Li, L., Vuong, K. and Wadhwa, E., 2020. Fake news detection with different models. arXiv preprint arXiv:2003.04978.

[48] Tschiatschek, S., Singla, A., Gomez-Rodriguez, M., Merchant, A., & Krause, A. (2018). "Fake News Detection in Social Networks via Crowd Signals." In Proceedings of the 2018 Web Conference (pp. 517-524).

[49] Shu, K., Wang, S., & Liu, H. (2017). "Exploiting Tri-Relationship for Fake News Detection." In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 995-1004).

[50] Wu, Y., Liu, B., Zhang, Y., Yao, Y., & Zhang, A. (2018). "False Information Detection on Social Media with Graph Convolutional Networks." In Proceedings of the 2018 Conference on Information and Knowledge Management (pp. 1733-1742).

[51] Ahmed, H., Traore, I., & Saad, S. (2017). "Detecting Opinion Spams and Fake News Using Text Classification." Security and Privacy, vol. 1, no. 1, pp. 1-15.

[52]Combatting-fake-news-analysis-of-submissions-to-the-fake-news https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.meccsa.org.uk%2Fnl%2Fthree-d-issue-28-combatting-fake-news-analysis-of-submissions-to-the-fake-news-inquiry%2F&psig=AOvVaw3M0sIJumkshH1Heplaj5GH&ust=1723446470559000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCICIjNaw7IcDFQAAAAAdAAAAABAE