Machines and Algorithms
http://www.knovell.org/mna

Research Article

Fake News Detection using NLP and ML Techniques

Muhammad Irfan[1, *], Drakhshan Bokhat[1], and Rabia Bajwa[1]

[1]Computer Engineering Department, University of Engineering and Technology, Lahore, 54890, Pakistan
[*]Corresponding Author: Muhammad Irfan. Email: engr.mirfan32@gmail.com
Received: 06 May 2023; Revised: 23 May 2024; Accepted: 08 July 2024; Published: 1 August 2024
AID: 003-02-000040

Abstract: Effortless access to the bulk of information and its spread is becoming a major issue these days. Many people rely on social networking sites and electronic media to get information. These sites might be used to spread information more quickly, especially fake information. This widespread has catastrophic results on individual and society. Identifying the fake news over numerous mass media platforms have rendered traditional machine learning algorithms less effective. Since fake news detection is vital, this study aims at analysing common machine learning algorithms- linear regressor, decision trees, random forest and Naïve Bayes, MLPC and LSTM and the ensemble methods- XGBoost and CatBoost, particularly to validate the efficiency of Kaggle dataset on fake news (fake-and-real-news-dataset). The results reveal a surprising lead of ensemble methods and LSTM over other algorithms in this dataset, particularly the XG Boost Classifier achieved the highest accuracy of 92% in validation and 100% in training. Similarly, Recall score of CatBoost is higher than others.

Keywords: Fake News; Text Processing; Machine Learning;

1. Introduction

Exponential growth and informal access to the information available has made it complicated to differentiate between incorrect and correct news. The prevalent diffusion of data has ironically promoted an exponential increase in its deformation. [1]. Endorsing data validity in modern times is becoming an issue of raising concern. The propagation of any false information in society has determinantal effects on a community. Nowadays, search engines and social media have become the primary sources of news for many people, surpassing print media. It can be said undoubtedly that publishers have no control on information flow due to bulk of information and news accesses through electronic media. Social media possesses unmatched reach and influence, enabling it to impact billions of people worldwide. Fake data entail verifiable, misleading, and purposefully false information. There is a severe need of a system to ensure that the information exchanged is genuine or spreader is honest. With no active monitoring the news spread can be molded by the people or community for their own needs and can cause detrimental effects on a community [2]. Misinformation often spreads more rapidly than factual news, as it is often tailored to be more engaging and attractive to audiences. Different types of fake news include fabricated political stories, misleading ads, sensationalized rumors, and mock or satirical articles. Google publicized an innovative facility named "Google News Initiative" intended to track and eliminate misinformation. During the 2016 US presidential election, numerous articles circulating online contained false information and misleading data, which spread rapidly and gained widespread attention. As a result, distrust took hold,

increasing partial reviews and backing up biased public perception. Such misleading information poses biased opinion that creates hoax among people mind therefore controlling the decision e.g., the political election [3]. As a result of such serious concern about social and national damage, considerable research attention is already being devoted to this field. Advances in natural language processing and machine learning have empowered the creation of complicated models capable of classifying between true and false news stories [4].

Developing a fake news detection model is challenging. The goal of these models is to classify news articles as fake or real by outlining insights gained from a comprehensive review of previously verified fake and real news stories. The readiness of large news datasets is decisive, and news articles can be sourced from various online platforms, including search engines and social media sites. Still, defining the news authenticity can be stimulating and time-consuming task if done by hand. On the other hand, fake news detection can be viewed as a binary classification problem or a fine-grained classification task. Following 2017, researchers have made rigorous efforts to improve model performance using assorted widely available datasets. Although no single benchmark dataset has been widely adopted, several prominent datasets are commonly used are CREDBANK, BuzzFeedNews, LIAR, PHEME, ISOT, PolitiFact, BS Detector and Kaggle.[5].

The substantial amount of heterogeneous content is available and advancement of natural language processing and machine learning techniques we can improve, pace up, and standardise the methodical process of such problems. Unstructured text data holds valuable insights, which can be uncovered through the synergistic application of NLP and ML. This paper demonstrates the efficiency of the combined approach in classifying fake news, using a labeled dataset from Kaggle. Through a comprehensive evaluation of existing techniques, we developed a better model to improve classification performance. Our methodology necessitated calculating TF-IDF features prior to model training, which was ensured using a variety of machine learning techniques. The model achieved the highest accuracy score achieved is 92% when keras vectorizer is used on data set. All the model results were above 95% with simple TF-IDF vectorization.

Section II explains the background knowledge and survey of the relevant content. Proposed model is given in Section III which discourses the approach to detect the fake news from the real ones using basic NLP processing on dataset and applying LSTM to tarin and test the model. Section IV concludes the discussion, and references are provided in Section V.

2. Literature Review

Reference [2] provides an exhaustive examination of the key characteristics involved in detecting fake news, along with a complete analysis of various machine learning techniques and their applications among different news categories. Comparative studies on fake news detection approaches and datasets showed that content type, whether textual or image-based, impacts the choice of method, eventually leading to better problem-solving and improved classification performance. The work on the image type news is still an open challenge as the reliability and impact of the spread depends a lot on the content type i.e., text or image. News articles frequently feature multimedia components, such as images and videos, that are not inevitably connected to the content, but are designed to tempt readers with clickbait tactics.

The paper [5] presents analysis of four machine learning techniques based on accuracy score and convergence time. The Naïve Bayes, the neural network, the random forest, and the decision trees algorithms are used for fake news detection. The dataset was trained and analyzed to determine which algorithm executes well. Naïve Bayes algorithm performed good on textual data. The algorithm utilizes Bayes' theorem to calculate conditional probabilities, enabling it to classify between events or classes based on individual text occurrences. Notably, this approach yields better results compared to other algorithms when used on this dataset. The accuracy is decent, and the convergence/training time of the Naïve Bayes is 1.3s which is best as compared to the other algorithms: Random Forest 54.3, neural network 420.2 and Decision tree 520s.

The study [6] developed and evaluated various fake news classification models using machine learning, deep learning, and natural language processing techniques. The scholars compared the performance of four traditional algorithms (binomial logistic regression, naive Bayes classifier, support vector machines, and random forest) and three neural network models (CNN with GlobalMaxpool, CNN with DepNetwork, and LSTM). They also examined the impact of NLP methods such as TFIDF, Word2Vec, and Tokenization on model performance. The results steered that random forest achieved the highest accuracy among traditional methods, while CNN with GlobalMaxpool went best among neural network models. The use of NLP methods significantly improved model accuracy, with traditional methods achieving over 85% accuracy and neural network models achieving over 90% accuracy.

In [7], authors concluded that bots' usage can be perceived as facilitators of information broadcast, either with bad intention or the good ones. They don't benefit from any type of entry; however computational capabilities of the bots are better at disseminating data at faster rate as compared to human beings. In context of its popularity, easy manufacturing, and due to simple usage, they are adopted by many users. There are many ways to improve information authentication but, it requires a great deal of preprocessing of related elements and topologic assessment of items. The reviewed state-of-the-art automatic detection models are composites of natural language approaches and machine learning techniques. According to the authors, natural language processing methods employed in literature serve as a foundational step rather than a complete solution but are commonly incorporated into broader machine learning solutions. This systematic literature review recommends that a hybrid approach combining traditional techniques, orchestrated by a neural network, is the most operative method. They also insisted on unification of different terminologies and definition of fake news domain in lieu of uniform domain ontology. The dearth of consensual information could mislead suppositions and judgments.

The study in [8] explores fake news detection through a two-step approach: classification and discovery. The classification step outlines the fundamental concepts and principles of fake news, while the discovery step reviews existing methods for detecting fake news using supervised learning algorithms. The results show that Naïve Bayes achieves accuracy rates ranging from 74% to 96.8%, while SVM and neural networks reach accuracy scores of up to 99% on Kaggle and various datasets. A study on 2012 Dutch elections fake news on Twitter employed 8 supervised machine learning classifiers. The decision tree algorithm yielded the best results, achieving an impressive F score of 88%. Additionally, some detection models utilizing N-gram analysis attained the highest accuracy score of 92%. Most research papers have utilized the Naïve Bayes algorithm, achieving prediction precision between 70-76%. While these studies primarily employed qualitative analysis, focusing on sentiment analysis, titles, and word frequency repetition, the authors suggest that incorporating quantitative approaches, such as POS textual analysis, could enhance feature extraction and improve precision results using random forest algorithms. The proposed feature set includes a range of linguistic and stylistic metrics, including Lexical features: total words, total unique words, and Type/Token Ratio, Syntactic features: number of sentences and average sentence length, Character-based features: number of characters and average word length, Semantic features: part of-speech tags (nouns, prepositions, adjectives, etc.)

Fake news term is used of any fabricated, doctored, or fake content shared via any medium. Authors of [9] state that unverified news has been escalating at a rapid rate which is alarming for society. It is becoming very difficult to filter the real news from the fake ones. They proposed a solution to detect the fake news automatically through linguistic analysis and machine learning approaches.

Jain et. al., mentioned in [10] that there must be a sophisticated system for detecting fake news as most of the persons use their mobiles to read the news on social media platforms and same medium is playing an important role in spread of wrong information within no time. This paper focuses on detection models that make use of natural language processing and machine learning techniques. like Naïve Bayes Classifier, SVM etc. and achieved state-of-the-art results.

In [11] the discussion is focused on fake profiles on the social media platforms which are spreading wrong information on the internet, and most of the people believe that information blindly. Bots are employed which are generated via Machine Learning techniques and behave like real users. There must be

a model that can differentiate fake accounts and real accounts. A very little research exists on detecting the fake profiles created by the human being, especially on SMP's. The previously formed methods though feature engineering is not very useful to detect fake accounts. Every human being has unique thinking and different style to use the social media account. So, bots are always not successful, as they use specific patterns all the time.

The analysis in [12] enforces the development of intelligent systems that can be used to solve complex problems based on implicit knowledge. This survey emphasizes the use of knowledge engineering in development of fake news detection model. According to the authors existing solutions are helpful in overcoming the stated issue but data driven approaches and knowledge engineering can improve the efficiency of existing solutions.

This paper [13] conducts a thorough evaluation and comparison of various fake news detection methods, spanning traditional machine learning approaches, such as Naive Bayes, and popular deep learning approaches, including CNN and RNN. The study reviews a diverse range of data items, including visual, user, network, post, knowledge, style, and stance-based features. The evaluation is performed on three feature sets, and the results underscore the importance of feature selection in achieving optimal accuracy and performance.

While simple approaches yield promising results, the study highlights the potential for complex models to significantly improve accuracy. Reference [14] used probabilistic latent semantic analysis to detect fake news. The study also exhibited a comprehensive comparative analysis of existing literature and estimated various machine learning and deep learning approaches on three datasets. The comparison revealed that deep learning techniques go beyond traditional machine learning techniques in performance, with Bi-LSTM achieving an accuracy score of 95%. The analysis primarily focused on textual data but can be extended to image data and heterogeneous datasets.

The primary objective of the research presented in [15] is to develop a deep learning-based style for detecting fake news. Given the disturbing prophecy by Gartner that most people in mature economies will consume more false information than true information by 2022, automated fake news detection has become a pressing task. The proposed approach addresses the boundaries of existing binary classification models by introducing a neural network architecture, TF-IDF-DNN and BoW-DNN, which accurately predicts the stance between a given pair of headlines and article body, accomplishing an accuracy score of 94.21% on test data.

Reference [16] presents a comprehensive comparison of recent benchmark datasets and experimental results obtained using various methods. It presented fake news detection problem as challenging practical problem of NLP and discussed the existing NLP solutions and their weaknesses. They recommended the appropriate handling of non-textual data and demanded to check whether the handcrafted features can be employed with neural networks. Comparing multiple techniques on different data sets, LSTM based models achieved higher accuracy on LIAR data set as compared to CNN.

On FEVER dataset attention-LSTM has the best score. The experimentation centered on semantic matching of each sentence from reclaimed pages and the entitlement. Another dataset analyzed is Fakenewsnet. This dataset is collected from two different sources: Buzzfeed and Politifact. It mostly comprises of social engagements data from the articles of Twitter. Castillo. The maximum achieved accuracy score was given by GCN. This method takes advantage of graph-based data that encodes associations between news stories and their publishers, using this information as input for a CNN to evaluate the integrity of news articles.

Authors encourage extending this problem of binary classification to multiclass classification problem, where news must not be only categorized as strictly true or false, but the labels must include half true, mostly false etc. Concluding, the accuracy scores can be improved if meta-data including speaker credibility and social engagements' information is also considered.

3. Proposed Methodology

This section presents the experimental methodology, which includes dataset preparation, workflow design, and model training. Figure 2 provides a visual representation of the experimental process. The procedure includes three main stages: data preprocessing, preparation of machine learning and neural network models using traditional and neural network approaches, and a evaluation of the used techniques.

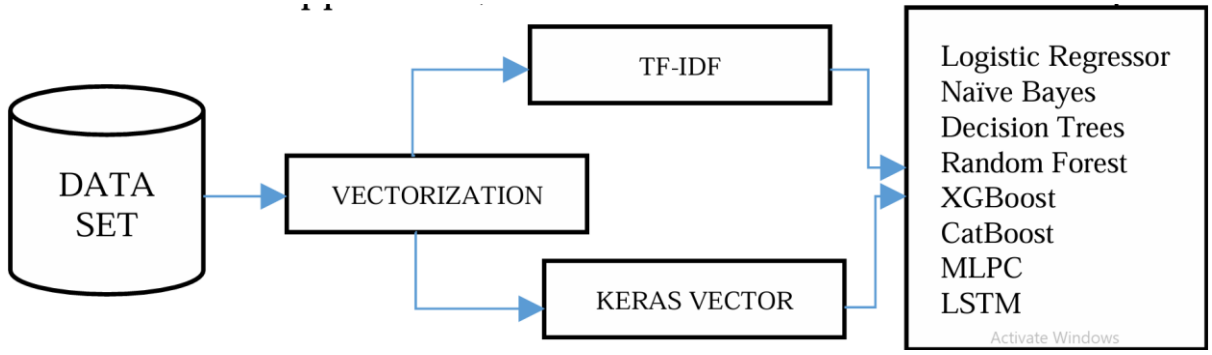

Figure 1: Workflow of Fake News Classification Using Feature Extraction Methods and Machine Learning Models

3.1. Data Set

The Kaggle dataset is used for model training. The dataset comprises 40,000 articles, evenly divided between fake and real news, with each category consisting of approximately 20,000 articles. This dataset consisted of two csv files: fake with 21417 samples and true news with 23481 news samples. Both records consisted of four features. The class label assigned to fake and real news as 0 and 1 respectively, and files are merged for preprocessing of data as a single unit.

3.2. Data Pre-Processing

We investigated to check the missing values in dataset as it affects the overall performance algorithms; no missing values were found. Then data is visualized on subject type for getting the insight. It is clear from the Figure 1 that the data set contains fake news of some categories and real news for other type of categories. The fake and real news are not categorized within the same subject.

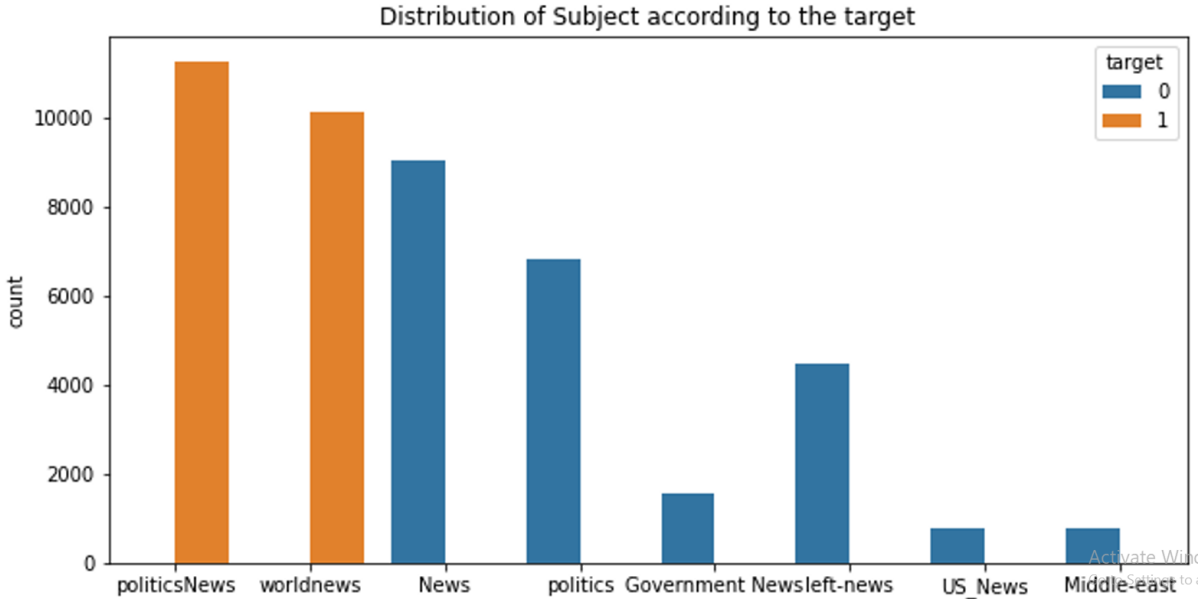

Figure 2: Distribution of True vs Fake News in the Different Subjects

In the next steps all the attributes are concatenated in one content text (news subject + news title + news text) and html content is removed. Next all the punctuation and special characters are removed. Non alphabetic characters are replaced with spaces and all the text is converted into lower case letters. After tokenizing the text into individual words, stop words are removed from the dataset. This is because stop words, such as common function words, are equally likely to appear in both true and fake news articles, representing them useless in distinguishing between the two. Following, we applied lemmatization is applied to bring back multiple form of words/tokens to their common root.

3.3. Text Analysis

After the data is ready, frequent words are looked at using the word cloud. For that all the tokenized text is converted into strings in separate columns because it is to be used later for model training and text analysis. From the word cloud it's clear that there is a detach in progressively diminishing rate of recurrence. The frequency is either elevated or muted in fake news as compared to real ones. iv. Vectorization of Processed data After the examination, the term frequency-inverse document frequency is computed using TFIDF-vectorizer and data is spitted into training and testing samples as 80-20. Another approach of vectorization is deployed in which keras tokenizer is used. This vectorizer enables the conversion of text into a numerical format, creating either a sequence of integers or a vector comprising binary coefficients that correspond to each token. After updating the internal vocabulary for manuscript, text corpus is converted into a sequence of integers and padded to ensure that all sequences in a list have the same length.

3.4. Model Training

The classification models used in this research were Logistic Regression, Random Forest, Naïve Bayes, Decision Trees, CatBoost, XGBClassifier, MLPC, and LSTM. Training these models involved tuning their parameters, resulting in varied outcomes. Subsequently, the trained models were tested on a separate dataset, with a train-test split ratio of 80:20.

4. Results

The scope of this study is to analyze and classify the Kaggle dataset, which confines labeled fake and real news articles. The data is first pre-processed using natural language techniques, and then various evaluation techniques are applied. The results of these evaluations, which involve six different algorithms, are presented in a confusion matrix. The development and testing of the model involved the use of eight machine learning algorithms: Logistic Regression, Naïve Bayes, Decision Trees, Random Forest, XGBoost, CatBoost, MLP, and LSTM. The data set is given in two different forms of vectors and simple TF_IDF vectorization shows better results as compared to Keras tokenizer. Only LSTM performed better in the case of keras tokenizer, where MLPC has shown the lowest accuracy score. The results are summarized in the table below.

Table 1: Results

Model	**Accuracy**		**Precision**		**Recall**	
Logistic Regression	60	99	59	98	51	99
Naives Bayes	58	94	56	94	59	93
Decision Tree	75	99	67	99	95	99
Random Forest	85	99	86	99	82	99
XG Boost Classifier	92	99	88	99	94	99
Cat Boost Classifier	90	100	85	100	97	100
MLPC	54	99	83	99	7	99
LSTM	80		76		8	

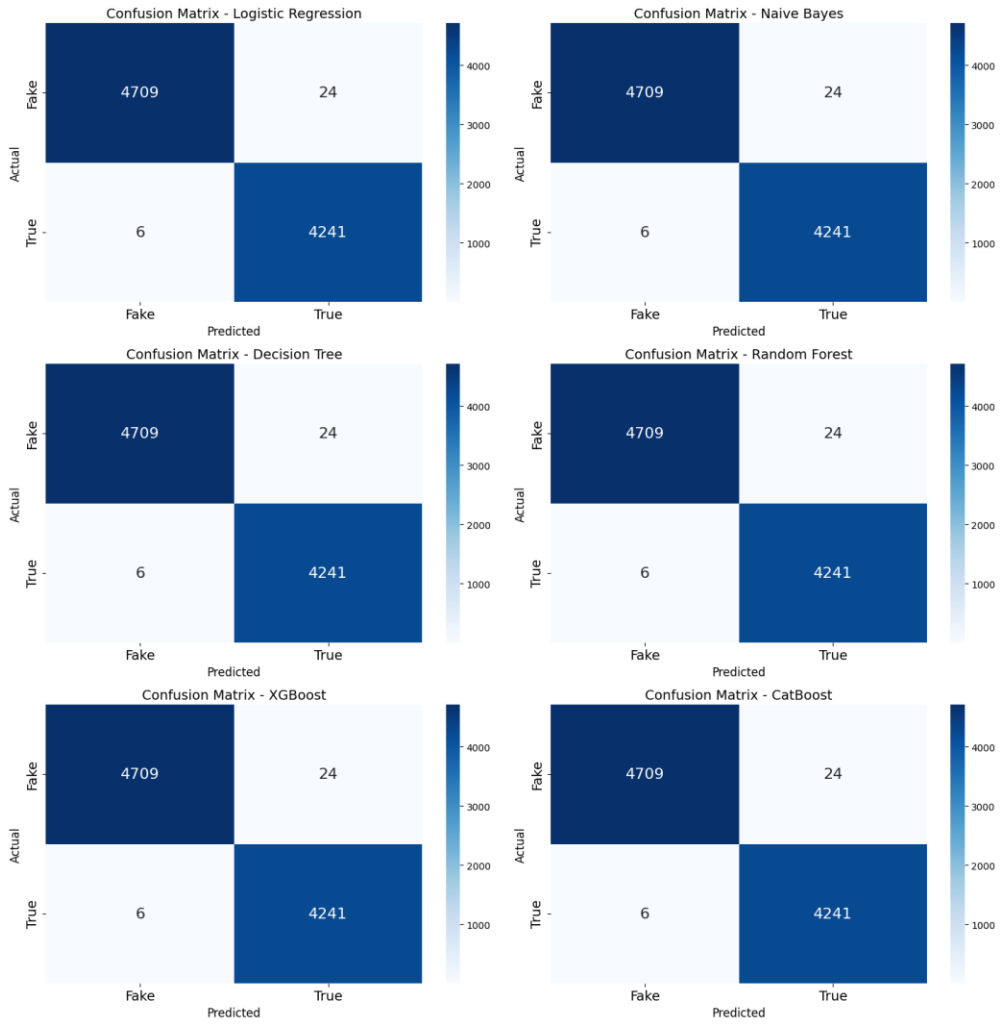

Figure 3: Confusion Matrices of Different Machine Learning Models for Fake News Classification

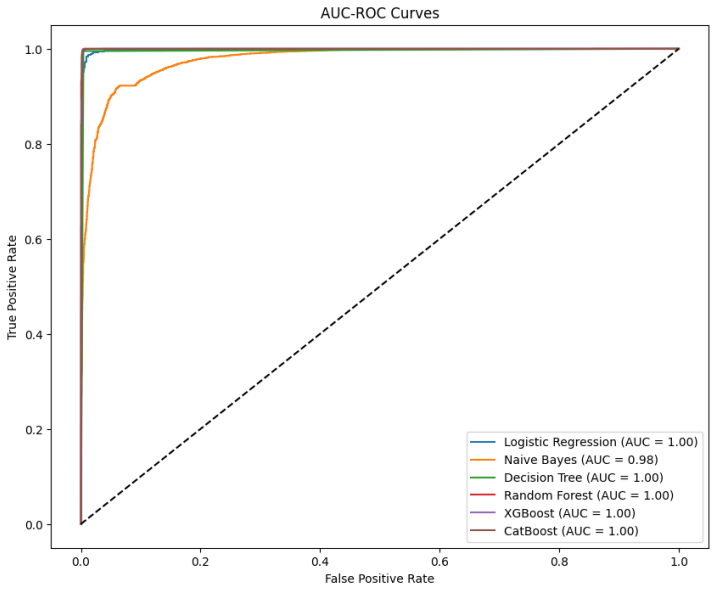

Figure 4: AUC-ROC Curves of Selected Machine Learning Models for Fake News Classification

The confusion matrices for selected classification models are presented in figure 3. Each confusion matrix evaluates the corresponding model's performance in distinguishing fake news (0) from true news (1). The figure shows that all models perform exceptionally well on the selected dataset, showing high accuracy. We also generate ROC for the selected models as shown in figure 4. Although the Naïve Bayes model performs slightly worse (AUC = 0.98) but most of the models show perfect discrimination (AUC = 1.00) True Positive Rate (TPR) and False Positive Rate (FPR). It is likely due to strong learning from the dataset. AUC values close to 1.00 indicate high accuracy, but models should be tested on new data to ensure generalization and avoid overfitting concerns.

5. Conclusion

This study points to develop an effective approach for detecting fake news, employing the combined strengths of NLP and ML techniques. In the first step the data textual data is preprocessed using NLP techniques. After cleaning the data, it is vectorized using keras vectorization technique and simple TFIDF computation. Results were better in the case of TF-IDF vectorization. The maximum accuracy score achieved is 92% using XGBoost Classifier while MLPC has the lowest accuracy score when applied to keras vectorized data set. On the TF-IDF feature vectorst, the maximum accuracy score is given by CatBoost classifier and all the classifiers have values above 95%. For future different feature vectors analysis and K-cross validation can be applied to achieve better results in case of the models that failed to achieve better accuracy.

References

[1] S. I. Manzoor, J. Singla and Nikita "Fake news detection using machine learning approaches: A systematic review." *2019 3rd international conference on trends in electronics and informatics (ICOEI)*. IEEE, 2019.

[2] M. Fire, D. Kagan, A. Elyashar, and Y. Elovici, “Friend or foe? fake profile identification in online social networks,” *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–23, 2014

[3] M. Jiang, P. Cui, and C. Faloutsos, “Suspicious behavior detection: Current trends and future directions,” *IEEE Intelligent Systems*, vol. 31, no. 1, pp. 31-39, 2016

[4] Z. Khanam, B. N. Alwasel, H. Sirafi, and Mamoon Rashid. "Fake news detection using machine learning approaches." In *IOP conference series: materials science and engineering*, vol. 1099, no. 1, p. 012040. IOP Publishing, 2021.

[5] A. Albahr and M. Albahar. "An empirical comparison of fake news detection using different machine learning algorithms." *International Journal of Advanced Computer Science and Applications* 11, no. 9, 2020.

[6] C. M. Lai, M. H. Chen, E. Kristiani, V. K. Verma and C. T. Yang. "Fake news classification based on content level features." *Applied Sciences* 12, no. 3: 1116, 2022.

[7] C. Shao, G. L. Ciampaglia, O. Varol, K. C. Yang, A. Flammini and F. Menczer. The spread of low-credibility content by social bots. *Nature communications*, *9*(1), p.4787, 2018.

[8] M. S. Looijenga. "The detection of fake messages using machine learning." Bachelor's thesis, University of Twente, 2018.

[9] V. Singh, R. Dasgupta, D. Sonagra, K. Raman, and I. Ghosh. "Automated fake news detection using linguistic analysis and machine learning." In *International conference on social computing, behavioral-cultural modeling, & prediction and behavior representation in modeling and simulation (SBP-BRiMS)*, pp. 1-3. 2017.

[10] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, "A smart system for fake news detection using machine learning." In *2019 International conference on issues and challenges in intelligent computing techniques (ICICT)*, vol. 1, pp. 1-4. IEEE, 2019.

[11] N. Singh, T. Sharma, A. Thakral, and T. Choudhury. "Detection of fake profile in online social networks using machine learning." In *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, pp. 231-234. IEEE, 2018.

[12] M. R. Islam, S. Liu, X. Wang and G. Xu. "Deep learning for misinformation detection on online social networks: a survey and new perspectives." *Social Network Analysis and Mining* 10, no. 1: 82, 2020.

[13] Y. Liu, and Y. F. Wu. "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks." In *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1. 2018.

[14] H. Padalko, V Chomko and D. Chumachenko Padalko,. A novel approach to fake news classification using LSTM-based deep learning models. *Frontiers in big Data*, *6*, p.1320800, 2024.
[15] A. Thota, P. Tilak, S. Ahluwalia and N. Lohia. Fake news detection: a deep learning approach. *SMU Data Science Review*, *1*(3), p.10, 2018.
[16] A. H. J. Almarashy, M. R. Feizi-Derakhshi and P. Salehpour. "Enhancing fake news detection by multi-feature classification." *IEEE Access* 11: 139601-139613, 2023.