



A comparative study of machine learning algorithms for fake news detection

Alaa Mohammed Jawish

Table of contents

Abstract	2
Introduction	3
Background	3
Problem statement	4
Research objective	5
Research scope	6
Literature review	7
Research Methodology	10
Limitations	11
Expected Outcome	11
Gantt chart	12
References	13

Abstract

The rapid growth of social media has led to increase in the spread of fake news, which can significantly influence public opinion and decision-making. As a result, detecting fake news has become an important research topic in the field of artificial intelligence. This study represents a comparative analysis of different machine learning and deep learning approaches used for fake news detection.

The research focuses on evaluating traditional machine learning models such as Logistic Regression, Support Vector Machine, and decision tree, as well as advanced deep learning and hybrid models. It also examines the impact of different word embedding techniques, including TF-IDF, Word2vec, and FastText, on model performance.

The findings indicate that deep learning and hybrid models generally achieve higher accuracy, while the traditional models offer better computational efficiency and faster execution time. In addition, the study highlights the limitations of current approaches, particularly those that rely on textual data and ignore other types of data.

This research provides a better understanding of the strengths and weaknesses of different approaches and offers useful insights for improving fake news detection systems and guiding future research in this field.

1.1 Introduction

Nowadays, social media has become an integral part of people's lives. It's a fertile ground for connecting people, creating and sharing information, and staying up to date on trending events. Millions of users rely on online sources for information, which makes it easier for spreading of false information to reach a wide audience in short time. Fake news can influence people's decision, create confusion, and even cause social problems. As a result, developing automated systems that can accurately distinguish between real and fake content is now a critical necessity. However, despite the progress in this field, there are still several limitations in existing models.

Therefore, this research aims to explore and analyze different approaches for fake news detection, compare their performance, and identify the key challenges and limitations. The goal is to provide a clear understanding of the current state of fake news detection and suggest how to improve existing models.

1.2 Background

Evaluation of fake news detection

Fake news is not a new phenomenon, but it's increased with spread of the internet and social media platforms. With the development of digital platforms, online fake news has become a major concern nowadays. Over time researchers started developing automated methods to detect fake news using machine learning techniques. AI techniques have been successfully used in recent years to solve different prediction and classification problems in large range of research fields.

Machine learning and deep learning approaches

Early efforts in automated detection relied on traditional machine learning algorithms like logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree and Naïve Bayes. These models are valued for their speed and low computational requirements. However, with development of fake news, researchers have moved to deep learning architectures such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), which can capture more complex patterns and improved performance in many cases. Additionally, pre-trained language models such as BERT have enhanced the ability to understand context and meaning in text data.

Challenges in fake news detection

Despite the progress in fake news detection techniques, several challenges still exist. One of the main issues is that many models rely on textual data and ignored other data type like images and videos.

Another challenge is the variation in datasets. Some models perform well on specific datasets but fail to generalize to others.

Therefore, there is a need for more comprehensive approaches that can handle different types of data and provide more reliable results.

1.3 Problem statement

In today's world various developments in the technology have led to nuance that “Data is the oil” of the 21st century. There has been a drastic shift in the source of News consumption from Print media to Social media (Shalini Pandey,2022). The increased prevalence of fake news is a worrying phenomenon that has the ability to influence individuals' decisions, opinions and that may lead to severe influence on both people and society (OUASSIL,2022). As the widespread presence of fake news on social media continues to challenge the trustworthiness of online information, it becomes increasingly important to develop effective measures to address this problem (EHTESHAM HASHMI,2024). Existing studies have proposed many approaches for fake news detection, including traditional Machine learning algorithms such as logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree and Naïve Bayes. As well as deep learning models like Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) model and hybrid model combined with Combination of Word Embedded Techniques like FastText and Word2vec. While traditional Machine learnings models offer the advantage of fast execution and lower computation cost, that make them suitable for real-time applications with limited resources. However, this models often fail to capture the complex patterns in text. On other hand, deep learning and hybrid models have achieve better accuracy especially when combined with word embedded techniques. However, these models need high computational resources and long training time. In addition, recent studies have shown that advanced pre-trained language model such as BERT and RoBERTa can achieve a higher performance compared to traditional machine learning models (Junaed Younus,2021). The significance of this problem is further underscored by the common limitation in existing literature, which relies primarily on textual features and not on other sources as images and videos, which limit over all effective. Additionally, there is a trade-off between accuracy and computational efficiency, making it difficult to determine the best approach.

So, this research aims to address this gap by conducting a comparative analysis of different Machine learning and deep learning models to identify the best approach.

1.4 Research objective

1-To compare different Machine learning and deep learning algorithms in order to find the most suitable method for fake news detection.

- To identify the strength and weaknesses of each approach.
- To evaluate the classification accuracy and precision of traditional model against hybrid deep learning models.

2- To investigate the impact of various word embedding techniques in accuracy.

- To evaluate the effect of word embedding like FastText, TF-IDF, and Word2vec.

3- To evaluate the computational efficiency and execution time.

- To compare the time taken for execution between traditional models and hybrid models.
- To analyze the trade-off between speed and accuracy.

4- To propose an optimized framework for effective fake news detection.

5- To identify the current limitation of the method and suggest possible improvements.

- To analyze the limitation of relying only on textual data without considering multimedia content such as images and videos.

1.5 Research scope

Analysis of existing approaches: This research focuses on analyzing different approaches used in fake news detection, including traditional machine learning algorithms and deep learning models, and transformer-based models (BERT and RoBERTa). The study focuses on evaluating the performance of these models based on accuracy, precision, and computational efficiency.

Comparison of models: the research includes comparative analysis between different models such as Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree, and deep models like CNN and BiLSTM.

Evaluation of word Embedding techniques: The study examines the impact of different word embedding techniques such as FastText and Word2vec on model performance, and how they effect on the fake news detection.

Computational efficiency analysis: this research also analyzes the computational efficiency of the models by comparing execution time and resource requirements. It aims to understand the trade-off between accuracy and speed.

Limitations of current approaches: The research highlights the limitations of existing models, especially those that rely on textual data and do not consider multimedia content such as images and videos. The result may vary depending on the dataset used and may not fully generalize to all real-world scenarios.

2 Literature review

Fake news detection has become an important issue in recent years due to the rapid spread of mis information across the internet and social media platforms. Many studies have focused on applying machine learning and deep learning techniques to address this problem.

Gaurav Singh and Kamalnathan Selvakumar (2023) presented a comparative study of hybrid machine learning approaches for fake news detection. The fundamental issue this paper addresses is detecting/filtering out misleading news from online websites. In this study they used two datasets (ISOT and Kaggle), after preprocessing the data they implemented multiple Machine Learning algorithms, including single learning classifiers and ensemble learners, as well as Natural Language Processing, to identify the best-performing algorithms capable of classifying the news articles as fake or authentic with the highest accuracy. As result they found that Ensemble classifier learners, i.e., XGBoost and AdaBoost, performed relatively better than individual weak learners.

In their 2021 study, Kadek Sastrawan, Bayupati, Dewa Made Sri Arsa they analyze several deep learning methods performances combined with the current state of the art word embeddings on benchmark fake news datasets. Each pre-trained word embeddings is combined with the respective deep learning methods, namely CNN, Bidirectional LSTM, and ResNet, to determine their performance in detecting fake news. Furthermore, they also empirically show that the performance of deep learning in detecting fake news can be boosted by selecting proper word embeddings. The test results show that data augmentation has a positive effect, especially in improving model performance consistency. Several deep learning methods were also evaluated in this study using four different datasets. Each deep learning method is combined with pre-trained word embeddings, namely Word2Vec, GloVe, and fastText. Based on the test results, Bidirectional LSTM outperformed CNN and ResNet on all test datasets.

In their 2022 article in Journal of physics Shalini Pandey, Sankeerthi Prabhakaran, Subba Reddy and Dinesh Acharya, the algorithms such as K-Nearest Neighbor, Support Vector Machine, Decision Tree, Naïve Bayes and Logistic regression Classifiers to identify the fake news from real ones in a given dataset and also have increased the efficiency of these algorithms by pre-processing the data to handle the imbalanced data more appropriately. By using Word2vec it is observed that processing of text for computation is time consuming. Apart from this it's easier to execute the classifiers with a good accuracy report. The experimental result show that classifiers have achieved an accuracy of 89.98% for KNN, 90.46% for Logistic Regression.

In their 2024 study Published in: IEEE Access, EHTESHAM HASHMI, SULE YILDIRIM YAYILGAN, MUHAMMAD MUDASSARYAMIN, SUBHAN ALI, AND MOHAMED ABOMHARA, they introduce a robust approach for fake news detection, utilizing three publicly available datasets. The study integrates FastText word embeddings with various Machine Learning and Deep Learning methods, further refining these algorithms with regularization and hyper parameter optimization to mitigate overfitting and promote model generalization. Notably, a hybrid model combining Convolutional Neural Networks and Long Short-Term Memory, enriched with FastText embeddings, surpassed other techniques in classification performance. explainable AI modeling was employed using Local Interpretable Model-Agnostic Explanations, and Latent Dirichlet Allocation to gain deeper insights into the model's decision-making process. The results show that a hybrid model combining CNN with LSTM layers, augmented with FastText embeddings, outperforms other models in accurately classifying news articles. The use of explainable AI through LIME and LDA has not only justified the transparency of the detection process but also provided valuable interpretative insights.

Mohamed-Amine OUASSIL, Bouchaib CHERRADI, Soufiane HAMIDA (2022) present a new deep learning method to detect fake news based on a combination of different word embedding techniques and a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) model. The study trained the classification model on the unbiased dataset. The first task of this approach was representing words to meaningful numerical vectors using combination of different word embedding techniques. The second task was training the proposed hybrid model based on CNN and BiLSTM architectures. The obtained results show an improvement in terms of accuracy and precision when compared to traditional machine learning algorithms and related work results. The simple concatenation of the different pre-trained embedding models increases the dimension of embedded vectors. Moreover, that leads to a high computational complexity.

In 2025 study published in the International Journal for Research in Applied Science & Engineering Technology, Ritwik Mallavarapu compares the performance of two algorithms (Logistic Regression and Naïve Bayes) in identifying fake news based on accuracy, precision, recall, F1-score, and the time taken to execute. the computational cost of each algorithm is compared to establish its usability for large-scale applications. The results indicate fundamental trade-offs between speed and precision, stressing the importance of optimal models. the Results show that Logistic Regression achieved higher accuracy. However, Naïve Bayes demonstrated faster execution time.

Junaed Younus Khana, Md. Tawkat Islam Khondakera, Sadia Afroz, Gias Uddin, Anindya Iqbal (2021), conducted a benchmark study to assess the performance of different applicable machine learning approaches on three

different datasets where they accumulated the largest and most diversified one. They explored a number of advanced pre-trained language models for fake news detection along with the traditional and deep learning ones and compared their performances from different aspects for the first time to the best of their knowledge. They find that BERT and similar pre-trained models perform the best for fake news detection, especially with very small dataset. Hence, these models are significantly better option for languages with limited electronic contents. Their findings highlighted that transformer-based models, particularly RoBERTa reach over 90% accuracy even with limited training data, making them ideal for resource-constrained environments.

Jumana Jouhar, Anju Pratap, Neharin Tijo, Meenakshi Mony (2024) explored machine learning models using the ISOT dataset. They focused on comparing different machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost. The results indicated that XGBoost achieved the best performance in terms of accuracy and F1-score, highlighting the importance of selecting appropriate model.

In their 2024 study Mutaz A. B. Al-Tarawneh, Omar Al-ir, Khaled S. Al-Maaitah, Hassan Kanj and Wael Hosny Fouad Al, systematically investigates the effectiveness of different word embedding techniques TF-IDF, Word2Vec, and FastText when applied to a variety of machine learning (ML) and deep learning (DL) models for fake news detection. Leveraging the Truth-seeker dataset, which includes a diverse set of labeled news articles and social media posts spanning over a decade, they evaluated the performance of classifiers such as Support Vector Machines (SVMs), Multilayer Perceptron's (MLPs), and Convolutional Neural Networks (CNNs). their analysis demonstrates that SVMs using TF-IDF embeddings and CNNs employing TF-IDF embeddings achieve the highest overall performance (99.03%) in terms of accuracy, precision, recall, and F1 score. This study emphasizes the importance of selecting appropriate embedding techniques based on the model architecture to maximize fake news detection performance.

Muhammad Nadeem, Parchamdar Abbas, Wei Zhang, Sumaira Rafique, and Sundas Iqbal (2024) propose a hybrid framework for fake news detection, combining Natural Language Processing (NLP) techniques and machine learning algorithms. Using Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction, and classifiers such as Logistic Regression (LR), Naïve Bayes (NB), and Support Vector Machines (SVM), the model integrates Maximum Likelihood Estimation (MLE) with Logistic Regression to achieve 95% accuracy and 92% precision on a Kaggle dataset. The results highlight the potential of combining statistical and NLP approaches to improve fake news detection accuracy.

3.1 Research Methodology

This research follows a comparative approach to analyze different machine learning and deep learning models used for fake news detection. The methodology is designed to evaluate the performance of these models and identify the most suitable approach based on different needs.

Phase1: Literature Review

In this phase involves an extensive literature review to identify and synthesize existing research on fake news detection. The focus is on machine learning, deep learning, and hybrid models, as well as word embedding techniques such as FastText and Word2vec. The purpose of this phase is to identify current trend, challenges, and research gaps in the field.

Data sources: academic dataset as IEEE Xplore, Google Scholar, and ScienceDirect.

Selection criteria: Recent studies focusing on fake news detection.

Analysis techniques: comparative analysis to identify strengths, weaknesses, and limitations of different approaches.

Phase2: Model selection and data preparation

In this phase different machine learning and deep learning models are selected for comparison. The data is prepared using preprocessing techniques such as cleaning and feature extracting. Text representation methods such as TF-IDF and word embeddings are applied to convert text into numerical format.

Phase 3: Model implementation and evaluation

In this phase the different machine learning models such as Logistic Regression, support Vector Machine, decision tree, and Nive Bayes are implemented. In addition, deep learning models such as CNN and BILSTM are applied. Hybrid approaches that combine multiple techniques are also considered.

The performance of each model is evaluated using metrics such as accuracy, precision, recall, and F1-score. In addition, execution time is measured to evaluate computational efficiency.

Phase 4: Comparative analysis

In this phase, the results of different models are compared to identify the strengths and weakness of each approach. The analysis focus on the trade-off between accuracy and computational efficiency.

Phase 5: optimization and improvement

Based on the results, possible improvements are suggested to enhance the performance of fake news detection models. This includes selecting the most suitable model and optimizing feature representation techniques.

Limitations

This research methodology acknowledges potential limitations, including the study is limited to text-based fake news detection and does not include other type of media. In addition, the result of the models depends on the datasets used, which effect the generalization of the findings. Finally, some advanced models such as deep learning and hybrid approaches require high computational resources and longer training time, which may not be suitable for all applications.

3.2 Expected Outcome

The research is expected to provide a clear comparison between different machine learning and deep learning models used for fake news detection. This comparison will help identify the most suitable approach based on accuracy, precision, and overall performance.

Performance of different approaches: it is expected that deep learning and hybrid models will achieve higher accuracy compared to traditional machine learning models. However, traditional models may perform better in term of execution time and computational efficiency.

Impact of feature representation: the study is expected to show that the performance of fake news detection models depends on the feature representation techniques used, such as TF-IDF and word embedding methods like FastText and Word2vec.

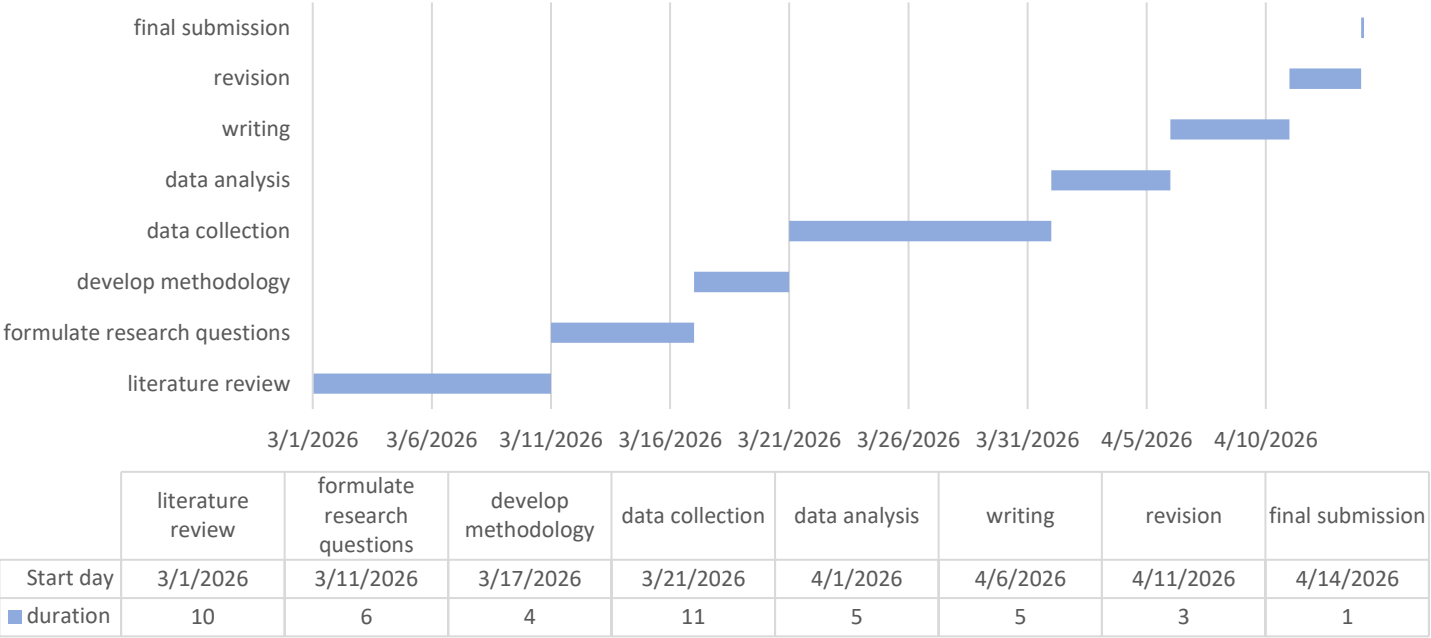
Identification of limitation: the research is expected to highlight the limitations of current approaches, especially those that rely only on textual data and do not consider other type of data.

Recommendations for improvement: the study is expected to provide useful recommendations to improve the fake news detection systems by selecting appropriate models and optimizing feature representation techniques.

Contribution to future research: this study is expected to contribute to future studies by providing a better understanding of the strengths and weaknesses of different approaches and guiding further improvements in the field.

3.3 Gantt chart

Gantt Chart



References

1. Singh, G.J., Selvakumar, K. (2023). *A Comparative Study of Hybrid Machine Learning Approaches for Fake News Detection that Combine Multi Stage Ensemble Learning and NLP-based framework*
<https://www.techrxiv.org/doi/full/10.36227/techrxiv.21856671.v3>
2. Sastrawan, K., Bayupati, I.P.A., and Arsa, D.M.S. (2021). *Detection of fake news using deep learning CNN–RNN based methods*
<https://www.sciencedirect.com/science/article/pii/S2405959521001375>
3. Pandey, S., Reddy, N.V.S., and Acharya, D. (2022). *Fake News Detection from Online media using Machine learning Classifiers.*
<https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012027/meta>
4. HASHMI, E., YAYILGAN, S.Y., MUDASSARYAMIN, M., SUBHAN ALI, AND ABOMHARA, M. (2024). *Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI.*
<https://ieeexplore.ieee.org/abstract/document/10477989>
5. OUASSIL, M.A., CHERRADI, B., HAMIDA, S. (2022). *A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model.*
https://www.researchgate.net/profile/Soufiane-Hamida/publication/365118023_A_Fake_News_Detection_System_based_on_Combination_of_Word_Embedded_Techniques_and_Hybrid_Deep_Learning_Model/links/6366d7cc431b1f530071d4c0/A-Fake-News-Detection-System-based-on-Combination-of-Word-Embedded-Techniques-and-Hybrid-Deep-Learning-Model.pdf

6. Mallavarapu, R. (2025). *Accuracy and Execution Time in Fake News Detection: A Comparative Analysis of Logistic Regression and Naïve Bayes* https://www.researchgate.net/profile/Ritwik-Mallavarapu/publication/389449473_Accuracy_and_Execution_Time_in_Fake_News_Detection_A_Comparative_Analysis_of_Logistic_Regression_and_Naive_Bayes/links/6853edd824267473b778e01d/Accuracy-and-Execution-Time-in-Fake-News-Detection-A-Comparative-Analysis-of-Logistic-Regression-and-Naive-Bayes.pdf
7. Khan, J.Y., Khondaker, M.T.Y., Afroz, S., Uddin, G., Iqbal, A. (2021). *A benchmark study of machine learning models for online fake news detection*. <https://www.sciencedirect.com/science/article/pii/S266682702100013X>
8. Al-Tarawneh, M.A.B., Al-ir, O., Al-Maaitah, KH.S., Kanj, H. and Aly, W.H.F. (2024). *Enhancing Fake News Detection with Word Embedding: A Machine Learning and Deep Learning Approach*. <https://www.mdpi.com/2073-431X/13/9/239>
9. Jouhar, J., Pratap, A., Tijo, N., Mony, M. (2024). *Fake News Detection using Python and Machine Learning*. <https://www.sciencedirect.com/science/article/pii/S1877050924006252>
10. Nadeem, M., Abbas, P., Zhang, W., Rafique, S., and Iqbal, S. (2024). *Enhancing Fake News Detection with a Hybrid NLP-Machine Learning Framework*. <https://www.icck.org/article/abs/tis.2024.461943>