

A comparative study of machine learning algorithms for fake news detection

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.

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.

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 technique in accuracy.
 - To evaluate the effect of word embedding like FastText 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 frame work for effective fake news detection.
- 5- To identify the current limitation of the method and suggest possible improvement.

Literature review:

Fake news detection has become an important issue in recent years because the rapid spread of misinformation across the internet and social media. Therefore many studies have focused on applying Machine learning and deep learning techniques to address this problem.

Machine learning approaches:

Traditional Machine learning algorithms have been widely used for detecting the fake news algorithms such as logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree and Naïve Bayes. They have shown a good performance especially when combined them with proper preprocessing techniques as highlighted by (J. Singh,2023) and (Pandey,2022). These models are fast and need lower computational cost, make them suitable for large applications and limited resource environment. However, they often may not fully capture the complexity of the textual data.

Deep learning and hybrid models:

Recent studies have explored deep learning models such as Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). (K Sastrawan,2021) show that these models have demonstrated improved performance compared to traditional methods. Ouaissil, (2022) highlight that combining different word embedding techniques can improve precision, though it increases computational complexity.

Explainable AI and model interpretation:

Recent trends have introduced explainable AI to the field. HASHMI (2024) utilized LIME and LDA to better understand model decisions and used to provide insights into how models classify news as fake or real.

Challenges and limitations:

There are several challenges in this field. Most of the current studies rely basically on text only, which limits their ability to handle other types of media. Also, there is a trade-off between accuracy and efficiency, as traditional models are faster but less accurate in complex cases. On other hand the deep learning shows better result but it requires higher computational resources.

Implications for future research:

The literature reviewed that no single approach is optimal for all applications. while the traditional Machine learning models offer efficiency, deep learning models and hybrid approaches provide better performance. future research needs to do a comparative study that comprehensively evaluates these methods to determine the most suitable approach depending on different requirements such as accuracy, speed and available resources.

References:

Gaurav J. Singh, Dr. 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>

Kadek Sastrawan, I.P.A. Bayupati*, Dewa Made Sri Arsa (2021) Detection of fake news using deep learning CNN–RNN based methods

<https://www.sciencedirect.com/science/article/pii/S2405959521001375>

Shalini Pandey, N V Subba Reddy and Dinesh Acharya (2022) Fake News Detection from Online media using Machine learning Classifiers

<https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012027/meta>

EHTESHAM HASHMI, SULE YILDIRIM YAYILGAN, MUHAMMAD MUDASSARYAMIN, SUBHAN ALI, AND MOHAMED ABOMHARA. (2024) Advancing Fake News Detection: Hybrid Deep Learning with FastText and Explainable AI <https://ieeexplore.ieee.org/abstract/document/10477989>

Mohamed-Amine OUASSIL, Bouchaib CHERRADI, Soufiane HAMIDA (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](https://www.researchgate.net/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)

Ritwik Mallavarapu (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](https://www.researchgate.net/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)