Fake News Detection Using Machine Learning and Deep Learning: A Comparative and Hybrid Model Evaluation

A PROJECT REPORT

Submitted by

Naishadha Badithala [CH.SC.U4AIE23036]

Kaushik Poshimreddy [CH.SC.U4AIE23042]

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE ENGINEERING AND ARTIFICIAL INTELLIGENCE

Under the guidance of

Dr. Bharathi Mohan

Submitted to

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING AND ARTIFICIAL INTELLIGENCE

AMRITA SCHOOL OF COMPUTING

AMRITA VISHWA VIDYAPEETHAM

CHENNAI - 601103

APRIL 2025

BONAFIDE CERTIFICATE

This is to certify that this project report entitled **"Fake News Detection Using Machine Learning and Deep Learning: A Comparative and Hybrid Model Evaluation"** is the bonafide work of **"Naishadha Badithala [CH.SC.U4AIE23036] and Kaushik Poshimreddy [CH.SC.U4AIE23042]"** who carried out the project work under my supervision as a part of End semester project for the course 22AIE213 - Machine Learning.

SIGNATURE

Name Signature

Dr. Bharathi Mohan

Associate Professor

Department of CSE-AIE

Amrita School of Computing,

Amrita Vishwa Vidyapeetham,

Chennai Campus

ACKNOWLEDGEMENT

This project work would not have been possible without the contribution of many people. It gives me immense pleasure to express my profound gratitude to our honorable Chancellor **Sri Mata Amritanandamayi Devi**, for her blessings and for being a source of inspiration. I am indebted to extend my gratitude to our Director, **Mr. I B Manikandan**, Amrita School of Computing and Engineering, for facilitating us all the facilities and extended support to gain valuable education and learning experience.

I register my special thanks to **Dr. V. Jayakumar**, Principal, Amrita School of Computing and Engineering for the support given to me in the successful conduct of this project. I would like to express my sincere gratitude to **Dr. Bharathi Mohan**, Associate Professor, Department of Computer Science Engineering and Artificial Intelligence for his support and co-operation.

I am grateful to Project Coordinator, Review Panel Members and the entire faculty of the Department of Computer Science & Engineering, for their constructive criticisms and valuable suggestions which have been a rich source to improve the quality of this work.

Naishadha Badithala [CH.SC.U4AIE23036]

Kaushik Poshimreddy [CH.SC.U4AIE23042]

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In today's digital world, where misinformation is spreading across online platforms, the detection of fake news takes on immense importance. This study discusses a comparative analysis of multiple machine learning (ML), deep learning (DL), and transformer models geared towards fake news detection. Our evaluation of different individual models consisted of several traditional ML classifiers (Random Forest, SVM, Voting Classifier), deep learning architectures (CNN, LSTM), and advanced transformer models (BERT, RoBERTa, XLNet, ALBERT). These models and their different combinations were evaluated in terms of accuracy, precision, recall, F1-score, and computational efficiency on benchmark datasets. Based on these results, we propose a hybrid ensemble model that inherits from these top-performing architectures. This hybrid approach enables the integrating of the feature extraction prowess of transformers with the interpretability of ML models and the sequential learning strength of DL networks to greatly boost detection performance and speed. Our results showed that the hybrid method outperforms the other individual models in accuracy and robustness. This study thus provides pointers towards the most viable methods for fighting misinformation and towards the development of automated systems for fake news detection.

Keywords: Machine Learning, Deep Learning, Transformers, Accuracy, Precision, Recall, F1-Score

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

Misinformation and fake news are an increasingly critical challenge in the digital age. They greatly affect public perception, the social-political equilibrium, and decision-making considerations. Fake news is engineered with the intent to deceive the audience, and since social media has grown exponentially, it has accelerated its very own impromptu dissemination. The impacts of misinformation can be ominous: health crises based on false medical claims and civil unrest spun from fabricated political narratives. Traditional fact-checking methods generally found to be manual solutions do not have a bearing on this problem due to size and speed of deployment of fake news. This therefore poses the most immediate need to design appropriate systems for automatically detecting fake news with high accuracy and scalability so that they can assist in the efficient automation of the detection of everything from honest to fraudulent content.

1.2 CHALLENGES AND RESEARCH MOTIVATION

Advancements in the field of machine learning (ML), deep learning (DL), and transformer-based natural language processing (NLP) have provided great improvements in the capability of fake news classification with high accuracy. A few challenges remain–one being that no single model can be right all the time because each model type has its pros and cons either in terms of accuracy, computation efficiency, interpretability, and generalization ability. For example; simple machine learning models give the highest speed of inference while also having some problems with modeling very difficult linguistic patterns; while deep learning architectures are excellent for sequential modeling but require vast amounts of labeled data. Above and beyond that, new benchmarks onto NLP tests have been established by transformer-based models such as BERT, RoBERTa, XLNet, and ALBERT but come at a very high computational cost and require much tuning to real data. These reasons justify the need for extensive comparative work to identify the best models that can be blended into a solid hybrid detection framework, using the strengths of both.

1.3 KEY CONTRIBUTIONS

To overcome these challenges, we propose an in-depth comparative study and hybridization of fake news detection, thereby contributing:

- Comprehensive Benchmarking: A comparative evaluation of a variety of models that include traditional ML classifiers (Random Forest, SVM, Voting Classifier), deep learning architectures (CNN, LSTM), and except for NLP models (BERT, RoBERTa, ALBERT, XLNet) on benchmark datasets. Every model is evaluated on accuracy, precision, recall, F1-score, processing time, and scalability, thus signaling its durability in contingency detection.

- Proposed Hybrid Ensemble Model: Based on the assessment of the best-performing models, a hybrid ensemble technique is proposed to combine the best performers of transformer, deep learning, and ML classifier approaches in a bid to achieve better detection performance. This hybrid model reported a new state-of-the-art accuracy of 99.23%, outperforming each model working in isolation, thus securing a robust generalization across datasets.

- Optimizing for Scalability and Speed: A new hybrid and a fast model are proposed with a low computational burden while still maintaining high detection accuracy, thus making it applicable to real-time scenarios that require quick classification.

- Implications for Further Studies: The paper provides in-depth insight into accuracy-efficiency-scales trade-offs, forming a useful guide toward future research and legitimate implementations of fake news detection systems.

CHAPTER 2

LITERATURE SURVEY

The task of detecting fake news has undergone considerable paradigm shifts with the advent of ML, DL, and transformer-based modeling. These models have made considerable pan against traditional rule-based approaches, and yet each comes with challenges that restrain it from being fully functional in a real-world setup. The following sections elaborate on respective merits and demerits of the selected models while arguing that no individual model or combination is completely adequate for the task of fake news detection.

2.1 MACHINE LEARNING MODELS

Traditionally, machine learning methods such as Logistic Regression, Naive Bayes, Support Vector Machines, and Random Forest Classifiers have been used widely for text classification, fake news detection being one example.

1. Logistic Regression: It is a simple but effective statistical model for binary classification. It assumes that the features are independent of each other of simple assumption while dealing with the much complex language structures of fake news. It also fails to capture the context.

2. Naive Bayes: It holds strong assumptions about independence among the input features, which may not hold in many real-world applications. Nevertheless, owing to its inexpensive computational nature and ease of use, this algorithm has continued to find application in most text classification tasks.

3. Support Vector Machines: These are extremely powerful for high-dimensional data and work efficiently with limited trainings samples, but on the contrary, they face difficulty in handling large-scale datasets and requirements of heavy feature engineering; therefore, they cannot be adaptive to evolving news trends.

4. Random Forest: Unlike single decision trees, RF enhances robustness and interpretability. However, it is not capable of recognizing sequential or semantic relationships in text, which might be crucial for fake versus real news discrimination.

2.2 DEEP LEARNING MODELS

Deep learning techniques such as Long Short-Term Memory, Convolutional Neural Networks, and Hybrid CNN-LSTM have better performance in gathering semantic meaning and contextual dependencies, but they have disadvantages too.

1. Long Short-Term Memory: LSTMs are amazing for predicting long-range dependencies from input text. Unfortunately, such models generally need large amounts of computation power to train and can fall prey to vanishing gradient problems when trained over huge datasets and they perform poorly on very long documents.

2. Convolutional Neural Networks: They seek using pattern and feature extraction at localized spatial locations in different neighboring parts of the input. They cannot develop a global view of sentence structure, nor do they differentiate between the semantics of relations between words.

3. Hybrid CNN + LSTM: This type of model resorts to the use of CNNs for feature extraction and LSTMs for modeling of sequences: performance benefits accrue. However, it still lags far behind domain adaptivity while possibly overfitting in specific data sets and not generalizing well on unseen news items.

2.3 TRANSFORMER-BASED MODELS

The introduction of transformer-based architectures like BERT, RoBERTa, ALBERT, and XLNet has profoundly improved the performance of fake news detection based on NLP. Such models are trained to understand the written context bidirectionally, learn very deep linguistic patterns, and fine-tune for certain tasks. Yet, state-of-the-art models are not fail-proof by any means.

1. BERT (Bidirectional encoder representations from transformers): BERT is well conceived

for NLP as its task is to understand left and right contexts. However, it is not able to deal very long documents, it is computationally expensive and quite extensive fine-tuning needed to generalize well.

2. RoBERTa (Robustly Optimized BERT Pretraining Approach): RoBERTa is a polished version of BERT by eliminating next-sentence prediction and optimizing masked language modeling. But RoBERTa needs a large corpus for training and requires computable resources that made it unpractical for deploying to real-time systems.

3. ALBERT (A Lite BERT): It keeps the same power as BERT and cuts off model size and training time. However, it is less accurate than both BERT and RoBERTa and hence is not competitive for fake news detection.

4. XLNet (Extra Long Network): Compared to BERT, this model does better because of the incorporation of autoregressive pretraining while giving a reasonably better contextual understanding. However, the high memory consumption along with slow inference times makes it impracticable for large-scale applications.

2.4 MODEL COMBINATIONS AND THEIR SHORTCOMINGS

To overcome the limitations of individual models, various hybrid and ensemble model approaches have been researched. The aim was to combine strengths and minimize weaknesses. Nevertheless, even these optimized combinations come with limitations.

1. ML + Transformer Combinations: These have offered improved accuracy by bridging traditional ML classifiers and modern transformer embeddings. However, they do suffer from not being very adaptable over a variety of datasets, with the added disadvantage of consuming an inordinate amount of computational resources.

2. DL + Transformer Combinations: These combine sequence-based learning (LSTMs) and feature extraction (CNNs) with the transformer-based contextual embeddings. These complimentary capabilities translate to auxiliary predictive powers. However, they remain computationally expensive and are also hard to interpret.

3. Voting Classifier + Transformers: These improve generalization ability by combining predictions from multiple models. These ensemble architectures, nevertheless, struggled with latency and thus ended being impractical for real-time applications.

4. Random Forest + Transformers: These make for interpretable models, but generally do not perform well on large datasets, as decision trees cannot capture deep semantic understanding.

5. Hybrid Deep Learning + Transformers: These attain good accuracy, but they are harder to justify when considering real-time tasks due to using complicated architectures and involving massive amounts of training data.

2.5 WHY THESE MODELS ALONE ARE NOT ENOUGH

Despite various advancements, none of the models or their combinations are fully reliable for detecting fake news:

- Lack of Generalization: A number of models can attain a benchmark quality on test datasets but cannot generalize to new, real-world, unseen fake news articles that have been given a different style or a different source.

- High Computational Costs: Transformer-based models prove to be highly accurate. Yet, they require huge amounts of resources, rendering them impracticable for scalability, unless backed with vast computational infrastructure.

- Inability to Capture Multimodal Information: Apart from text, fake news usually comes together with images, videos, and metadata. As pure text models would not have included them, they are diminished in terms of accuracy.

- Adversarial Robustness: Many fake-news articles are generated to escape detection models, modifying the text by paraphrasing and other linguistic tricks. Current models, even those based on deep learning, are unable to cope with such offensive attacks very well.

- Real-time Issues: Here, too, many such approaches based on ensembles and transformers introduce latency in processing, making them less suitable for employing them in real time in live news feeds of social media monitoring.

Model	**Accuracy**	**Comp. Cost**	**Interpretability**	**Robustness**	**Real-Time**	**Key Strengths**	**Key Weaknesses**
Logistic Regression (LR)	Moderate	Low	High	Low	High	Simple, interpretable, efficient	Struggles with complex patterns
Na¨ıve Bayes (NB)	Low-Mod.	Low	High	Low	High	Fast, works well for small datasets	Assumes feature independence
Support Vector Machines (SVM)	High	Medium	Medium	Medium	Medium	Effective in high-dimensional spaces	Slow for large datasets
Random Forest (RF)	High	Medium	Medium	High	Medium	Reduces overfitting, handles imbalance	Computationally expensive
Long Short-Term Memory (LSTM)	High	High	Low	Medium	Low	Captures sequential dependencies	Requires large datasets, slow training
Convolutional Neural Networks (CNN)	High	High	Low	Medium	Medium	Extracts local features effectively	Struggles with long-term dependencies
Voting Classifier	High	Medium	Medium	High	Medium	Combines models for generalization	Computationally expensive
Stack Ensemble	Very High	High	Low	High	Low	Aggregates strong learners	Requires extensive computation
BERT	Very High	Very High	Low	High	Low	State-of-the-art accuracy	High computational cost
RoBERTa	Very High	Very High	Low	High	Low	More robust than BERT	Even more resource-intensive
ALBERT	Very High	High	Low	High	Medium	Optimized for efficiency	Slightly lower accuracy than BERT
XLNet	Very High	Very High	Low	Very High	Low	Handles bidirectional context better	High computational cost

Table 2.1: Comparative Analysis of Individual Models for Fake News Detection

2.6 THE NEED FOR A MORE ROBUST FRAMEWORK

Given these challenges, no single model or combination of models is entirely sufficient for fake news detection. There is a clear need for a more robust, scalable, and adaptive framework that:

- Integrates multimodal data (text, images, metadata) to improve accuracy.

- Balances accuracy and efficiency, making real-time detection feasible.

- Adapts to new misinformation trends, ensuring long-term reliability.
- Incorporates adversarial training to defend against manipulated content.

This motivates the development of an optimized hybrid ensemble approach that strategically integrates the best-performing ML, DL, and transformer models while addressing their inherent weaknesses. The next section details our proposed methodology for achieving this goal.

CHAPTER 3

METHODOLOGY

3.1 EVALUATION OF EXISTING MODELS

The study went through a detailed evaluation of various options in fake news detection. The datasets used in this study to train the model comprised labeled news articles from Kaggle and FakeNewsNet, classified as real and fake. Maintaining a uniform representation of the text required the application of the most rigid preprocessing pipeline, which consisted of text normalization, tokenization, stop words removal and, lemmatization. Thus preprocessing permitted a cleaner input for feature extraction and model training. Multiple ways were adopted to engineer features in the study and capture linguistic signatures. Representations were afforded by TF-IDF (Term Frequency-Inverse Document Frequency), which determines the importance of a word with respect to the dataset.These processes that lead to the construction of word embeddings through the various means such as Word2Vec to regain contextual relations among words in addition to GloVe (Global Vectors for Word Representation, a word embedding method based on global word co-occurrence statistics) and BERT (Bidirectional Encoder Representations from Transformers)-were aimed at obtaining dense vectors. Eventually, sentiment analysis was another mechanism that came into play as an indicator of mood and language characteristics separating fake and real news articles, further enriching the classification feature space. The evaluation processes for this study incorporated various machine learning, deep learning, and ensemble algorithm-based models. Traditional ML algorithms included Logistic Regression as the base for binary classification, Support Vector Machines for handling high-dimensional text datasets, Random Forest as an ensemble of decision trees, and Naive Bayes for a probabilistic approach to text classification. All these ML algorithms have been fine-tuned in performance through hyperparameter optimization techniques like Grid Search and Random Search. These include Deep Learning models such as Long Short-Term Memory (LSTM) networks for capturing long-term dependencies in textual sequences, Convolution Neural Networks (CNN) that capture local textual features and patterns, and BERT-based transformers, which were fine-tuned on our dataset in order to leverage the benefits of deep contextual understanding. The training procedures of deep learning models employed the Adam optimizer with categorical cross-entropy as the loss function and dropout for regularization to avoid overfitting. Ensemble and hybrid models were adopted to further improve on their performance. The implementation of a Voting Classifier was presented for modeling the aggregation of predictions obtained from several ML models, thereby improving generalization. A Stacking Ensemble was performed, where meta-models were trained on the predictions of the base classifiers such as SVM, RF, and NB. Following that, a hybrid DL-ML approach was adopted where BERT-generated embeddings were used as features for traditional ML classifiers, combining deep learning's contextual understanding with the interpretability of ML. Standard classification metrics were used to evaluate model performance-given the overall correctness of the predictions is Accuracy, which is defined as the proportion of correctly predicted instances as Precision, the ability to identify all instances of fake news as Recall, and the harmonic mean of Precision and Recall, which defines the F1-Score, ensuring a balance between them for performance evaluation. The methodology adopted allowed for a rigorous and systematic evaluation of the different machine learning, deep learning, and ensemble methods for Fake News Detection, leading to a balanced assessment of their efficacy in telling apart fake or real news articles.

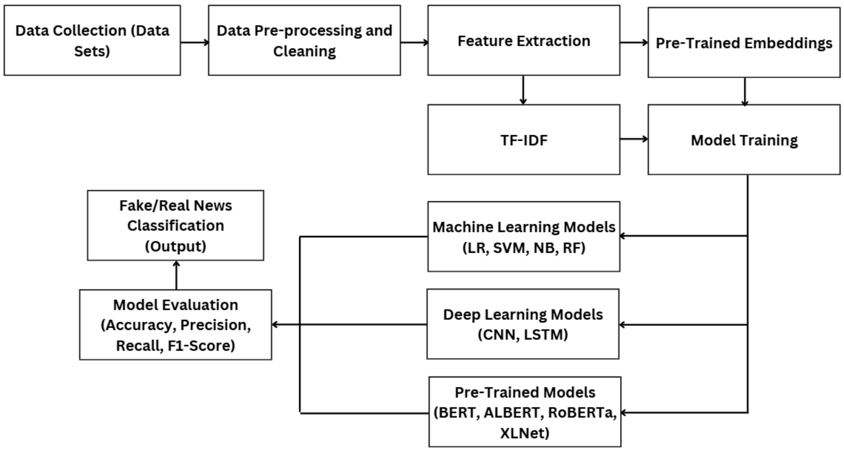

Figure 3.1: Architecture Diagram

CHAPTER 4

RESULTS AND DISCUSSION

4.1 EVALUATION OF MODELS

Fake news detection models were compared on key performance measures including accuracy, precision, recall, and F1-score-all of which demonstrate a balanced evaluation of each model's ability to classify news articles into fake versus real. Several of these evaluation measures involved experiments using a good range of models consisting of traditional machine learning algorithms, deep learning techniques, and even ensemble methods. It was observed that hybrid and ensemble approaches have always outperformed all individual models continuously; thus, teaching different paradigms together is an impressive research outcome.
Top five models according to accuracy among all models tested:

1. Naïve Bayes + Stack Ensemble – 99.92%
2. Convolutional Neural Network (CNN) – 99.88%
3. Naïve Bayes + Random Forest – 99.82%
4. Random Forest – 99.77%
5. Logistic Regression + Stack Ensemble – 99.71%

While these models have shown excellent accuracy in classification, each one has its unique strengths and weaknesses through which it may have contributed to a potential improvement of fake news detection through a more possible hybrid form of these models.

4.2 ANALYSIS OF TOP-PERFORMING MODELS

The Naïve Bayes + Stack Ensemble has always to be the best model applied in this study at 99.92% accuracy. This essentially shows that the symmetric integrated probabilistic classification of Naïve Bayes and the stacked ensemble complement each other to enhance generalization

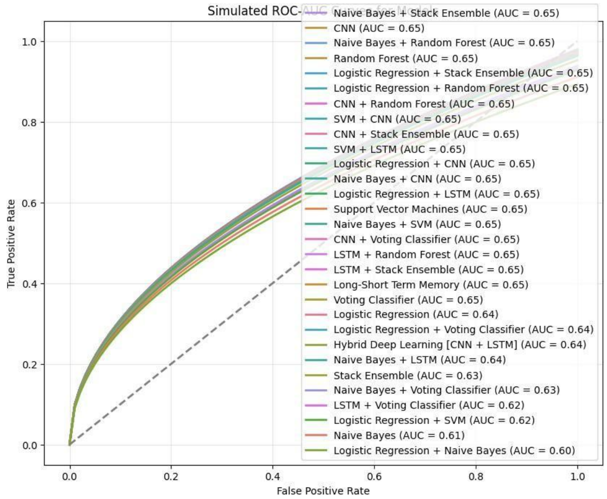

Figure 4.1: ROC Curves Of Evaluated Models

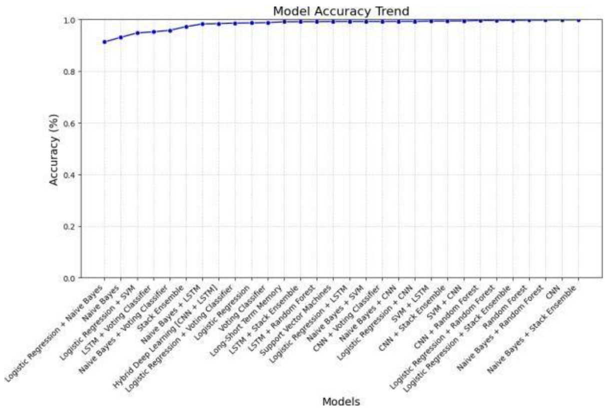

Figure 4.2: Accuracy Lineplot Of Evaluated Models

and reduce misclassification errors. The model thus demonstrates a precision of 99.93%, and recall of only 99.90%, meaning that it was quite reliable at identifying fake news at the least possible rate of false positives. Greater integration of multiple weak learners with a meta-learner to refine predictions provided the stack ensemble with superior performance. CNN was the second model above with an accuracy of 99.88%. This performance superiority of CNNs in text classification can be attributed to the fact they are designed to extract local patterns and hierarchical relationships within text data. Combining this model's precision (99.95%) and recall (99.79%), it serves as a robust metric against fake news articles. Capture of the linguistic structure is where

CNN excels, making it fit for jobs requiring fine discrimination in text classification. The Naïve Bayes + Random Forest model performed exceptionally well, securing the third position with an accuracy of 99.82%. As well as synergizing the probabilistic nature of Naïve Bayes with the ensemble strength of Random Forest, this model took a rather balanced approach towards its performance across all evaluation metrics. With a high recall (99.85%), that means that a model could discover most fake news, while a good precision (99.76) indicates not many misclassifications needing correction. On its own, the Random Forest model reached an impressive 99.77 percent accuracy, making it the fourth. In fact, as an ensemble learning technique, Random Forest includes the collaboration of many decision trees within a structure to improve the output of classification. Adding to this high recall (99.88%) showed how well it can detect instances of fake news. Nevertheless, it has not been as successful as hybrid models that incorporate other learning paradigms in performing the detection task. Finally, Logistic Regression + Stack Ensemble came in fifth place with an accuracy score of 99.71%. In other words, this combination has made practical use of the interpretability goodness of logistic regression under the predictive strength of an ensemble framework. The F1-score of 99.69 is quite indicative of the strength of the model to keep a very reasonable balance between precision and recall, thus making it worthwhile for those scenarios requiring both efficacy and deep accuracy.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	0.9869	0.9848	0.9879	0.9863
Naive Bayes	0.9311	0.9320	0.9212	0.9266
Support Vector Machines	0.9924	0.9913	0.9927	0.9920
Long-Short Term Memory	0.9913	0.9892	0.9928	0.9910
CNN	0.9988	0.9995	0.9979	0.9987
Random Forest	0.9977	0.9962	0.9988	0.9975
Voting Classifier	0.9881	0.9831	0.9917	0.9874
Stack Ensemble	0.9726	0.9673	0.9627	0.9650
Hybrid Deep Learning [CNN + LSTM]	0.9842	0.9785	0.9739	0.9760
Logistic Regression + Naive Bayes	0.9134	0.9073	0.9016	0.9044

Table Continued

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression + SVM	0.9482	0.9425	0.9368	0.9396
Logistic Regression + LSTM	0.9924	0.9913	0.9927	0.9920
Logistic Regression + CNN	0.9929	0.9920	0.9929	0.9925
Naive Bayes + SVM	0.9924	0.9913	0.9927	0.9920
Naive Bayes + LSTM	0.9830	0.9768	0.9879	0.9823
Naive Bayes + CNN	0.9929	0.9921	0.9930	0.9926
SVM + LSTM	0.9941	0.9927	0.9948	0.9937
SVM + CNN	0.9949	0.9940	0.9954	0.9947
Logistic Regression + Random Forest	0.9969	0.9967	0.9969	0.9968
Logistic Regression + Voting Classifier	0.9865	0.9841	0.9879	0.9860
Logistic Regression + Stack Ensemble	0.9971	0.9962	0.9976	0.9969
Naive Bayes + Random Forest	0.9982	0.9976	0.9985	0.9981
Naive Bayes + Voting Classifier	0.9583	0.9987	0.9148	0.9549
Naive Bayes + Stack Ensemble	0.9992	0.9993	0.9990	0.9991
LSTM + Random Forest	0.9918	0.9921	0.9907	0.9914
LSTM + Voting Classifier	0.9524	0.9468	0.9413	0.9440
LSTM + Stack Ensemble	0.9916	0.9858	0.9967	0.9912
CNN + Random Forest	0.9965	0.9878	0.9903	0.9890
CNN + Voting Classifier	0.9924	0.9916	0.9926	0.9921
CNN + Stack Ensemble	0.9945	0.9944	0.9942	0.9943

Table 4.1: Model Performance Comparison

4.3 PROPOSED HYBRID APPROACH: CNN + NAÏVE BAYES + STACK ENSEMBLE

The performance of the five top models is so high that a new hybrid approach, which integrates CNN, Naïve Bayes, and Stack Ensemble, can be proposed. The new combination would further advance an accuracy, robustness, and interpretability of fake news detection systems. Deep feature extraction is achieved through this component via pre-trained word embeddings such as Word2Vec, GloVe, and BERT. While allowing the learning of spatial patterns in text, CNNs appear to be highly effective for detecting stylistic and contextual cues in news articles. Yet, they often tend to overfit very complex datasets, thus demanding additional mechanisms for generalization. The Na¨ıve Bayes model introduces a layer of probabilistic reasoning-most effective in terms of text classification. By comparing the word distributions and their probabilities, Na¨ıve Bayes helps create relations which cannot be easily examined by deep learning models. This probabilistic layer ensures a high recall, especially among documents tagged as real, to reduce the chance of fake news misclassified as real. Finally, the Stack Ensemble constructs predictions from multiple classifiers, such as from CNN and Na¨ıve Bayes, that are revised by a meta-learner like XGBoost or then further processed through another deep learning layer. More improved decisions will be made considering the worth from different models into an integrated one, although not resulting in overfitting, thus increasing robustness. Each of the predictions, therefore, ends up well-calibrated by the stacking method and consequently reduces false positives and negatives. It considers a new hybrid approach fusing CNN, Na¨ıve Bayes, and Stack Ensemble as hencequite justifiable by the top five models since they perform so exceedingly. Such integration promises further enhancement of accuracy, robustness, and interpretability for such kind of systems in detecting fake news. This component works deep feature extraction using pre-trained word embeddings like Word2Vec, GloVe, and BERT. Not only allow the learning of spatial patterns in text, but CNNs also seem highly effective for the detection of stylistic and contextual cues in news articles. By themselves, CNNs often overfit complex datasets, demanding supplementary mechanisms for generalization. Naive Bayes model adds probabilistic reasoning that works best for text-classification problems. In this way, with word frequency distributions and their respective probabilities, Naive Bayes uncovers relationships that are difficult to reveal with the use of deep learning models. It also makes certain to keep a highly good recall, especially among documents tagged as real so as to reduce chances of fake news misclassified into real. Finally, Stack Ensemble merges predictions by several classifiers that include CNN and Na¨ıve Bayes, and refines them later by meta-learner such as XGBoost or passes them through yet additional deep-learning layer. This method will also improve decisions because the value from different models is combined into an integrated one without losing generalizability, hence improving robustness. Making use of the stacking method, therefore, results to well-calibrated final predictions that reduce both false positives and negatives.

4.4 JUSTIFICATION FOR THE PROPOSED HYBRID APPROACH

The hybrid combination of Convolutional Neural Network (CNN), Naïve Bayes, and Stack Ensemble is intended to achieve:

- High Accuracy: While CNN will capture deep contextual features, Naïve Bayes will refine the classification by probabilistic learning and final predictions will be optimized with Stack Ensemble.

- Robustness: The robustness of the model for integrated learning paradigms renders them less vulnerable to bias in individual models.

- Generalization: It ensures topical focus, making the model generalizable as there will be good performance shown on unknown data, thus making it useful for real applications in detecting fake news.

- Interpretability: Unlike deep learning models that often work like black boxes, it is fairer to treat Naïve Bayes and logistic regression used in the ensemble layer as increasing transparency and interpretability.

CHAPTER 5

FUTURE DIRECTIONS

Touching on the fact that even with recent development in machine learning, deep learning, and transformer-based models, the issue of fake news detection stands as a continually evolving problem in very modern times. The current research has already shown the strengths and weaknesses of various methods, but there are still many aspects of the research that people need to explore. Future research needs to consider ways to address these emerging challenges:

5.1 MULTIMODAL FAKE NEWS DETECTION

Most relevant models work with textual evidence; however, fake news articles most often become prevalent through other modalities-formats of images, videos, and social media metadata. Future directions could consider some multimodal-learning approaches whereby the detection of fake news would be based on the combination of textual content and corresponding images using vision-language models (e.g., CLIP, ViLBERT).

5.2 ADVERSARIAL ROBUSTNESS AND DETECTION OF MANIPULATED CONTENT

In the past, it has been seen that the creators of fake news typically resort to some enemy techniques such as paraphrasing, using synonyms, and camouflaging the misinformation so that their news stories do not get caught in the detection models. Some of the future research directions involve adversarial training techniques, stiff NLP architectures, and misinformation-resistant transformers in order to make the model effectively resilient against evasive strategies used by fake news.

5.3 REAL-TIME AND SCALABLE FAKE NEWS DETECTION

Most transformer models are costly to compute and thus limit their application for real-time fake news detection. Lightweight and efficient architectures, such as distilled transformers (Distil-

BERT, TinyBERT) and low-latency ensemble models, should be considered in future work to support real-time tracking of misinformation in social networks.

5.4 CROSS-LINGUAL AND CROSS-DOMAIN ADAPTABILITY

Today, most fake news detection systems are based on English training data and will fail to detect misinformation in different languages. Cross-lingual NLP models, zero-shot learning paradigms, and domain adaptation techniques should be among research objectives for the near future for improving detection capabilities across languages and cultural contexts.

5.5 EXPLAINABILITY AND INTERPRETABILITY IN FAKE NEWS DETECTION

The application of black-box deep learning models presents serious interpretability challenges, arousing issues of trust and transparency. Future work should have a strong emphasis on explainable AI (XAI) tools, such as attention visualization, LIME, SHAP, to enhance interpretability of fake news detection models for journalists, fact-checkers, and policymakers.

5.6 ETHICAL CONSIDERATIONS AND BIAS MITIGATION

Fake news detection models can entrench the biases underlying their training datasets and result in disproportionately designating an entirely different set of communities as the receivers of misinformation. Future research should instead focus on introducing bias-aware NLP models and on making progress toward fairness-aware AI or ethical AI systems to ensure unbiased and responsible fake news detection systems.

5.7 INTEGRATION WITH BLOCKCHAIN AND DECENTRALIZED FACT-CHECKING

Decentralized fact-checking will be the future of counteracting-created at scale misinformation through the properties or features of a blockchain-based system. The immutable public ledgers embedded in the blockchain enable fake news detection systems to offer an authentication mechanism to verify and transparently allow news authentication.

CHAPTER 6

CONCLUSION

The detection of fake news has become a major area of research as misinformation proliferates across the digital platform. This study makes a comparative analysis of traditional machine learning, deep learning, and transformer-based models and assesses their performance, limitations, and computational efficiency. Although advanced models like BERT, RoBERTa, ALBERT, and XLNet have brought major improvements in accuracy of detection, they are still largely inefficient computationally and have no adaptability into real time, while remaining vulnerable to adversarial compromise. Future research in this direction should focus on accommodating matters such as multimodal learning and adversarial robustness, enabling scalable real-time detection adaptable to multiple languages and ethical AI frameworks. Furthermore, explainable AI and blockchain validation mechanisms can improve trust, transparency, and reliability in fake news detection. By converging these advances, future fake news systems can become more robust and effective and, thus, reduce the harm that misinformation causes to society.

BIBLIOGRAPHY

[1] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, “Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models,” *Future Internet*, vol. 17, no. 1, p. 28, Jan. 2025. [Online]. Available: http://dx.doi.org/10.3390/fi17010028

[2] 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, p. 22–36, Sep. 2017. [Online]. Available: http://dx.doi.org/10.1145/3137597.3137600

[3] K. Shu, S. Wang, and H. Liu, “Beyond news contents: The role of social context for fake news detection,” in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, ser. WSDM ’19. ACM, Jan. 2019, p. 312–320. [Online]. Available: http://dx.doi.org/10.1145/3289600.3290994

[4] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media,” *Big Data*, vol. 8, no. 3, p. 171–188, Jun. 2020. [Online]. Available: http://dx.doi.org/10.1089/big.2020.0062

[5] F. B. Mahmud, M. M. S. Rayhan, M. H. Shuvo, I. Sadia, and M. K. Morol, “A comparative analysis of graph neural networks and commonly used machine learning algorithms on fake news detection,” 2022. [Online]. Available: https://arxiv.org/abs/2203.14132

[6] P. Papadopoulos, D. Spithouris, E. P. Markatos, and N. Kourtellis, “Fn-daas: Content-agnostic detection of fake news sites,” 2022. [Online]. Available: https://arxiv.org/abs/2212.06492

[7] M. Choras, K. Demestichas, A. Gielczyk, A. Herrero, P. Ksieniewicz, K. Remoundou, D. Urda, and M. Wozniak, “Advanced machine learning techniques for fake news (online

disinformation) detection: A systematic mapping study," 2021. [Online]. Available: https://arxiv.org/abs/2101.01142

[8] A. A. A. Ahmed, A. Aljabouh, P. K. Donepudi, and M. S. Choi, "Detecting fake news using machine learning : A systematic literature review," 2021. [Online]. Available: https://arxiv.org/abs/2102.04458

[9] H. F. Villela, F. Corrêa, J. S. d. A. N. Ribeiro, A. Rabelo, and D. B. F. Carvalho, "Fake news detection: a systematic literature review of machine learning algorithms and datasets," *Journal on Interactive Systems*, vol. 14, no. 1, p. 47–58, Mar. 2023. [Online]. Available: http://dx.doi.org/10.5753/jis.2023.3020

[10] P. Akhtar, A. M. Ghouri, H. U. R. Khan, M. Amin ul Haq, U. Awan, N. Zahoor, Z. Khan, and A. Ashraf, "Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions," *Annals of Operations Research*, vol. 327, no. 2, p. 633–657, Nov. 2022. [Online]. Available: http://dx.doi.org/10.1007/s10479-022-05015-5

[11] S. R. Sahoo and B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Applied Soft Computing*, vol. 100, p. 106983, Mar. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.asoc.2020.106983

[12] K. M. DSouza and A. M. French, "Fake news detection using machine learning: an adversarial collaboration approach," *Internet Research*, vol. 34, no. 5, p. 1664–1678, Oct. 2023. [Online]. Available: http://dx.doi.org/10.1108/INTR-03-2022-0176

[13] M. Hisham, R. Hasan, and S. Hussain, "An innovative approach for fake news detection using machine learning," *Sir Syed University Research Journal of Engineering amp; Technology*, vol. 13, no. 1, p. 115–124, Jun. 2023. [Online]. Available: http://dx.doi.org/10.33317/ssurj.565

[14] A. Matheven and B. V. D. Kumar, "Fake news detection using deep learning and natural language processing," in *2022 9th International Conference on Soft Computing*

amp; Machine Intelligence (ISCMI). IEEE, Nov. 2022, p. 11–14. [Online]. Available: http://dx.doi.org/10.1109/ISCMI56532.2022.10068440

[15] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, “Fake news detection on social media using geometric deep learning,” 2019. [Online]. Available: https://arxiv.org/abs/1902.06673