I.J. Information Technology and Computer Science, **2024, 6, 89-100**
Published Online on December 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2024.06.07

Performance Evaluation of Fake News Detection Models

Afeez Ayomide Olagunju*
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
E-mail: olagunjuafeez@gmail.com
ORCID iD: https://orcid.org/0000-0001-8811-8414
*Corresponding author

Iyabo Olukemi Awoyelu
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
E-mail: iawoyelu@oauife.edu.ng
ORCID iD: https://orcid.org/0000-0003-3523-6753

Received: 17 August 2024; Revised: 20 September 2024; Accepted: 17 October 2024; Published: 08 December 2024

Abstract: The rapid spread of misinformation on social media platforms, especially Twitter, presents a challenge in the digital age. Traditional fact-checking struggles with the volume and speed of misinformation, while existing detection systems often focus solely on linguistic features, ignoring factors like source credibility, user interactions, and context. Current automated systems also lack the accuracy to differentiate between genuine and fake news, resulting in high rates of false positives and negatives. This study investigates the creation of a Twitter bot for detecting fake news using deep learning methodologies. The research assessed the performance of BERT, CNN, and Bi-LSTM models, along with an ensemble model combining their strengths. The TruthSeeker dataset was used for training and testing. The ensemble model leverages BERT's contextual understanding, CNN's feature extraction, and Bi-LSTM's sequence learning to improve detection accuracy. The Twitter bot integrates this ensemble model via the Twitter API for real-time detection of fake news. Results show the ensemble model significantly outperformed individual models and existing systems, achieving an accuracy of 98.24%, recall of 98.14%, precision of 98.42%, and an F1-score of 98.24%. These findings highlight that combining multiple models can offer an effective solution for real-time detection of misinformation, contributing to efforts to combat fake news on social media platforms.

Index Terms: Fake News Detection, Ensemble Learning, Deep Learning, Twitter Bot, Social Media Platforms.

1. Introduction

Social networking sites, such as Twitter, have revolutionized the dissemination of information by providing a platform for effortless content sharing. The vast user base of these networks further amplifies the reach and virality of shared content. Consequently, the prevalence of both genuine and fake news on these platforms has created a pervasive challenge in distinguishing between reliable and misleading information. Despite the efficacy of conventional fact-checking techniques in assessing news veracity, the substantial volume and swift dissemination of fake news on social media undermine the effectiveness of these approaches [1]. The exponential proliferation of misinformation necessitates a more effective strategy to address the issue.

Existing fake news detection models or algorithms have explored various ways in which fake news can be detected and stopped on social networking sites like Twitter [2-4], still, fake news spread across the social networking sites meaning these models or algorithms are not implemented in a way to stop the spread of fake news efficiently. In social networking sites like Twitter, users should be provided with ways in which they can detect fake news on the go and be able to spread the truth to debunk the hoarse. A Twitter bot for automated fake news detection provides an avenue for fake news to be detected by users on the go, automatically and immediately so that the fake news can be debunked in time before it causes chaos.

Twitter is a social networking platform that enables users to connect and transmit brief messages known as "tweets". It is characterized by its microblogging format, where users can share text-based content of up to 280 characters (until recently that it allows longer characters), along with images, videos, and links. Twitter is the fourth most visited online platform. 69 per cent of Twitter users admit that they prefer Twitter over other platforms to stay informed about events

[5]. It has been discovered that over 50 per cent of Americans access news from social media while over 23 per cent use Twitter to access news [6]. In Nigeria, 31.6 million people use social media; many of these users access their news from Twitter. A report from Twitter advertisement shows that Nigeria has 4.95 million active users as of early February 2023 [7].

A Twitter bot is an automated program or script that interacts with the Twitter platform. It is intended to carry out particular duties or operations on Twitter without requiring direct human intervention. Twitter bots can be created to perform various actions, such as posting tweets, retweeting content, following or unfollowing users, liking tweets, replying to mentions, and more [8]. They can also be used to disseminate news with little or no human involvement [9]. An automated Twitter bot can employ deep learning algorithms and natural language processing methods to analyse massive quantities of tweets, discover patterns suggestive of fake news, and provide to users, real-time notifications.

[10] defined fake news as the behaviour of intentionally deceiving the public by publishing and spreading false information. Fake news is disastrous in turning the people against the government or bringing lack of trust in government policies. It can influence people's opinion towards another direction mostly for the personal gains of the poster [10]. It can pose serious health dangers and negatively affect the lives of people who accept the news as true such as the case of Corona Virus Disease of 2019 (COVID 19) fake news spread around during the pandemic [11].

People don't just accept fake news as truth. Some psychological phenomena make people accept fake news as true. Confirmation bias is one of many psychological phenomena that are mostly attributed to the spread of fake news such that people would only believe what they want to believe [12]. This is a situation in which people have an affinity to fall for fake news that appeals to their personal beliefs. They easily accept and spread this information. Another psychological phenomenon is Truth bias which makes people believe the information they are accessing is true until they find it suspicious. Research shows that people with a “truth bias” “think about the truth” when interacting with others and “judge social information as accurate, only to abandon this notion when something in the situation causes suspicion” [13]. Another psychological phenomenon is the “mass appeal” phenomenon in which when a person sees that certain information is accepted and spread by the majority, they tend to develop a soft spot for that information and accept it as a truth. These psychological occurrences make it hard to stop the spread of fake news in society.

Fake news is disseminated by mostly four categories of people on the internet and the categories are innocent users who thought the information they are sharing is credible, trolls whose intention is to deliberately wreak havoc on an individual, society or organization, bots which are automated accounts or applications designed to spread fake news and cyborg, a combination of human and bot to spread misinformation [10]. The use of bots and cyborgs has made fake news detection to become a humongous task, especially for traditional fact-checkers because this news is spread at an unprecedented rate which might be too overwhelming for human fact-checkers to detect in time.

Many current fake news detection methods explore the use of techniques or combinations of techniques to detect fake news. Some search engines use only descriptive terms, while others use web analytics, credibility, and more. It has been shown that using language analysis alone will not be effective in detecting fake news because using speech analysis will lead to many false positives in detecting fake news [2]. While most fake news is designed to stimulate imagination, curiosity, and discontent, social media users can’t really tell the difference between real news and fake news well crafted.

2. Related Works

The identification of fake news on Twitter has attracted significant focus in recent studies, employing various approaches to tackle the widespread problem of misinformation. These studies have made significant contributions by exploring different facets of fake news detection and highlighting the complexities involved.

The study in [14] focused on COVID-19 misinformation on Twitter, aiming to evaluate author credibility as a way to determine news authenticity. The Information Gain approach ranked significant attributes, whereas the K-Nearest Neighbour (KNN) algorithm was utilised for classification. This approach achieved an accuracy of 91% and established a correlation between author credibility and the presence of hoaxes. However, concerns were raised about the credibility metric based on follower count, which could potentially lead to false positives.

The study in [4] adopted a broader approach, analysing both textual content and user profile information from Twitter. This study utilised deep learning techniques, including LSTM, Hierarchical Attention Network, and Natural Language Processing (NLP), to provide probabilities for verifying the authenticity of articles. With an accuracy of 93.4%, this method improved upon previous fake news detection techniques. Nevertheless, the reliance on user characteristics such as verified status or follower count raised concerns about possible false positives.

The research in [15] tackled the general issue of disinformation and fake news on Twitter using machine learning algorithms like XGBoost to categorise user characteristics. Deep learning techniques, such as CNN-RNN and BERT, were integrated for the classification of tweet text, resulting in a remarkable accuracy of 98.53%. However, users were required to exit the Twitter platform to verify news authenticity, indicating a need for real-time automated detection.

The study in [16] performed a comparative analysis of machine learning, deep learning, and transformer models for the detection of fake news. The research employed the ISOT Fake News Dataset and many classifiers, such as Random Forest, Naïve Bayes, SVM, LSTM, and GRU. The results showed that deep learning classifiers outperformed others, achieving an accuracy of 98.7%. However, the study did not address news sources or provide an automated solution for social media fake news detection.

The research in [17] focused on detecting breaking rumours on social media using the Bi-directional LSTM-RNN model. This model attained a precision of 0.87, a recall of 0.88, and an F1 score of 0.87. The sole dependence on linguistic characteristics for rumour identification prompted enquiries over its efficacy across various forms of fake news.

The study in [18] analysed linguistic characteristics that distinguish real news from fabricated news, evaluating multiple machine learning classification techniques, including Support Vector Machine (SVM), Decision Tree, Random Forest, and Logistic Regression. XGBoost was found to be the most effective, with an accuracy of 98.1%. However, this study primarily focused on linguistic cues and did not incorporate network analysis or source credibility, which are also essential for fake news detection.

The research conducted in [2] proposed an enhanced classification model for detecting fake news on social media by a stacking ensemble method utilising SVM, RNN, and Random Forest algorithms applied to the PHEME dataset. The proposed model achieved an accuracy of 81.9%, precision of 82%, and sensitivity of 82%. Despite these promising results, the model remained in the simulation phase and had not been implemented into a functioning system like a website or bot.

The study in [3] explored unsupervised fake news detection using Bayesian probability graphical models and Gibbs Sampling. This approach achieved an accuracy of 75.9%, outperforming other unsupervised techniques and performing slightly better than basic supervised methods with n-grams. However, it did not surpass the performance of newer supervised learning methods and raised concerns about potential biases due to reliance on user opinions and credibility.

The study in [19] presented the LIAR dataset to remedy the deficiency of annotated benchmark data for false news identification. The research employed a Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory Network (Bi-LSTM) layers to capture interdependencies across metadata vectors. Despite the valuable dataset, the model's accuracy did not exceed 30%, indicating room for improvement.

[20] conducted recent research on the detection of fake news in Arabic tweets with deep learning techniques. The study examined integrating contextualized word embeddings (ARBERT and MARBERT) with CNN and BiLSTM architectures. The MARBERT-CNN model attained accuracy, precision, recall, and F1 scores of 0.7467, 0.7490, 0.7467, and 0.7400, respectively. This research is notable for its application to Arabic tweets, addressing misinformation in the Arabic-speaking world. However, its performance in practical real-world deployment remains uncertain.

[21] investigated the categorization of bogus news utilizing transformer-augmented LSTM and BERT. The study, conducted on the FakeNewsNet dataset, demonstrated that the BERT + LSTM model surpassed baseline methods like as TCNN-URG and SAFE, achieving an accuracy of 88.75% on the PolitiFact dataset and 84.10% on the GossipCop dataset subset of FakeNewsNet. Despite the promising results, challenges remain in distinguishing between fake and real news when linguistic similarities are high, and the dataset size used in the study is somewhat limited.

3. Methodology

This research aims to create a Twitter bot that automates the identification of fake news, allowing users to authenticate material and assist in reducing the dissemination of misinformation on Twitter. The work aims to locate a substantial labelled dataset on false news, establish a fake news detection model utilizing deep learning methods, install a Twitter bot based on the constructed model, and assess the model's performance to verify its efficacy in real-world applications.

A methodical process was employed to accomplish these aims. The TruthSeeker dataset, a publicly available resource tailored for fake news identification, was employed for its extensive labelled data, essential for model training and evaluation. The fake news detection model was developed using Python and machine learning toolkits such as Scikit-learn, TensorFlow, and PyTorch. Deep learning algorithms like BERT, CNN, and Bi-LSTM, that according to literature are better than machine learning algorithms were applied, alongside feature extraction, linguistic analysis, source credibility assessment, and contextual analysis to enhance accuracy [22]. The Twitter bot was developed utilizing the Twitter API and Tweepy to automate the identification of fake news. This bot functioned as an independent account for the verification of fake news, incorporating detection capabilities via Tweepy for the automated classification of news as authentic or fraudulent. The model's efficacy was assessed using metrics such as accuracy, precision, recall, and F1-score. A distinct test dataset, separate from the training data, was employed to assess prediction accuracy. Comparative analysis with existing models was conducted to measure its effectiveness.

3.1. TruthSeeker Dataset

The initial phase involves gathering a dataset of tweets containing both genuine and fake news. The dataset is sourced from the Kaggle Data Science Community, specifically the TruthSeeker dataset. This dataset includes labelled and categorized fake and real news, with a focus on political and popular event news relevant to the research scope. As the largest and most comprehensive ground-truth dataset for fake news detection on Twitter, TruthSeeker stands out for its focus on Twitter while also being a valuable resource for related platforms. Distinguished by its vast size, it boasts an extensive collection of 134,198 labelled tweets, setting it apart as the most comprehensive ground-truth dataset for social media misinformation detection. This dataset goes beyond just the tweet text by including detailed metadata on the posting users, such as credibility scores, bot likelihood, and user influence metrics, followers count and many other fields. Such rich data facilitates a thorough analysis of how fake news propagates, offering valuable insights into its spread [23].

Fig 1 shows the snapshot of the TruthSeeker dataset with some of its fields and records.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	majority_	statement	BinaryNur	tweet	followers_	friends_c	favourites	statuses_	listed_cou	following	embeddir	BotScore	BotScoreB	cred
0	TRUE	End of evi	1	@POTUS	4262	3619	34945	16423	44	0	[[0 0 0 0 0	0.03	0	0.540794
1	TRUE	End of evi	1	@S0SickR	1393	1621	31436	37184	64	0	[[0 0 0 0 0	0.03	0	0.462177
2	TRUE	End of evi	1	THE SUPRI	9	84	219	1184	0	0	[[0 0 0 0 0	0.03	0	0.096774
3	TRUE	End of evi	1	@POTUS	4262	3619	34945	16423	44	0	[[0 0 0 0 0	0.03	0	0.540794
4	TRUE	End of evi	1	@OhCom	70	166	15282	2194	0	0	[[0 0 0 0 0	0.03	0	0.29661
5	TRUE	End of evi	1	I've said th	29010	5081	120924	89474	405	0	[[0 0 0 0 0	0.03	0	0.850958
6	TRUE	End of evi	1	As many f	33	21	432	62	0	0	[[0 0 0 0 0	0.03	0	0.611111
7	TRUE	End of evi	1	@Thoma	919	620	131682	216319	71	0	[[0 0 0 0 0	0.03	0	0.597141
8	TRUE	End of evi	1	@Socialis	18002	15754	969539	497128	48	0	[[0 0 0 0 0	0.03	0	0.533298
9	TRUE	End of evi	1	@daysofa	0	0	2	132	0	0	[[1 0 1 1 0	0.971	1	0
10	TRUE	End of evi	1	BREAKING	200231	3529	44820	86430	1279	0	[[0 0 0 0 0	0.03	0	0.982681
11	TRUE	End of evi	1	@aplemk	9	287	706	2654	0	0	[[0 1 0 0 0	0	0	0.030405
12	TRUE	End of evi	1	@LINDAK	211	418	12437	6615	0	0	[[0 0 0 0 0	0.03	0	0.335453
13	TRUE	End of evi	1	No	2389	2417	1446	10044	17	0	[[0 0 0 0 0	0.03	0	0.497087
14	TRUE	End of evi	1	The	157	354	18785	1789	0	0	[[0 0 0 0 0	0.03	0	0.307241
15	FALSE	End of evi	1	@Riguy_	1	16	58	200	0	0	[[0 1 0 0 0	0.03	0	0.058824
16	TRUE	End of evi	1	@Andreal	90513	657	11944	6590	343	0	[[0 0 0 0 0	0.03	0	0.992794
17	TRUE	End of evi	1	@dhiggin	128	1025	19927	12211	0	0	[[0 0 0 0 0	0.03	0	0.111015
18	TRUE	End of evi	1	Ohio cong	2476	3062	6456	7685	1	0	[[0 0 0 0 0	0.03	0	0.447093
19	TRUE	End of evi	1	@anthony	80	46	1042	5435	2	0	[[0 0 0 0 0	0.03	0	0.634921
20	FALSE	End of evi	1	@Jim_Jor	8	19	54	1644	0	0	[[0 0 0 0 0	0.03	0	0.296296
21	TRUE	End of evi	1	@POTUS	9147	4984	437694	552990	124	0	[[0 0 0 0 0	0.03	0	0.6473
22	TRUE	End of evi	1	@Jim_Jor	30	226	10346	2952	0	0	[[0 0 0 0 0	0.03	0	0.117188
23	TRUE	End of evi	1	HUD Sec F	27058	2768	11413	87834	842	0	[[0 0 0 0 0	0.03	0	0.907195
24	TRUE	End of evi	1	@Jeff_A_	2033	4991	98311	57030	6	0	[[0 0 0 0 0	0.03	0	0.289436

Fig.1. TruthSeeker dataset

3.2. Fake News Detection Model

This study's development of the Fake news detection model was a thorough procedure employing many analytical methods to precisely categorize content as authentic or fraudulent. The model was constructed on the basis of three fundamental processes: language examination, source reliability assessment, contextual evaluation. Linguistic analysis focused on the textual features and language patterns indicative of fake news, while source credibility analysis assessed the trustworthiness of the information's origin. Contextual analysis examined the surrounding circumstances and background information related to the news content, ensuring that the model could accurately interpret nuanced details. Utilizing deep learning algorithms such as BERT, Bi-LSTM, CNN, and their stacked ensemble, these processes constituted the foundation of the fake news detection model, facilitating great accuracy and reliability in differentiating between genuine and deceptive information. The proposed study intends to create a Twitter bot for the identification of misinformation, with a particular emphasis on trending news events. The framework employs a combination of diverse deep-learning algorithms following data preprocessing. Fig 2 illustrates the design of the proposed framework.

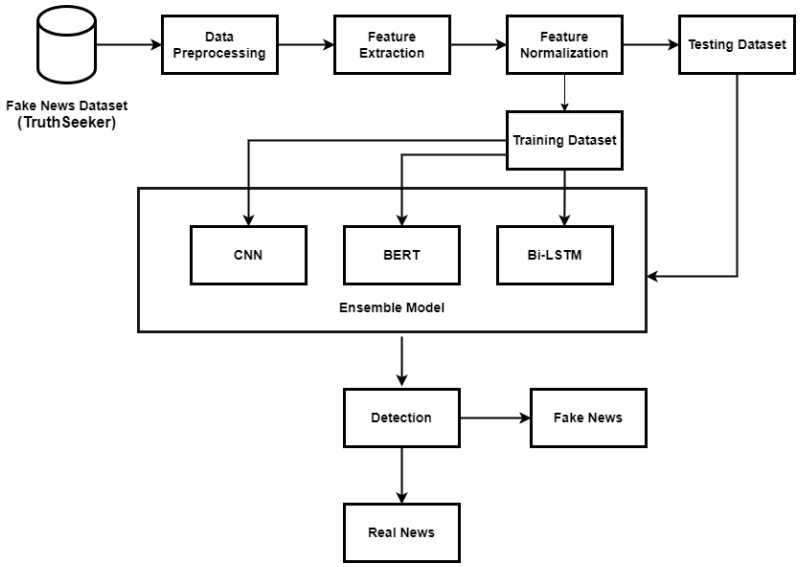

Fig.2. Proposed architecture for the fake news detection

3.3. Fine-tuning BERT Model

The Bidirectional Encoder Representations from Transformers (BERT) model, created by Google in 2019, was incorporated into the fake news detection system utilizing the TruthSeeker dataset. The BERT base uncased model was

used, which simplifies the preprocessing phase and handles the varied case usage in social media text. The TruthSeeker dataset underwent comprehensive preprocessing, including removing URLs and expanding contractions. The BERT base uncased tokenizer was used to convert raw text into tokens.

Fine-tuning involved adapting the pre-trained BERT model to detect fake news using transfer learning approach. The model was trained on the TruthSeeker dataset, divided into training, testing and validation subsets, to optimize parameters and minimize the loss function. The AdamW optimizer was employed for its adaptive learning rate and weight regularization. Hyperparameters such as learning rate, batch size, and epoch count were meticulously optimized to attain optimal performance.

Following training, the BERT model's efficacy was assessed utilizing criteria including accuracy, precision, recall, and F1-score. The fine-tuned BERT model illustrated in Fig 3 demonstrated great accuracy and reliability in categorizing news articles as either fake or authentic.

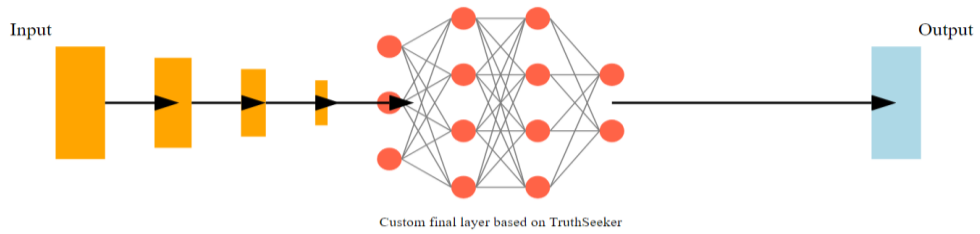

Fig.3. Fine-tuned BERT model

3.4. CNN Model

The CNN model, a crucial tool in identifying fake news, is designed to process and classify news tweets effectively. The model's architecture includes several distinct layers, including an embedding layer that converts input text into dense vectors, multiple convolutional layers that apply filters to extract features, and a ReLU activation function to incorporate nonlinearity. Max-pooling layers minimize the dimensionality of feature maps, focusing on the most important characteristics to reduce computational complexity and avoid overfitting. The final layer is a softmax layer that outputs the probability of each class (fake or real).

The training method entails optimizing the CNN model to precisely classify tweets as either fake or real. Data preparation entails preprocessing the TruthSeeker dataset, selecting features such as bot score, credibility score, and influence score, and partitioning the dataset into training and validation sets. The model is constructed with a suitable loss function and optimizer, and trained on the training set utilizing batch processing. The validation set monitors the model's performance and guards against overfitting.

The CNN model's proficiency in detecting false news stems from its capacity to discern subtle patterns in textual data and its effective training procedures. The model's performance is assessed on the validation set post-training using conventional measures, including accuracy, precision, recall, and F1-score.

The CNN model is quite good at detecting false news because of its capacity to identify little patterns in textual data. The model's ability to effectively classify news tweets and aid in reducing the spread of false information on social media platforms is attributed to its utilisation of convolutional layers' advantages and suitable training methodologies.

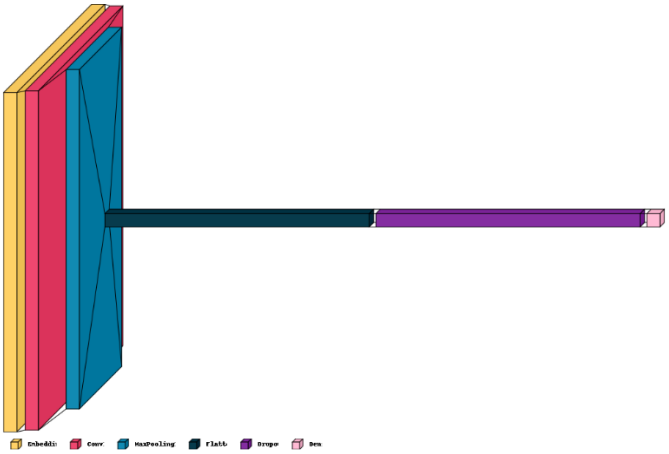

Fig.4. CNN model

3.5. Bi-LSTM Model

The Bi-LSTM model is a valuable tool for detecting fake news, as it captures dependencies in both forward and backward directions within a sequence. The model architecture comprises an embedding layer, bidirectional LSTM layers,

dropout layers, a fully connected layer, and an output layer. The initial layer converts input data into dense vectors, encapsulating the semantic meaning of words.

The Bi-LSTM layers analyse the data in both forward and backward orientations, collecting dependencies from preceding and subsequent contexts. Dropout layers mitigate overfitting by deactivating a proportion of input units during training. The output from the final Bi-LSTM layer is passed to a fully connected layer, which integrates features extracted by the Bi-LSTM layers. The final layer is a softmax layer that outputs the probability of each class (fake or real).

The training process involves data preparation, splitting data, model compilation, model training, and evaluation. The model is trained using batch processing and modified weights based on discrepancies between expected and actual labels. The model's performance is assessed using common measures like accuracy, precision, recall, and F1-score.

Fig 5 shows the depiction of the different layers of the Bi-LSTM model according to [24].

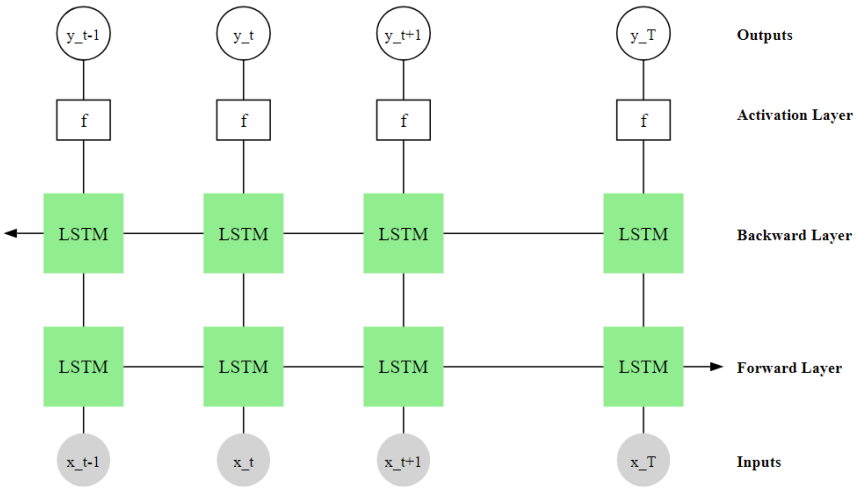

Fig.5. The Bi-LSTM model

4. Results and Discussion

This section analyses the performance of this fake news detection models. The results obtained from various deep learning algorithms, including BERT, CNN, and Bi-LSTM, as well as the ensemble model that combines their strengths were examined and the results of the interface and twitter bot implementation were discussed.

4.1. Results of the Fine-tuned BERT Model

The BERT model, a deep learning architecture, has exhibited remarkable efficacy in the identification of false information. The model attained a remarkable accuracy of 0.9796, signifying that it accurately categorized the majority of tweets in the sample. It was equally good at detecting true positive cases and minimising false positive cases. The model's recall and precision were both 0.9796, with a balanced performance shown by the F1 score of 0.9796.

The confusion matrix offers a comprehensive insight into the model's efficacy, indicating a predominance of true positives and true negatives, with false positives and false negatives being negligible. This distribution underscores the model's capacity to sustain elevated accuracy and dependability across both categories.

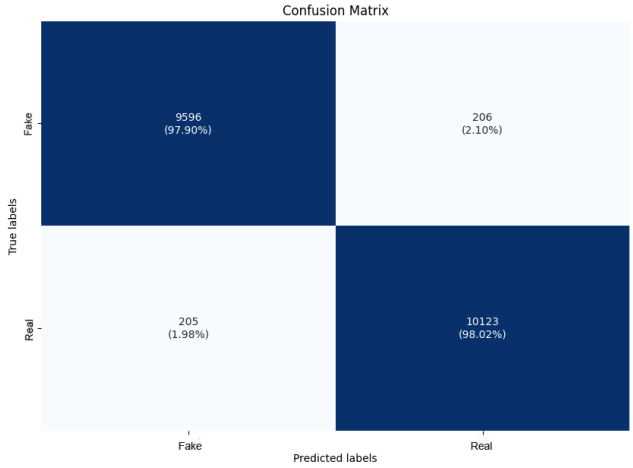

Fig.6. BERT's model confusion matrix

The BERT model exhibits robustness in its performance on unfamiliar data, showcasing its capacity to generalize beyond the training sample. The capacity to generalize is essential in practical applications, where the model faces novel and varied news pieces. The model's deep learning architecture, which includes multiple layers of transformers capable of understanding and processing complex linguistic structures, allows it to effectively apply its learned knowledge to new, unseen examples, making it a reliable choice for fake news detection.

4.2. Results of the CNN Model

The CNN model has remarkable efficacy in identifying bogus news, achieving an overall accuracy of 0.9786, which reflects a well-balanced performance between precision and recall. The model achieved an F1 score of 0.9791, signifying a substantial quantity of true positives and true negatives, accompanied by negligible false positives and false negatives. This signifies that the CNN model is adept at accurately distinguishing between bogus and genuine news pieces.

The CNN model's architecture, designed to capture local patterns in the text, played a crucial role in its performance. By concentrating on particular phrases and terms commonly linked to misinformation, the model could discern subtle distinctions between authentic and fabricated news pieces. The capability to discern local elements and patterns in the text enabled the CNN model to attain elevated precision and recall, enhancing its total accuracy.

Upon assessment with novel data, the CNN model exhibited robust performance, demonstrating its capacity to generalize beyond the training dataset. This resilience is crucial for practical applications, as the model will face a range of novel and different news pieces. The CNN model's superior accuracy, precision, recall, and F1 score, along with the insights from the confusion matrix, illustrate its efficacy in identifying fake news.

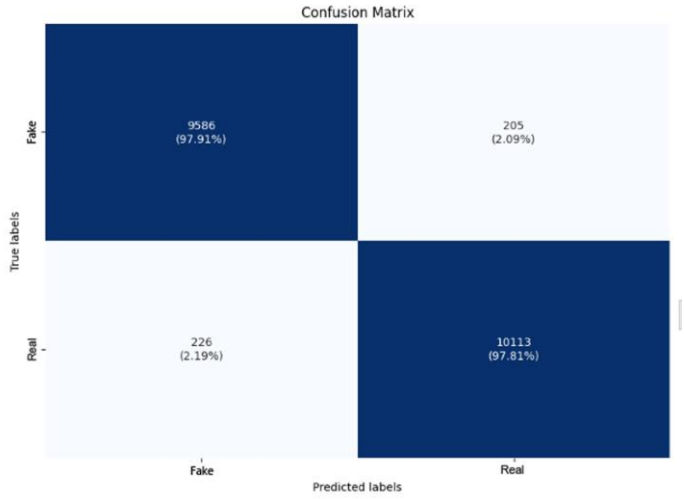

Fig.7. CNN model’s confusion matrix

4.3. Results of the Bi-LSTM Model

The Bi-LSTM model, a reliable tool for detecting fake news, has exhibited remarkable performance metrics. It achieved an impressive accuracy of 0.9805, accurately classifying the majority of tweets within the dataset. The model's recall was recorded at 0.9873, indicating its ability to identify true positive cases of fake news with high precision. Precision was 0.9750, slightly lower than recall, indicating a high level of accuracy in identifying fake news without excessively misclassifying real news as fake. The F1 score, the harmonic mean of precision and memory, was 0.9811, indicating a balanced performance between precision and recall.

The confusion matrix for the Bi-LSTM model elucidates its performance, indicating a substantial quantity of true positives (accurately identified fake news) and true negatives (accurately identified real news), alongside a limited number of false positives (real news incorrectly classified as fake) and false negatives (fake news incorrectly classified as real). This distribution validates the model's efficacy in reducing instances of overlooked fake news while preserving an acceptable level of precision.

When tested on unseen data, the Bi-LSTM model demonstrated strong generalization capabilities, crucial for real-world applications where the model encounters new and diverse news articles. The architecture of the Bi-LSTM model, which includes bidirectional layers that capture information from both past and future sequences in the data, contributes to its strong performance. The training process further refines these capabilities, enabling the Bi-LSTM model to become a powerful tool for fake news detection.

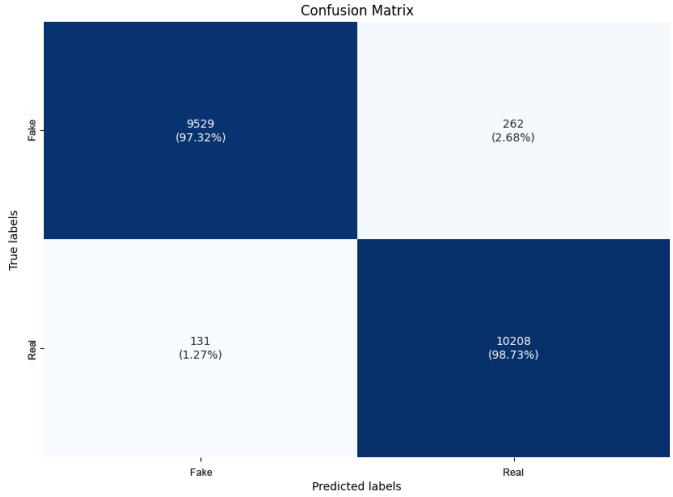

Fig.8. Bi-LSTM model's confusion matrix

4.4. Results of the Ensemble Model

The Ensemble model, integrating the advantages of BERT, Bi-LSTM, and CNN architectures, has exhibited outstanding efficacy in the detection of fake news. Its overall accuracy was 0.9843, indicating higher precision in classifying news tweets compared to individual models. This enhanced accuracy results from leveraging the combined predictive power of individual models, minimizing limitations inherent in each. The F1 score of the Ensemble model was 0.9845, indicating an ideal equilibrium between precision and recall. The accuracy value of 0.9852 demonstrates the model's remarkable proficiency in properly identifying genuine news articles with little false positives, whereas the recall value of 0.9839 reflects the model's effectiveness in detecting a significant proportion of fake news instances with few false negatives.

The Ensemble model's architecture capitalizes on the complementary strengths of the BERT, Bi-LSTM, and CNN models, harnessing their diverse approaches to capture a broad spectrum of features and patterns in the text, leading to more accurate classifications. When tested on unseen data, the Ensemble model maintained exemplary performance, demonstrating its robustness and adaptability in real-world scenarios. The strong performance on unseen data highlights the Ensemble model's proficiency in learning and applying complex patterns indicative of fake news.

The Ensemble model's superior accuracy, precision, recall, and F1 score, together with the comprehensive insights from the confusion matrix, validate its efficacy and dependability in detecting fake news. The Ensemble approach integrates the qualities of various models, alleviating individual shortcomings and improving overall performance.

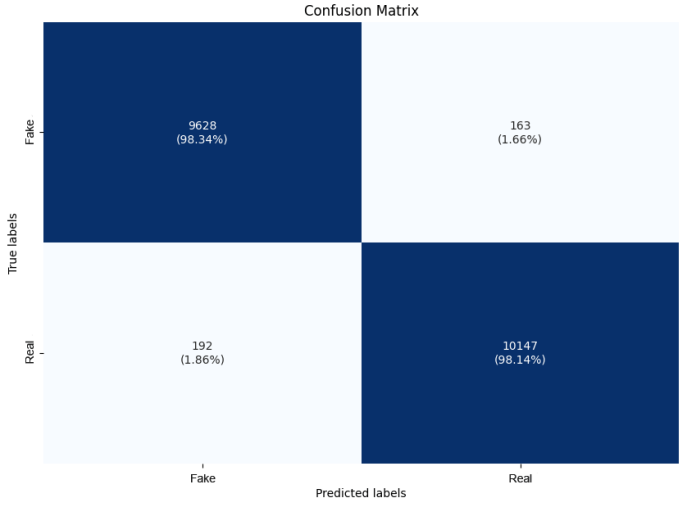

Fig.9. Ensemble model's confusion matrix

4.5. Comparison of Model Performance

The performance of four models - BERT, Bi-LSTM, CNN, and an Ensemble model - in detecting fake news is examined. The BERT model achieved high accuracy with a precision and recall value of 0.9796, indicating its effectiveness in identifying both real and fake news. However, it slightly underperforms compared to the Bi-LSTM and Ensemble models. The Bi-LSTM model showed the highest recall value of 0.9873, indicating its robustness in detecting

fake news. The CNN model also performed well, achieving high precision and recall values, indicating its effectiveness in identifying true positive instances of fake news. However, its slightly lower recall compared to Bi-LSTM suggests it may miss some instances of fake news

The Ensemble model, integrating the advantages of BERT, Bi-LSTM, and CNN, surpassed all individual models, achieving an accuracy of 0.9824, an F1-Score of 0.9828, a precision of 0.9842, and a recall of 0.9814. The ensemble approach efficiently reduces false positives and false negatives, hence enhancing the reliability of news tweet classification. This comparative analysis underscores the need of utilizing different models to attain enhanced performance in intricate tasks like fake news identification. The ensemble model exhibited superior results, underscoring the advantages of integrating many models to improve the accuracy and reliability of fake news detection systems. The results of these models show that integrating multiple models through an ensemble approach significantly enhances the accuracy and reliability of fake news detection, offering a more robust solution for real-time misinformation identification on social media platforms.

Table 1. Results of each model showing the comparison of each model according to the evaluation metrics

S/N	Models	Accuracy	F1-Score	Precision	Recall
1.	BERT Model	0.9796	0.9796	0.9796	0.9796
2.	Bi-LSTM	0.9805	0.9811	0.9750	0.9873
3.	CNN	0.9786	0.9791	0.9801	0.9781
4.	Ensemble	0.9824	0.9828	0.9842	0.9814

4.6. Comparison of the Model with Existing Methods

The ensemble model created in this study surpassed three existing fake news detection models from the literature, exhibiting higher performance across all criteria. The ensemble model attained the best accuracy of 0.9824, surpassing the results of [21], [15], and [20]. The ensemble model demonstrated exceptional performance in the F1-score, with a value of 0.9828, which signifies a harmonious balance between precision and recall. Precision evaluates the model's capacity to accurately identify fake tweets, with the ensemble model achieving the greatest precision of 0.9842. This signifies that the ensemble model possesses superior proficiency in mitigating false positives relative to the other models. Recall assesses the model's capacity to identify all pertinent false news tweets, with the ensemble model achieving the greatest recall of 0.9814. The model in [21] had a recall of 0.9000, markedly inferior to the model in [15] at 0.9781, and the model in [20] at 0.7467. In conclusion, the ensemble model exhibited enhanced performance across all criteria relative to current literature, highlighting its potential as a formidable option for addressing misinformation on social media platforms.

Table 2. Comparison of proposed model's results to existing works based on the evaluation metrics

S/N	Models	Accuracy	F1-Score	Precision	Recall
1.	The Proposed Model (Ensemble)	0.9824	0.9828	0.9842	0.9814
2.	Rai *et al*. (2022)	0.8875	0.9000	0.9100	0.9000
3.	Sharma *et al*. (2022)	0.9786	0.9791	0.9801	0.9781
4.	Alyoubi *et al*. (2023)	0.7467	0.7400	0.7490	0.7467

4.7. The Fake News Detection User Interface

The Gradio interface, a user-friendly library for building interactive interfaces for machine learning models, was used to create a user interface for detecting fake news. The platform enables users to submit news tweets and obtain predictions regarding their authenticity as real or fake.

The Gradio Blocks API was used to create a flexible and dynamic interface. The interface consists of a textbox for inputting the news tweet, a textbox for displaying the prediction, and an HTML element for providing a link to the Twitter bot account where the prediction can be posted. Users can input a tweet into the provided textbox and click the "Detect" button to get a real-time prediction.

The "Detect on Twitter" button leverages the Twitter API to post the prediction result on a designated Twitter bot account. This integration allows for broader dissemination of the fake news detection results, enhancing the system's utility and reach.

The Gradio interface bridges the gap between complex machine learning predictions and user-friendly interactions, making the fake news detection system accessible and practical for everyday use.

Fig.10. Interface of the deployed model

4.8. The Twitter Bot

The Twitter bot CANNBot was created utilizing the Twitter API and the Tweepy library to identify fake news. The bot uses a deep learning model to evaluate and forecast tweets, accepting the input language and producing predictions. The bot subsequently disseminates the results on Twitter, augmenting its real-time detection capabilities.

This is especially beneficial in combating the swift dissemination of disinformation on social media sites. The development approach entailed authenticating with the Twitter API, executing user interaction features, and including the deep learning model for real-time detection of fake news.

The CANNBot, an automated account, allows users to verify the veracity of news tweets, and the detection results are posted and shared on Twitter. This approach underscores the practical application of deep learning in addressing contemporary challenges like misinformation

Fig 11 shows the snapshot of the Twitter bot named CANNBot which is an automated account that enables interaction of the model developed with Twitter. It also enables users to verify the veracity of a news tweet and the results of the detection is posted and shared on Twitter so that the spread of fake news is mitigated before cause havoc on unsuspecting users of the platform. The system detects and updates on Twitter with an average time of less than 5 seconds.

Fig.11. Image of the twitter bot named CANNBot

5. Conclusions

This study effectively addressed the problem of fake news on Twitter by creating a Twitter bot capable of automatically identifying it. The hybrid strategy of incorporating BERT, Bi-LSTM, and CNN into an ensemble model outperformed standard models in detecting fake news. These algorithms combined contextual comprehension, sequence learning, and feature extraction to improve misinformation detection accuracy and reliability. The Twitter bot's real-time input on news validity was confirmed by its practical implementation.

The study also addressed automated fake news detection challenges and prospects. This paper establishes a framework for future advancements in this sector by examining model restrictions and establishing hybrid methodologies.

This research offers scalable solutions to fake news, particularly on social media, as it evolves. This research successfully integrates AI and deep learning to handle misinformation in the digital era, contributing to technical knowledge and practical applications.

To enhance the precision of fake news identification and contextual accuracy, Large Language Models (LLMs) should to be integrated with Retrieval-Augmented Generation (RAG). This would let systems analyse linguistic traits and cross-reference text with current external data to improve detection. A more complete approach would incorporate multilingual detection and computer vision for misrepresentation in photos, videos, and deepfakes. The model should also use continuous learning to adapt to misinformation patterns.

References

[1] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Information Processing and Management*, vol. 57, no. 2, pp. 102-125, 2020.

[2] B. Akinyemi, O. Adewusi, and A. Oyebade, "An improved classification model for fake news detection in social media," *Int. J. Inf. Technol. Comput. Sci.*, vol. 12, no. 1, pp. 34-43, 2020.

[3] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised fake news detection on social media: A generative approach," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 5644-5651, July 2019.

[4] V. Verma, M. Rohilla, A. Sharma, and M. Gupta, "Fake news detection on Twitter," in *Advances in Data and Information Sciences: Proceedings of ICDIS 2022*, Singapore: Springer Nature Singapore, 2022, pp. 141-149.

[5] A. Mitchell, "News on Twitter: Consumed by most users and trusted by many," *Pew Research Center*, Nov. 15, 2021. [Online]. Available: https://www.pewresearch.org/Journalism/2021/11/15/news-on-twitter-consumed-by-most-users-and-trusted-by-many/. [Accessed: Aug. 22, 2023].

[6] Pew Research Center, "Social media and news fact sheet," *Pew Research Center*, 2022. [Online]. Available: https://www.pewresearch.org/Journalism/fact-sheet/social-media-and-news-fact-sheet/. [Accessed: Oct. 20, 2023].

[7] S. Kemp, "Digital 2023: Nigeria – Datareportal – Global digital insights," *Datareportal*, 2023. [Online]. Available: https://datareportal.com/reports/digital-2023-nigeria. [Accessed: Jan. 19, 2024].

[8] R. J. Oentaryo, A. Murdopo, P. K. Prasetyo, and E. P. Lim, "On profiling bots in social media," in *Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I*, Springer International Publishing, 2016, pp. 92-109.

[9] T. Lokot and N. Diakopoulos, "News bots: Automating news and information dissemination on Twitter," *Digital Journalism*, vol. 4, no. 6, pp. 682-699, 2016.

[10] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017.

[11] Y. M. Rocha, G. A. De Moura, G. A. Desidério, C. H. De Oliveira, F. D. Lourenço, and L. D. de Figueiredo Nicolete, "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review," *Journal of Public Health*, pp. 1-10, 2021.

[12] K. Stahl, "Fake news detection in social media," California State University Stanislaus, 2018.

[13] V. L. Rubin, "Deception detection and rumor debunking for social media," in *The SAGE Handbook of Social Media Research Methods*, London: Sage, 2017, p. 342.

[14] N. I. Saputri, Y. Sibaroni, and S. S. Prasetiyowati, "Covid-19 fake news detection on Twitter based on author credibility using information gain and KNN methods," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 1, pp. 185-192, 2023.

[15] S. Sharma, M. Saraswat, and A. K. Dubey, "Fake news detection on Twitter," *International Journal of Web Information Systems*, vol. 18, no. 5/6, pp. 388-412, 2022.

[16] E. Amer, K. S. Kwak, and S. El-Sappagh, "Context-based fake news detection model relying on deep learning models," *Electronics*, vol. 11, no. 8, p. 1255, 2022.

[17] T. Ahmad, M. S. Faisal, A. Rizwan, R. Alkanhel, P. W. Khan, and A. Muthanna, "Efficient fake news detection mechanism using enhanced deep learning model," *Applied Sciences*, vol. 12, no. 3, pp. 17-43, 2022.

[18] P. K. Varshney, G. K. Wadhwani, and G. M. Upadhyay, "Systematic approach for fake news detection using machine learning," *NeuroQuantology*, vol. 20, no. 15, pp. 1819-1828, 2022.

[19] W. Y. Wang, "Liar, liar pants on fire: A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[20] S. Alyoubi, M. Kalkatawi, and F. Abukhodair, "The detection of fake news in Arabic tweets using deep learning," *Applied Sciences*, vol. 13, no. 14, p. 8209, 2023.

[21] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake news classification using transformer-based enhanced LSTM and BERT," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 98-105, 2022.

[22] Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). A comprehensive review on fake news detection with deep learning. *IEEE access*, *9*, 156151-156170.

[23] S. Dadkhah, X. Zhang, A. G. Weismann, A. Firouzi, and A. A. Ghorbani, "The largest social media ground-truth dataset for real/fake content: Truthseeker," *IEEE Transactions on Computational Social Systems*, 2023.

[24] Y. Verma, "Complete guide to bidirectional LSTM (with Python codes)," *Analytics India Magazine*, 2021. [Online]. Available: https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/. [Accessed: Feb. 9, 2023].

Authors' Profiles

Afeez Ayomide Olagunju is a postgraduate student of computer science at Obafemi Awolowo University, Ile-Ife, Nigeria, with a B.Sc. in Computer Science from Osun State University, Osogbo. His research focuses on online social networks, artificial intelligence, health informatics, augmented/virtual reality, geographic information systems, and gaming. He is a highly motivated researcher and computer professional who has experience in teaching, research and startup management.

Prof. I. O. Awoyelu is a Professor of Computer Science in the Department of Computer Science and Engineering, Faculty of Technology, Obafemi Awolowo University, Ile-Ife, Nigeria. She has over 20 years of teaching and research experience. Her research interests are in Data warehousing and Data Mining. She has successfully supervised 18 Masters and 5 PhD postgraduate theses. She has thirty-five publications to her credit. She is a member of Nigeria Computer Society (NCS), Computer Professionals of Nigeria (CPN) and Nigerian Women in Information Technology (NIWIIT).

How to cite this paper: Afeez Ayomide Olagunju, Iyabo Olukemi Awoyelu, "Performance Evaluation of Fake News Detection Models", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.6, pp.89-100, 2024. DOI:10.5815/ijitcs.2024.06.07