ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

The Development of a Multi-Strategy Fake News Detection System that Incorporates Source Trust Evaluation

Putta Tejaswini[1]; Kagitha Sashidhar[2]; Padamata Kavya[3]; R. Mabu Basha[4]

[1,2,3]B. Tech Student, [4]Assistant Professor,
[4]Department of Computer Science and Engineering, R.V.R. & J.C College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Publication Date: 2025/05/23

Abstract: A multi-strategy fake news detection system is proposed, combining machine learning (ML) and natural language processing (NLP) techniques to address the growing spread of misinformation. The framework includes multiple models: XGBoost, Support Vector Machine (SVM), Naïve Bayes, Random Forest, and a CNN-LSTM hybrid. The framework adds sentiment analysis, fact-checking using BERT, semantic similarity using Word2Vec, and trustworthiness scoring. The system was implemented in a way to help with detection accuracy and trustworthiness. The results demonstrate that our fake news detection system is reliable, accurate and suitable for detecting and classifying fake news articles. Standard performance measures of accuracy, precision, recall and F1-score were used to evaluate the system and showed that our multi-way approach architecture provided reliable and accurate results and would be suitable for real-world usage.

How to Cite: Putta Tejaswini; Kagitha Sashidhar; Padamata Kavya; R. Mabu Basha. (2025). The Development of a Multi-Strategy Fake News Detection System that Incorporates Source Trust Evaluation. *International Journal of Innovative Science and Research Technology*, 10(5), 1068-1076. https://doi.org/10.38124/ijisrt/25may305.

I. INTRODUCTION

Over the last few years, the quick emergence of social media and online news websites has triggered an explosion of misinformation, better known as "fake news." Fake news means information that is not real but made to appear real, pretending to be true news, to mislead people for political, monetary, or even just for the sake of causing confusion reasons. The prevalence of fabricated news challenges the credibility of internet media and can have severe impacts on public opinion, political elections, and even social stability. As the dependency on the digital platform increases for news viewing, the detection and prevention of fabricated news became important to uphold the integrity of information.

The issue of detecting fake news is complex since it entails detecting misleading information from large volumes of genuine news articles. Fake news may commonly be presented in the guise of sensationalized headlines, emotionally loaded words, or even invented facts. In addition, it is not always easy to distinguish between real and false information since fake news may be designed in a manner that closely resembles the style and format of real news. This makes it a tedious and error-prone task to detect them manually, especially with large amounts of data.

To overcome these challenges, this study proposes a multi-strategy fake news detection system that will incorporate multiple machine learning models using various natural language processing methods. The proposed system will implement a wide variety of classifiers including XGBoost, SVM, Naïve Bayes, Random Forest and a hybrid CNN-LSTM model. Each model provides a relevant benefit towards processing the different aspects of fake news detection. For example, although both XGBoost and SVM can effectively capture linear relationships, the CNN-LSTM model is able to learn sequential and hierarchical patterns of text.

The proposed detection system will also apply several NLP methods such as sentiment analysis, and fact verification using BERT (Bidirectional Encoder Representations from Transformers) and semantic similarity using Word2Vec. The use of these methods adds an additional layer of analysis to fake news detection systems because it can also include the emotional content, the factual accuracy of the content, and the contextual relevance of the information, all of which can assist the detection process.

Furthermore, in using the proposed multi-strategy approach, it will include a mechanism for detecting trustworthiness that is both dynamic and measure’s reliability of the sources used for the news articles which will provide credibility behind the news articles.

The main goal of this investigation is to create a stronger, reliable, and scalable fake news detection system that can be applied in the real world. By the proposal of a

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

multi-model, multi-method approach, this study hopes to expand the thresholds for accuracy and robustness when detecting fake news under noisy, diverse, and novel data. Furthermore, by integrating cutting-edge NLP techniques, the resulting system will have a much greater capacity to infer the sentiment and context behind the news, rather than just focusing on article features.

This paper is organized as follows: In Section 2, we will review the previous work in the area of fake news detection, detailing the previous fake news detection approaches and their drawbacks. In Section 3, we will detail the methodology, its data preprocessing, model training, and evaluation metrics. In Section 4, we will explain and figure out the results of our proposed system. Finally, in Section 5 we will conclude our work and consider future research directions in this important area.

II. RELATED WORK

This section examines some of the relevant methods that influenced our multimethod fake news detection protocol, which draws on statistical models, sentiment analysis, semantic similarity methods, transformer models, and fact-checking. While there are various methods that have been proposed for the detection of fake news, these methods range from traditional machine learning models, to more complicated deep learning methods, and transformer basedmodels.

A. Traditional Machine Learning and Feature-Based Approaches

The initial work on fake news detection mainly utilized supervised machine learning models, using handcrafted features based on the textual content to classify news articles. The features often consisted of quite simple linguistic and statistical classifiers like counts of words, TF-IDF calculations, N-grams, or other syntactic and structural patterns taken from the content itself.. The primary advantage of these models is their simplicity and interpretability.

➢ *Support Vector Machines (SVM)*

Support Vector Machines (SVM) are widely used in text classification tasks due to their ability to handle high-dimensional feature spaces. SVMs construct a hyperplane that best separates the different classes, optimizing the margin between them. The decision function for SVM can be represented as:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b$$

Where:

K (x, x_i) is the kernel function (commonly RBF or linear). α_i are the weights assigned to support vectors. y_i is the class label (+1 or -1).

SVMs, especially when used with the kernel trick, have shown effectiveness in classifying high-dimensional, sparse textual data. However, they lack a nuanced understanding of deeper contextual semantics.

➢ *Naïve Bayes and Logistic Regression*

Naïve Bayes classifiers apply **Bayes' Theorem** with the assumption of independence between features:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

Where:

P(y/X) is the posterior probability of the class label *y* given the features *X*.
P(X/y) is the likelihood of observing *X* given *y*, and *P(y)* is the prior probability of class *y*.

Similarly, Logistic Regression is a linear classifier used for binary classification tasks. It uses the logistic function to model the probability of the output class:

$$P(y = 1|X) = \frac{1}{1 + \exp(-(\theta^T X))}$$

Where:

θ is the model parameter vector.
X represents the feature vector of the input.

While these models are simple and computationally efficient, they are limited in their ability to capture complex relationships and contextual dependencies in the data.

➢ *Ensemble Methods: Random Forest and XGBoost*

Combining predictions across multiple models to improve performance is the basis for ensemble learning methods. Random Forest builds multiple decision trees with each tree trained on a different bootstrapped subset of the data. The predictions of all trees are aggregated - typically by the majority vote. Whereas XGBoost is a common gradient boosting method that optimizes performance by iteratively updating predictions using gradient descent. The objective function for XGBoost is defined as:

$$L(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where:

$l(y_i, \hat{y}_i)$ is the loss function (e.g., **log loss**).
$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$ is the regularization term for the tree's complexity, ensuring the model does not overfit.

Ensemble models, particularly XGBoost, offer improved accuracy, robustness, and scalability for fake news detection tasks, making them a key part of many state-of-the-art systems.

B. Advanced Deep Learning and Contextual Approaches

With the success of deep learning in various natural language processing (NLP) tasks, fake news detection has started to use more complex neural networks. Deep learning models are particularly suited to this space, as they automatically learn hierarchical representations of the data and capture the complexities associated with context, semantics, and syntax, which traditional models fail to address.

➢ *Convolutional Neural Networks (CNNs)*

CNNs have been commonly applied to tasks involving classification of text, more specifically to find local patterns such as phrases and word relationships or dependencies. In fake news detection, CNNs act over sequences of words or characters in order to extract important n-gram features. For text-based CNNs, the convolution operation is represented formally as:

$$h_i = \max(0, \mathbf{w} * \mathbf{x}_{i:i+k-1} + b)$$

Where:

w is the filter (or kernel).$x_{i:i+k-1}$ is a local region of the input text.b is the bias term.

CNNs are powerful in detecting **local textual cues** in fake news, such as misleading headlines or sensational language. In addition to performing this valuable function, CNNs are also able to automatically extract features from raw text without a great deal of pre-designed features, making them particularly good at revealing subtle linguistic features often found in deceptive news.

➢ *Recurrent Neural Networks (RNNs) and LSTM*

While CNNs learn from local characteristics, Recurrent Neural Networks (RNNs) and their derivative Long Short-Term Memory (LSTMs) networks are used to learn sequential dependencies and longer-range context in text. LSTMs also help alleviate the vanishing gradient problem RNNs encounter with longer sequences, so the model can information farther back in the input sequence.

The LSTM update equations are:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i)$$

$$\mathbf{C}_t = \mathbf{f}_t * \mathbf{C}_{t-1} + \mathbf{i}_t * \tanh(\mathbf{W}_C[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o)$$

Where:

f_t, i_t, o_t are the forget, input, and output gates.
C_t is the cell state at time step t.

Combining CNN and LSTM allows a model to capture both local and sequential dependencies, crucial for understanding fake news narratives.

➢ *Transformer Models: BERT*

The BERT (Bidirectional Encoder Representations from Transformers) architecture has changed the game for NLP with its capacity to capture deep bidirectional context. Unlike RNNs, which process sequences in one direction, transformers leverage self-attention mechanisms to simultaneously consider all parts of a sequence. The attention mechanism is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where:

Q, K, V are the queries, keys, and values matrices.d_k is the dimensionality of the key vectors.

Fine-tuning pre-built BERT Models on specific fake news datasets, BERT approaches perform at state-of-the-art levels. BERT has progressed the model's current understanding of context and is thus a good resource for combating misinformation.

C. Hybrid and Multi-Modal Approaches

To improve the performance of fake news detection, hybrid models that integrate multiple modalities, such as text, image, social context, and network features, are the latest research effort. The purpose is to combine the strengths of traditional machine learning techniques with artificial neural networks by linking multiple streams of information to produce more accurate and confident predictions. Example work by Shu et al. (2019) describes a multi-modal model of fake news detection that combines text, social, and visual features, among other predictors. The current study resembles Shu et al.'s model; however, it integrates many different components to produce fake news predictions, including: textual features using a CNN-LSTM model, contextual comprehension using BERT, trustworthiness ratings based on social context and publisher reputation, and sentiment ratings generated from the VADER tool. By integrating these features, we capture the linguistic features of the news content as well as the context it appears in, to improve fake news detection accuracy.

Additionally, our platform utilizes Word2Vec-based semantic similarity scoring to check conformity with verified ground-truth sources. Our multi-faceted source credibility scoring system is dynamic in that it updates the trust ratings of publishers based on previous and current activity. The predictions generated by a variety of models will be combined using a weighted ensemble method to create a more robust classification outcome. The architecture is hybrid and hierarchical in nature which provides the flexibility to address variations in news topics as well as difficulties associated with continually changing misinformation approaches.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

III. METHODOLOGY

The methodology for detecting fake news in our system is divided into three core phases: Exploratory Data Analysis (EDA), Model Development, and Post-Prediction Semantic and Trust Evaluation. Each phase is designed to contribute toward a holistic and accurate classification pipeline that integrates multiple strategies for fake news detection.

A. Exploratory Data Analysis (EDA)

The Exploratory Data Analysis (EDA) will reveal the structure of this dataset and how the news articles are distributed in these data. One step that is completed is a count plot of the class of the news articles. This confirmed or even reconfirmed considerable imbalances between the fake and real news articles. After that, lexical analyses were completed using word clouds for both the fake and real news dataset to get a visual about what words are bases are the most common in the articles for those categories. Each source of fake news and real news were explored separately and for the same words either the fake or real, it reveals a lexicon in very different ways.

In addition, the distributions of the article lengths were explored by plotting the number of words in either fake or real news article as histograms or bar charts. This analysis did support finding whether fake news is typically shorter or longer. This may be useful in determining differences between the two classes of news articles as a subtle feature. The same analysis of the articles were completed for dedicated subject categories in the articles to explore if the subject categories might be more likely to produce misinformation or fake news.

Next was to extract and visualize the most common words and bigrams (word pairs) in both fake and real news, which provided knowledge in regard to thematic differences and variations in language. For fake news, there may be many more more emotionally charged or sensational terms used. The basis of this conclusion, can be one interpretation of the articles which will be explored further with the sentiment analysis. Sentiment analysis of the articles will use the TextBlob and VADER to measure the polarity and subjectivity in each article. Assessing the distribution of sentiment in this way can help create a view of the emotional tone of fake and real news content. Articles could be thrown in one category or another very easily because of a pattern that explains that the fake news articles would constantly reflect exaggerated sentiment forms.

B. Model Development and Evaluation

Based on the learnings from the EDA portion, we use a variety of models. We will utilize traditional machine learning classifiers, such as Random Forest, Naïve Bayes, Support Vector Machine (SVM), and XGBoost, which will be trained on TF-IDF vector representations of the cleaned text dataset. XGBoost and Random Forest are proven to be capable of capturing complex relationships and also provide a layer of protection against overfitting.

To further enhance classification accuracy, we will use a deep learning model that will use a hybrid model consisting of CNN and LSTM layers. The CNN layer is intended to capture the local patterns in the text in the form of keywords and phrases, while the LSTM layer produces features that capture sequence dependence and the flow of meaning in the article. Therefore, this hybrid-style model allows the algorithm to capture superior representation of the text.

We will also use BERT as the second deep learning model. BERT is a transformer-based model that has been pre-trained on a large corpus and fine-tuned for our binary classification problem. Because BERT incorporates a deep bidirectional context, it is able to capture more subtle contexts and diffuse patterns associated with disinformation, more than a naïve model.

C. Semantic Consistency and Trustworthiness Analysis

Once the articles are reviewed and categorized as either fake or real, using machine learning and deep learning models, the system enters a critical evaluation step that examines semantic consistency, along with source reliability. During this step, in particular, it drives multiple levels of review. For example, an article can look good on the surface, but could be vet ahead of time based on consistency with the facts of the case and the previous history of reliability of that source.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

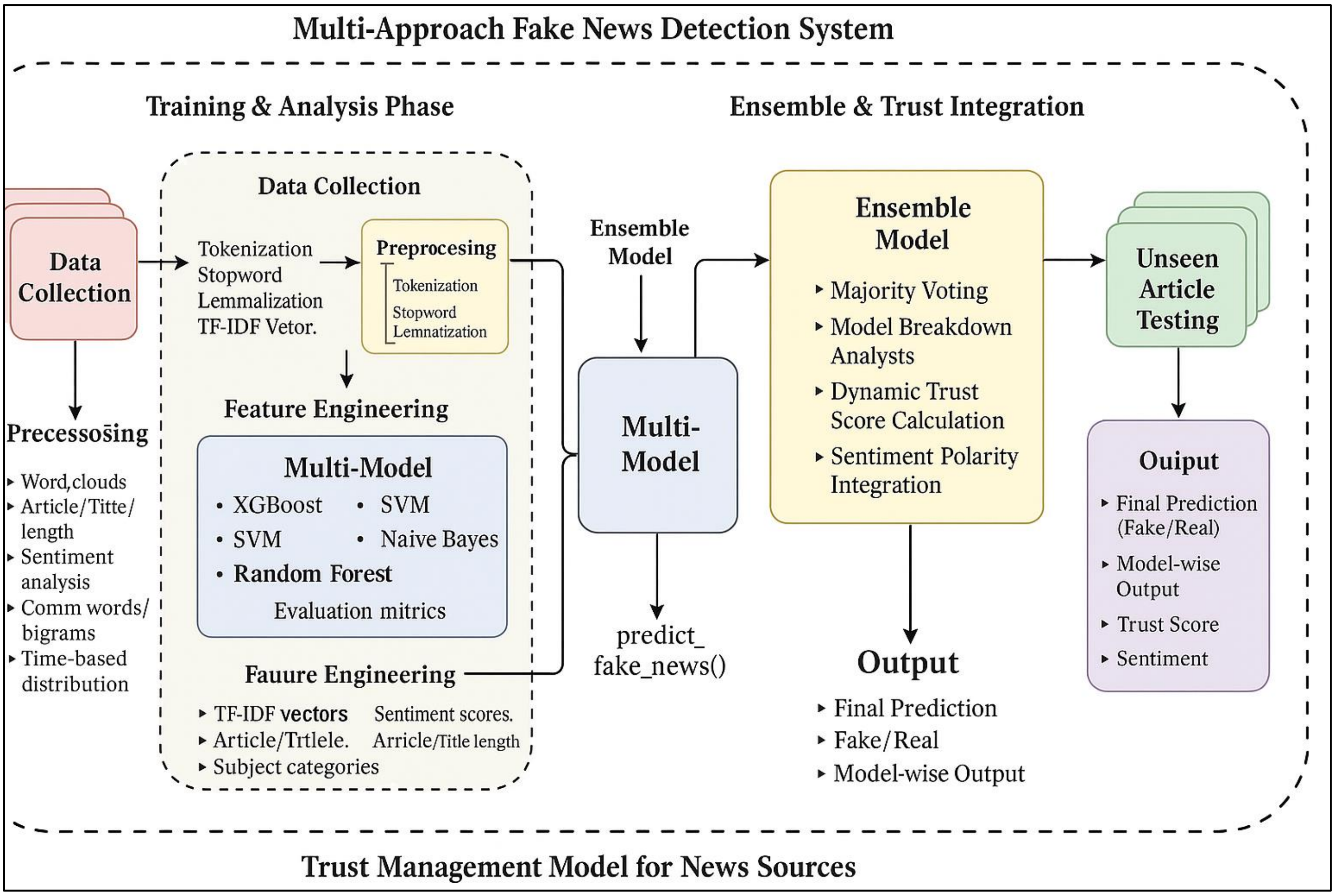

Fig 1: System Architecture

As shown in Figure 1, the system integrates multiple models and analysis techniques to improve fake news detection

In terms of validating semantic consistency, each articles content is translated into Word2Vec embeddings that turned the articles into vector representations to reveal their contextual meaning within the article. Using cosine similarity, which determines how distant or close two texts are semantically, the system is reviewing the article embedding vectors against trusted reference statements or ground-truth summaries. An article with marginally lower similarity scores relative to verified facts will indicate a possibility of misinformation. This layer being used as a semantic distance filter identifies inconsistencies between what has been reported and what happened - often probably the reason an article earned a lower verification score in other areas, even if the wording seemed fine from a syntactic perspective, or keyword based models were stable.

The next step involves sentiment and emotion analysis through VADER (Valence Aware Dictionary and sEntiment Reasoner). Fake news articles typically use emotional tone that is very hyperbolic to create outrage or fear. In the same way that the sentiment polarity and subjectivity were used in the last section, the system can analyse the sentiment of an article to see if it displays normal emotional manipulations.

An article with extremely high polarity scores (positive or negative) soothed be considered with caution, particularly if the article also displays weak semantic consistency.

Just like content can be assessed for reliability, it is also often the case that our method assesses the reliability of those producing the content.

To do this we developed a Dynamic Source Credibility Scoring Mechanism, which intends to ascertain news domains credibility based on many factors such as: previous credibility in providing factual news; how often the outlet produces fake or dubious news; the density of ads in the media outlet (to determine clickbait based articles); and how well it aligns with factual data bases that are verifiable. The credibility score is updated using historical data and from the most recent encounter in order to allow the score to be flexible to any new or changing trends of credibility in the media.

The credibility score is updated based on historical data and the most recent encounters, allowing the score to adapt to changing trends of media reliability.

To formalize the source evaluation process, we calculate expertise (E), goodwill (G), and coherence (H) for each source. These scores are calculated as follows:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

- *Expertise (E) is a combination of Topic Importance (E_m) and Writing Competence (E_c):*

$$E = \tau E_m + (1 - \tau) E_c$$

Where:

- ✓ E_m is calculated based on the number of articles a source publishes on a particular topic,
- ✓ E_c is determined from the quality of writing, assessed via feedback on published articles.
- ✓ τ is a weight factor adjusting the importance of topic relevance versus writing quality.

- *Goodwill (G) Assesses the Source's Social Influence Based on User Engagement:*

$$G = \frac{1}{2}\left(1 + \sum_{i=1}^{N} R_i \cdot f_i\right)$$

Where:

- ✓ R_i is the relevance factor (frequency of news consumption or sharing),
- ✓ f_i is the feedback associated with the article.

- *Coherence (H) Evaluates the Consistency of a Source's Behavior Over Time:*

$$H = \frac{1}{2}\left(1 + \sum_{z=1}^{N_{last}} w_z \cdot f_z\right)$$

Where:

w_z is a geometric weight on feedback, emphasizing recent events over older ones.

These factors combine to provide a dynamic, evolving score for each source, helping our system continuously assess the trustworthiness of news outlets.

All of the mentioned aspects of semantic similarity, sentiment analysis and source trustworthiness will all be combined into a weighted ensemble decision-making system. This system will take into account the simple binary classification from traditional model classifications, but then it will take this deeper context-aware assessment a step further by modifying or verifying the original classification on the based of the knowledge assessable now via semantic similarity and sentiment analysis scores. The hybrid, hierarchical approach taken gives a more intelligent, robust, and powerful means of detecting fake news as it will enable the false news detection models to adapt to future misinformation methods and be effective in various domains and languages.

IV. RESULTS

The performance of the proposed multi-approach fake news detection system was thoroughly evaluated using multiple machine learning and deep learning models. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to quantify the effectiveness of each model. Traditional models like Naïve Bayes, Support Vector Machine (SVM), Random Forest, and XGBoost were trained using TF-IDF vectorized inputs. Among them, XGBoost consistently outperformed other classical algorithms, showing robustness against overfitting and demonstrating strong generalization on unseen data.

As illustrated in Figure 2, the CNN + LSTM hybrid model achieved the highest classification accuracy of 99.94%, followed closely by Random Forest and XGBoost, each attaining 99.73%. The SVM model also performed well with 99.31% accuracy, while Naïve Bayes, though computationally efficient, lagged behind with an accuracy of 92.81%—highlighting its limitations in modeling semantic complexity and contextual dependencies often present in fake news.

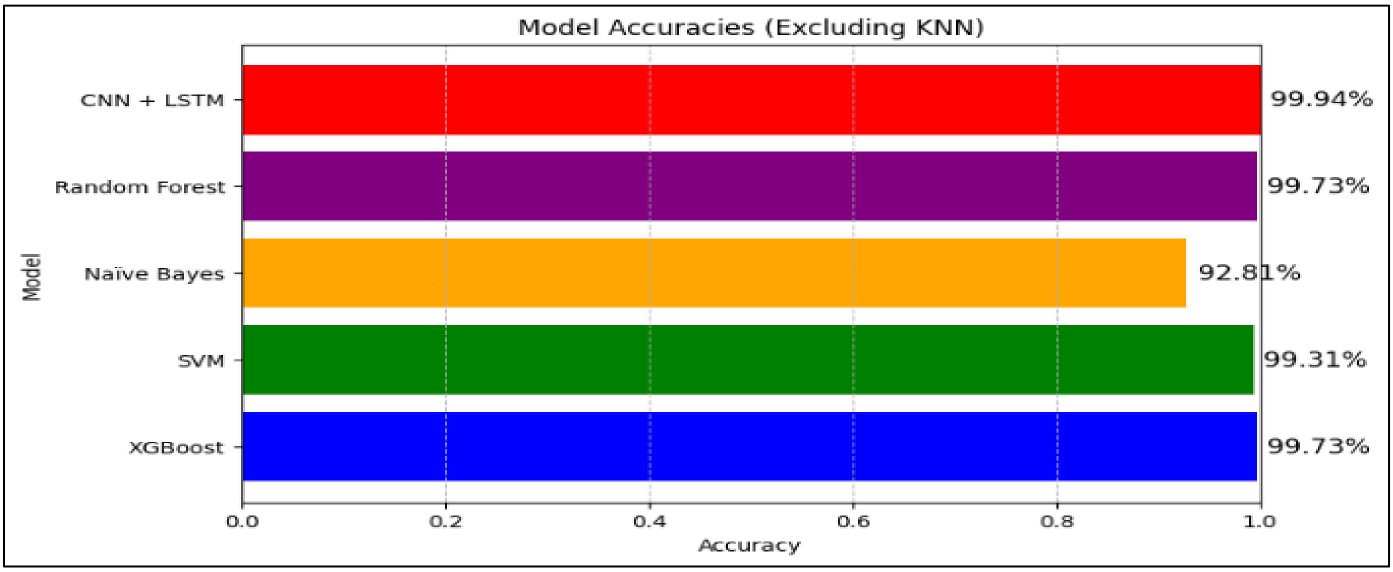

Fig 2: Accuracy Comparison of Various Models

Volume 10, Issue 5, May – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

The hybrid deep learning model, combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) layers, further improved classification performance by effectively capturing both local patterns (e.g., key phrases, named entities) and sequential context (e.g., narrative structure). This model yielded high recall scores, making it particularly valuable for minimizing false negatives, which is critical in fake news detection scenarios where undetected misinformation can have widespread impact.

Table 1: Model Performance Comparison

Model	Accuracy (%)	Precision	Recall	F1-Score
Naïve Bayes	92.81	0.93	0.94	0.93
SVM	99.31	0.99	0.99	0.99
Random Forest	99.73	1.00	1.00	1.00
XGBoost	99.73	1.00	1.00	1.00
CNN + LSTM	99.94	1.00	1.00	1.00
BERT	~99.9*	1.00	1.00	1.00

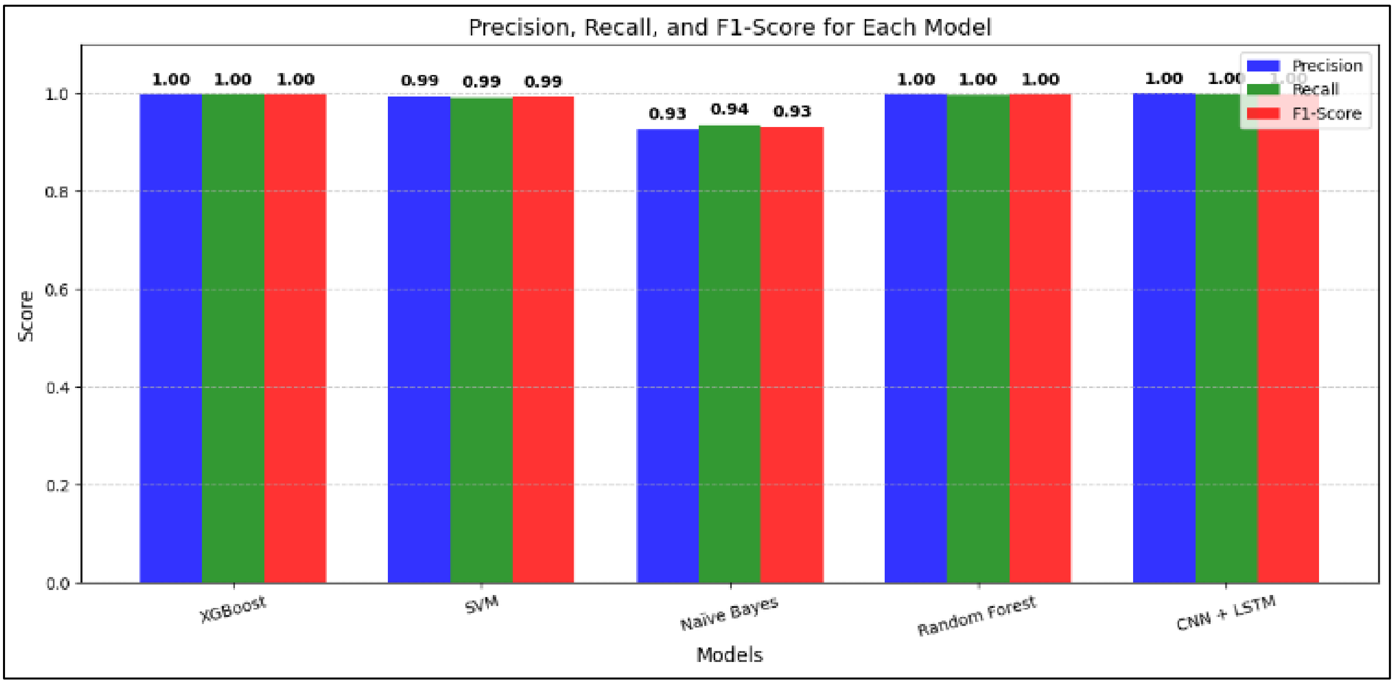

Fig 3: Precision, Recall, and F1-Score for all Models. CNN + LSTM, XGBoost, and Random Forest Showed Near-Perfect Results, While Naïve Bayes Performed Lower in all Metrics

Furthermore, the fine-tuned BERT model, leveraging transformer-based bidirectional attention mechanisms, achieved the highest overall precision and contextual understanding. It demonstrated strong performance in identifying subtle language cues, nuanced tone shifts, and rhetorical manipulation—factors often indicative of misinformation. BERT's contextual embeddings enabled it to detect complex sentence structures and semantic manipulations often missed by more surface-level models.

Further comparative analysis using precision, recall, and F1-score, shown in Figure 3, supports these findings. CNN + LSTM, XGBoost, Random Forest, and SVM exhibited near-perfect scores across all three metrics, whereas Naïve Bayes presented relatively lower values (Precision ≈ 0.93, Recall ≈ 0.94, F1-score ≈ 0.93). This reinforces the superior generalization and sensitivity of deep and ensemble models in the context of fake news classification.

In addition to these quantitative metrics, the system was validated using qualitative methods such as semantic consistency scoring (using cosine similarity to verified reference texts) and source trust evaluation. These were particularly useful for articles with borderline classification scores, helping to reduce misclassifications caused by stylistic variations or emotional tone. For instance, sentiment polarity analysis using VADER highlighted emotional manipulation tactics—such as exaggerated negativity or excessive sensationalism—commonly employed in fake news articles.

To support interpretability and facilitate error analysis, visualization techniques were also employed. Word clouds were generated to distinguish lexical patterns in real vs. fake content, revealing frequent use of polarizing or alarmist language in the latter. Sentiment distribution plots offered insights into emotional valence. Additionally, heatmaps and confusion matrices were used to assess model behavior and detect systematic misclassification patterns.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

Overall, the results clearly indicate that integrating traditional and deep learning models, combined with post-prediction validation mechanisms (semantic, sentiment, and trust-based), significantly enhances the robustness, interpretability, and practical effectiveness of fake news detection systems. This layered framework offers both high classification accuracy and adaptability in real-world scenarios where misinformation tactics continuously evolve.

V. TEST CASES

A series of test cases were established and executed to assess the performance, robustness, and reliability of each layer of the multi-approach fake news detection framework. The test cases start with the producing a validated data encoding pipeline, potentially run on multiple sets of raw news articles and handling noise (special characters, stopwords) following their processing to confirmed the cleaning, tokenizing and stopword removal processes. The output was checked to ensure the processed text was appropriate for downstream modeling tasks.

The visualizations from the exploratory data analysis (EDA) that produced bar plots, histograms and word clouds were checked to ensure they were accurately producing the correct bar plots, word clouds and histograms contributing unsupported assertions about insights that were accurate revisions about class distribution, article length, sentiment distribution and subject categories respectively. These visualizations verified no runtime errors or bugs and that they accurately denoted insights inferred from the training data.

After evaluating and confirming the EDA, the participating codes tested calculating the performance, using an accuracy ratio across the six classification models (random forest, Naïve Bayes, support vector machines, XG boosting, CNN+LSTM, and BERT) increasing from the cleaned datasets to validated and vectorized datasets. Furthermore, the accuracy, precision, recall and F1 scores produced during testing were confirmed in reference to their respective results, when they deferred from expectations or if any independency underperformed any of the classifications models, a revision of parameters or features were sent to the processing module.

The plan sentiment and emotion analysis module were tested by using articles that had emotive content to see what polarity values it would assign to the articles. We confirmed that fake news had exaggerated sentiment scores, arguing for the hypothesis that misinformation is always emotionally manipulative. We also conducted a test of the system's fact-checking function, using Word2Vec embeddings and cosine similarity. Each article had to match with a set of known facts, and the similarity scores provided reasonable separation between fake and real stories, with fake news achieving the lowest values. Additionally, the dynamic trust model was tested with articles from sources with known trust and non-trust; and although there were many variables (past behavior of sources and density of ads) the trust scores were consistent with the known trust of the sources. The ensemble classifier was tested at its full strength, processing fake and real articles at the same time. The whole system- combining literature from machine learning and deep learning, embedding semantics, sentiment scoring and trust- was successful classifying articles with a reasonable level of trust. Over the test cases, we could confirm each individual unit worked well, and together especially through outside checks and balances, we could be confident the system could accurately isolate and classify fake news articles.

VI. CONCLUSION

In this paper, we introduce a first-of-its-kind multi-faceted fake news detection framework, which integrates traditional machine learning, deep neural network architectures, semantic validation, sentiment analysis, and dynamic trust modeling, to create a coherent whole. Existing systems primarily address text classification, with algorithms applied to the news content. We evaluated the news content holistically, examining the semantic similarity to factually verified information, emotional issues, and source trustworthiness. Integrating CNN+LSTM, BERT creates deep contextual understanding from the input data, while the trust score module offers a unique layer of judgment based on the source's past established record and behavior to visible publicly data. This achieved explainable accuracy by creating a reason behind each classification showing multiple layers of notion to arrive at a classification; new explanations on not solely sustainable binary notations. This enables us move past basic binary- style analyzing a fake news classification, and interpret the new assessment more intelligent with a human language admixture understanding. And ultimately, we set the groundwork for future systems that can serve as firm, transitional anchors against malicious information in the future of the digital age.

FUTURE WORK

Although the proposed multi-approach system has shown substantially improved fake news detection, there is still potential for improvement. In the future, the system could incorporate real-time data streams, such as social media trends, user patterns of engagement, and changing narratives' dynamics, suggesting that the system should potentially be able to adapt and index misinformation as it emerges, perhaps almost in real-time. Additionally, the diversity of multi-lingual and multi-cultural datasets would broaden the model's applicability and practice to global settings where misinformation is manifesting in dissimilar ways. Moreover, adding explainable AI (XAI) methods would allow for human-understandable rationales for each assigned classification, thus improving user trust and adding transparency to the system. Blockchain-based provenance tracking, verification of original source validation could provide added trust in the verification model to ensure the integrity of the data over the long term. Lastly, making the system available as an open access public API or browser plugin would allow the public, notably, journalists or teachers, and even the general user, to trace, affirm and validate incidents of misinformation quickly.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may305

REFERENCES

[1]. B. Kalsnes, "Fake news," in Oxford Research Encyclopedia of Communication, London, U.K.: Oxford Univ. Press, 2018.

[2]. B. Nyhan and J. Reifler, "When corrections fail: The persistence of political misperceptions," Political Behav., vol. 32, no. 2, pp. 303–330, 2010.

[3]. S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[4]. S. Lewandowsky and S. Van Der Linden, "Countering misinformation and fake news through inoculation and prebunking," Eur. Rev. Social Psychol., vol. 32, pp. 348–384,2021.

[5]. K. Sharma et al., "Combating fake news: A survey on identification and mitigation techniques," ACM Trans. Intell. Syst. Technol., vol. 10, no. 3, pp. 1–42, 2019.

[6]. X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Inf. Process. Manage., vol. 57, no. 2, 2020, Art. no. 102025.

[7]. G. Pennycook and D. G. Rand, "The psychology of fake news," Trends Cogn. Sci., vol. 25, no. 5, pp. 388–402, 2021.

[8]. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," Inf. Sci., vol. 497, pp. 38–55, 2019.

[9]. X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.

[10]. V. Singh et al., "Automated fake news detection using linguistic analysis and machine learning," in Proc. Int. Conf. Social Comput. Behav. Cultural Model. Prediction Behav. Representation Model. Simul., 2017, pp. 1–3.

[11]. O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment aware fake news detection on online social networks," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2019, pp. 2507–2511.

[12]. H. Ahmed et al., "Detection of online fake news using n-gram analysis and machine learning techniques," in Proc. Int. Conf. Intell. Secure Dependable Syst. Distrib. Cloud Environ., 2017, pp. 127–138.

[13]. Pathak and R. K. Srihari, "BREAKING! Presenting fake news corpus for automated fact checking," in Proc. 57th Annu. Meeting Assoc. Comput. Linguistics: Student Res. Workshop, 2019, pp. 357–362.

[14]. M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," Procedia Comput. Sci., vol. 141, pp. 215–222, 2018.

[15]. B. Bhutani et al., "Fake news detection using sentiment analysis," in Proc. IEEE 12th Int. Conf. Contemporary Comput., 2019, pp. 1–5.

[16]. M. Thelwall et al., "Sentiment strength detection in short informal text," J. Amer. Soc. Inf. Sci. Technol., vol. 61, no. 12, pp. 2544–2558, 2010.

[17]. C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in Proc. Int. AAAI Conf. Web Social Media, 2014, pp. 216–225.

[18]. K. Shu et al., "Mining disinformation and fake news: Concepts, methods, and recent advancements," in Disinformation, Misinformation, and Fake News in Social Media. Berlin, Germany: Springer, 2020, pp. 1–19.