Fake Digital News Detection Using Machine Learning and Natural Language Processing

Ranjitha R[1], Bhoomika J Shetty[1], Yogish Naik G R[2]

[1] Lecturer, Sahyadri Science College, Kuvempu University, Shivamogga, Karnataka

[2] Professor, Department of computer science and MCA, Kuvempu University, Shivamogga, Karnataka

Abstract:

The way information is created, shared, used is completely changed with the advent of the quick development in digital communication technology along with online social networking sites. These technological developments have greatly increased access to information worldwide, but they have also sped up the distribution of false information and fake news via blogs, social media, and online news portals. By influencing public opinion, eroding confidence in reliable media sources, causing political instability, inciting public panic during crises, and adversely affecting social and economic well-being, the quick dissemination of false information presents significant problems to society. As a result, the automatic identification of false news has become a crucial research issue in artificial intelligence, machine learning, natural language processing domains.

With the massive volume and rapidity of online information, traditional manual fact-checking techniques are no longer adequate. Because of this, academics are concentrating more on creating clever automated systems that can accurately differentiate between real news and fake information. By identifying significant linguistic patterns and semantic elements in news items, the natural language processing techniques allow the computers to comprehend, interpret, analyze textual data. These resulting qualities can then be used by machine deep learning algorithms and machine to perform accurate text classification.

This article offers a thorough framework for detecting false news combining several machine learning and deep learning models with sophisticated NLP preprocessing approaches. Data preprocessing, tokenization, stop word removal, stemming, feature extraction using Term Frequency Inverse Document Frequency and word embedding, and classification using Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine, XG Boost, and Long Short-Term Memory networks are all part of the suggested methodology. These models' performance is assessed using commonly used measures, such as accuracy, precision, recall, and F1score. By successfully capturing contextual and sequential linkages within textual data, comparative research shows that deep learning approaches particularly LSTM outperform conventional machine learning methods. The results verify that a dependable and scalable method for identifying fake news and reducing the dissemination of false information across digital media platforms is to combine natural language processing with sophisticated machine learning algorithms.

Keywords:

Machine Learning, Natural Language Processing, Deep Learning, Text Classification, Feature Extraction, TF-IDF, LSTM, XG Boost, Support Vector Machine, Artificial Intelligence, Fake digital news detection.

1. Introduction:

The extensive use of digital communication technology over the internet has drastically changed how information is produced, shared, and consumed. blogs, online news portals, Social media platforms, and digital broadcasting services have done it possible for people all over the world to instantaneously access and exchange information. These technological advancements have enhanced communication and made information more accessible, but they have also made it possible for incorrect and misleading information to proliferate quickly. One of the biggest problems in the digital age is the unchecked spread of false information, which has an impact on people, businesses, governments, and society at large.

The term "fake news" describes information that has been purposefully created, altered, or misrepresented and published as real news with the intention of misleading readers. Such content is frequently produced with the intention of influencing public opinion, influencing governmental decisions, generating financial gains through online interaction, advancing ideological goals, or inciting social unrest. In contrast to unintentional false information, fake news is purposefully made to look authentic, making it difficult for readers to distinguish between false and real content. Fake digital news has spread more quickly during significant world events, such as elections, pandemics, natural catastrophes, and foreign conflicts. This has led to uncertainty, fear, and a decline in public confidence in reliable information sources.

This issue has been made worse by the growing reliance on digital media for news consumption. Information may quickly reach millions of users through social networking sites like Facebook, X (previously Twitter), Instagram, YouTube, and other online communication methods. Their recommendation algorithms often prioritize highly engaging content, regardless of its authenticity, thereby accelerating the spread of misinformation. Consequently, false news frequently propagates faster than verified information, making manual verification increasingly difficult and time-consuming.

The consequences of fake news extend beyond misleading individual readers. It can influence electoral outcomes, damage the reputation of individuals and organizations, create financial market instability, encourage hate speech, and negatively affect public health by spreading incorrect medical information. The COVID-19 pandemic demonstrated how misinformation regarding vaccines, treatments, and preventive measures could significantly hinder public health initiatives. These difficulties show how urgently sophisticated systems that can recognize false information before it spreads to a larger audience are needed.

Conventional fact-checking methods mostly depend on human specialists who manually confirm information by consulting trustworthy sources. Despite their great accuracy, these techniques are expensive+, labor-intensive, and unable to manage the massive amount of digital content produced on a daily basis. Manual verification is no longer a viable approach given the millions of news stories, blog entries, and social media updates that are published every day. Due to this result, automated systems for detecting fake digital news have emerged as crucial field of study in artificial intelligence.

Natural language processing, a branch of artificial intelligence offers computational methods that let robots comprehend, interpret, and analyze human language. Linguistic patterns, contextual connections, writing styles, semantic structures, and grammatical features can all be found in textual information using NLP techniques.

By examining the textual characteristics of news stories, these elements enable automated

systems to discern between authentic and fraudulent ones. Tokenization, lemmatization, text normalization, stemming, stop-word removal and feature extraction are common NLP preparation methods that enhance textual data quality prior to classification.

By identifying hidden patterns in previously labeled datasets, machine learning algorithms improve the identification of fake news. Machine learning models automatically detect discriminative textual characteristics that distinguish authentic news from fake news, as opposed to depending on manually created rules. Because they can analyze high-dimensional textual features created by methods like Term Frequency–Inverse Document Frequency, traditional supervised learning algorithms like Naive Bayesian, Decision tree, Logistic regression, Random forest, Support vector machine have shown strong performance in text classification problems.

The accuracy of detecting fake news has been substantially enhanced by recent developments in deep learning. Contextual dependencies and sequential linkages between words in news articles are efficiently captured by deep learning model, Long Short-Term Memory networks(LSTM). LSTM models automatically learn semantic representations from textual sequences through word embedding, in contrast to traditional machine learning algorithm that rely on manually extracted features. This improves classification performance and allows for a better understanding of linguistic context. These capabilities make deep learning particularly suitable for identifying subtle linguistic variations commonly observed in deceptive news articles.

The core mission of this research is to combine different deep learning and machine learning algorithms with natural language processing techniques to create an efficient and trustworthy false news detection system. Comprehensive text preprocessing, feature extraction using TF IDF and word embedding, model training, and comparative performance validation using Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Long Short Term Memory, XG Boost are all carried out by the suggested methodology. Widely recognized evaluation criteria, including as accuracy, precision, recall, and F1 score, are used to evaluate each model's efficacy.

By highlighting the pros and cons of different categorization algorithms, the findings of this study add to the developing corpus of knowledge in automated fake digital news identification. In order to lessen the dissemination of false information and improve the reliability of digital information, the suggested method can assist media companies, social media platforms, governmental organizations, fact-checking groups, and researchers. In order to create scalable, accurate, and real time fake news detection systems that can tackle among the most urgent issues in today's information driven society, the study also highlights the significance of fusing cutting edge Natural Language Processing techniques with contemporary machine learning approaches.

2. Literature Review:

The volume of online information has greatly expanded due to the quick development of social networking sites and digital media, which has made it simpler for people to access and share news in real time.

But this unparalleled accessibility has also made it easier for false information and fake news to proliferate widely. Researchers have focused a lot of effort on creating automatic fake news detection systems utilizing Machine Learning, Natural Language Processing and Deep Learning since false information can spread far more quickly than confirmed news. Many studies have investigated various computational methods to enhance the precision, effectiveness, and scalability of false news detection algorithms within the last ten years.

Shu et al. (2017) conducted one of the most thorough studies in this area and suggested that linguistic content should not be the only factor used to detect fake news. Instead, they suggested analyzing fake news from three complementary perspectives: news content, social context, and news propagation patterns. News content focuses on the linguistic characteristics of the article, including writing style, vocabulary, syntax, and semantic information. Social context examines user interactions such as comments, likes, shares, and user credibility, while propagation analysis studies how information spreads across online social networks. Their research demonstrated that integrating these multiple sources of information substantially improves the effectiveness of fake news detection systems compared with methods based only on textual analysis.

Horne and Adalı (2017) claim that fake news items have unique language traits that set them apart from real journalism. According to their research, inflated headlines, emotionally charged language, sensational expressions, repetitious terminology, and simplified sentence structures are frequently used in fake news items in an effort to grab readers' attention and promote online interaction.

The LIAR dataset, a benchmark dataset for identifying fake news, was presented by Wang, W. Y. (2017). The dataset includes brief comments that vary in their reliability. A helpful dataset for training and assessing machine learning models for the detection of false news was made available by the study.

Puri (2024) highlighted the growing problem of false information on digital and social media platforms in his thorough analysis of contemporary fake news detecting methods. In addition to natural language processing methods like tokenization, stemming, lemmatization, TF-IDF, stop words, and word embedding, the study looked at conventional machine learning algorithms like Logistic Regression, Naïve Bayes, Support Vector Machine, Random Forest and Decision Tree.

Reddy, O. (2024) introduced a Natural Language Processing based artificial intelligence system for detecting bogus news. In order to detect fake news from textual data, the study used AI and NLP algorithms.

Randhe, K., Vyavahare, N., Vishwakarma, R., and Kudale, S.(2025) suggested a Natural Language Processing (NLP)-based fake news identification system. The study processed news text and categorized it as authentic or fraudulent using NLP techniques.

In order to identify and categorize news articles as authentic or fraudulent based on their textual content, Sasirekha and Bhuvaneshwari (2025) suggested a false news detection system that uses Natural Language Processing. Important NLP procedures such text preprocessing, feature extraction, and the use of machine learning and deep learning methods to examine language patterns and contextual data were all covered in the study.

Ahire, P. S., Lokhande, N. M., and Gadekar, M. J. (2026) presented a deep learning and natural language processing-based fake news identifying system. In order to categorize news as authentic or fraudulent, the study concentrated on pre-processing news text and using deep learning models. The scientists came to the conclusion that integrating NLP and deep learning enhances the efficacy of identifying incorrect information and increases the accuracy of detecting fake news.

Due to their ease of use, good classification performance, and computational efficiency, a number of conventional machine learning methods that have been extensively used for the detection of fake news. One of the first probabilistic classifiers used for text classification applications is Naïve

Bayes. It is based on assumption that textual features are independent in order to estimate the likelihood of each class. Naïve Bayes works quite well for many document classification issues due to its computational efficiency and capacity to handle high-dimensional feature spaces, even if this assumption rarely holds for natural language data.

Additionally, logistic regression has been widely used because to its interpretability and robustness. Using weighted textual features, it models the likelihood that a news story falls into the true or phony category. Combining Logistic Regression with feature extraction techniques like Term Frequency Inverse Document Frequency has consistently resulted in dependable classification performance with comparatively low computing complexity.

The Support Vector Machine, another popular classifier, has shown remarkable success in high-dimensional text classification tasks. SVM finds the best decision boundary to maximize the distance between several classes. SVM successfully manages the thousands of sparse features that textual datasets often contain, and it has repeatedly produced excellent accuracy in research on the detection of digital fake news.

Classification performance has been further enhanced by ensemble learning techniques, which combine multiple weak learners to create a strong predictive model. Through bootstrap aggregation and random feature selection, Random Forest, an ensemble of decision trees, lessens overfitting, improving classification stability and generalization. Similar to this, XG Boost uses regularization procedures to avoid overfitting together with gradient boosting approaches that iteratively reduce prediction errors. One of the most effective algorithms for a variety of classification applications, including the identification of false news, is XG Boost because of its capacity to model intricate nonlinear relationships within textual data.

Natural language processing applications have seen a considerable transformation thanks to recent development in deep learning, which makes it possible to automatically learn features from raw textual input. Deep learning models learn semantic and contextual representations straight from text, in contrast to the conventional machine learning algorithms that need manually created features. The Long Short Term Memory network, a specific kind of Recurrent Neural Network intended to capture long-term dependencies within sequential input, is one of the most popular architectures.

LSTM models use memory cells and gating algorithms that selectively preserve or discard information over time to solve the vanishing gradient issue that is frequently present in traditional recurrent neural networks. This capability enables LSTM networks to understand contextual relationships among words, making them particularly effective for fake news detection, where the meaning of a sentence often depends on long range word dependencies rather than isolated keywords. Consequently, numerous studies have reported that LSTM models outperform conventional machine learning algorithms when sufficiently large training datasets are available.

The Kaggle Fake News Dataset and FakeNewsNet, in addition to the LIAR dataset, have gained popularity as benchmark datasets due to their bigger collections of news stories that include entire headlines, article bodies, author details, publication sources, and believability labels. Deep learning and machine learning models can learn more typical linguistic patterns from these datasets richer contextual information, increasing detection accuracy.

Recognizing the limitations of relying solely on textual content, Shu et al. (2019) extended earlier research by incorporating user behavior, source credibility, publisher reputation, and information propagation patterns into fake news detection models. Their findings demonstrated that multimodal detection frameworks generally outperform text only approaches because they capture

additional contextual information associated with the dissemination of misinformation. However, such systems require extensive social network data, higher computational resources, and more complex data collection processes, making them comparatively difficult to deploy in practical applications.

Despite significant progress, several challenges continue to affect fake news detection research. Misinformation evolves rapidly as creators continuously modify writing styles and dissemination strategies to evade automated detection systems. Distinguishing fake news from satire, opinion articles, parody, or partially true information remains particularly challenging because these categories often share similar linguistic characteristics. Furthermore, issues related to dataset imbalance, domain adaptation, multilingual content, explainability, fairness, and algorithmic bias continue to receive increasing research attention. Developing transparent and interpretable models is especially important in applications where automated decisions may influence public trust or policy implementation.

3.Methodology

3.1 Research Methodology and Research Design

The proposed fake digital news detection framework adopts a structured and experimental research methodology that combines Deep Learning, Machine Learning and Natural Language Processing, techniques to categories news articles as either real or fake. The research process is organized into series of systematic phases including data collection, text preprocessing, feature extraction, model development, training, performance evaluation, and prediction. Each phases plays vital role in enhancing the quality of the dataset and enhancing the overall accuracy and reliability of the classification models.

Because it allows for the implementation, training, and comparison of several classification algorithms under the same conditions, an experimental study methodology was chosen. This study evaluated the Long Short Term Memory architecture-based learning model with a number of conventional machine learning models, such as Logistic regression, Naïve Bayesian, Decision tree, Random forest, Support vector machine, and XG Boost. The objective is to determine the most effective model for fake digital news detection by analyzing using standard performance matrices

The first step in the suggested approach is gathering tagged news items from benchmark datasets that are accessible to the public. After that, a number of preprocessing techniques are applied to the raw textual data in order to eliminate unnecessary information and standardize the text. The cleaned text is transformed into numerical representations using feature extraction methods like word embedding and Term Frequency Inverse Document Frequency. These numerical characteristics are then used to train classification models that can recognize intricate linguistic patterns linked to both authentic and fraudulent news. The best-performing model is then chosen for real-time false news prediction after the trained models' classification performance is assessed using unseen test data.

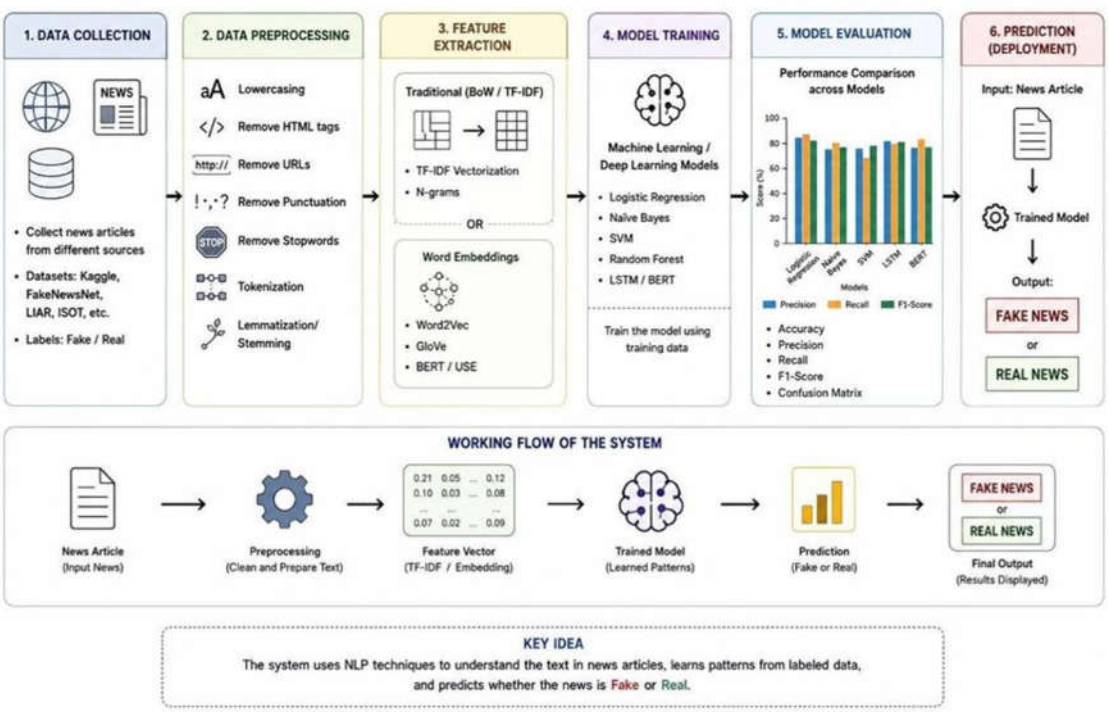

FIGURE 1. SYSEM DESIGN

3.2 Data Collection:

Any machine learning model's quality is mostly determined by the caliber, variety, and dependability of the training dataset. As a result, this study makes use of publicly accessible benchmark datasets that have been widely used in studies on false news detection. Two widely recognized datasets were considered for this research.

3.2.1 Wang (2017)'s LIAR Dataset

One of the most popular benchmark dataset for research on fake news detection is the LIAR dataset. It is composed of 12,836 brief political remarks that were gathered from the PolitiFact website. Professional fact-checkers manually verify each statement, classifying it into six veracity levels:

It is Mostly accurate

- Only Half True
- Almost Untrue
- Inaccurate
- Burning pants

Since binary categorization is the aim of this study, these six traits were split into two categories: Real and Fake. The Real category included statements marked as True and Mostly True, while the Fake category included statements marked as False, Barely True, and Pants on Fire. This conversion preserves trustworthy class labels while making the classification task simpler.

The LIAR dataset is particularly useful because it contains diverse political statements produced by multiple speakers under different contexts, allowing machine learning models to learn various linguistic patterns associated with truthful and deceptive information.

3.2.2 Kaggle Fake News Dataset

To improve the diversity of textual information, the Kaggle Fake News Dataset was also utilized. This dataset contains more than 20,000 labelled news articles collected from various online news sources.

- Each record typically includes:
- News headline
- Author name
- News article content

Classification label (Fake or Real) Compared with the LIAR dataset, the Kaggle dataset contains complete news articles instead of short statements. This allows deep learning models such as LSTM to capture contextual information across longer textual sequences. The inclusion of article headlines and body content also provides richer semantic information that improves feature extraction and overall classification performance.

The collected datasets were carefully examined to ensure consistency, remove duplicate records, eliminate incomplete entries, and verify the availability of class labels before proceeding to the preprocessing stage. Combining several benchmark datasets allows for a more thorough assessment of various classification methods and enhances the suggested false news detection system's capacity for generalization.

3.3Data Pre-processing:

Grammatical errors, HTML tags, punctuation marks, URLs, numbers, duplicate words, and other extraneous elements are frequently included in raw textual data from online news sources, which may have a detrimental effect on machine learning effectiveness. Thus, before feature extraction, a thorough preprocessing pipeline was put in place to convert the unprocessed news items into a clear, consistent, and machine-readable format.

3.3.1 Text Consolidation

Each news article generally consists of a title and the main body of the article. Before preprocessing, these two elements were combined into a single textual document in order to maintain all contextual information. Combining both fields enables the classification models to learn relationships between headlines and article content, thereby improving prediction accuracy.

3.3.2 Lowercasing

All characters in the combined text were converted into lowercase to eliminate case sensitivity during text processing. Without this step, words such as "Government," "government," and "GOVERNMENT" would be treated as different features despite representing the same concept. Lowercasing reduces vocabulary size and improves feature consistency throughout the dataset.

3.3.3 Noise Removal

Online news articles frequently contain HTML tags, hyperlinks, email addresses, hashtags, special symbols, punctuation marks, numerical values, and unnecessary whitespace. These elements generally do not contribute meaningful semantic information for fake news classification. Therefore, they were removed using regular expression techniques and text-cleaning functions to retain only linguistically relevant content.

3.3.4 Tokenization

This is a fundamental factor of natural language processing. Each cleaned news article was split up into discrete words, or tokens, during this procedure. Tokenization makes it possible to perform later NLP tasks including word embedding creation, stemming, stop-word removal, and feature extraction.

3.3.5 Stop word Removal:

Natural language contains many frequently occurring words such as the, is, are, was, and, of, for, and to. These words contribute very little discriminative information for text classification because they appear in almost every document.

Therefore, commonly occurring stop words were removed using predefined NLP libraries. Eliminating these terms reduces computational complexity while allowing the models to focus on more informative keywords such as election, fraud, government, vaccine, verified, and investigation, which are more relevant for fake news detection.

3.3.6 Stemming:

Words often appear in multiple grammatical forms that carry similar meanings. Stemming reduces these words to their common root form, thereby minimizing vocabulary size and improving model generalization.

For example:

Reporting → Report

Reported → Report

Reports → Report

This normalization process enables machine learning algorithms to treat different word variations as the same feature, reducing feature sparsity and improving classification efficiency.

3.3.7 Last Text Cleaning

The processed tokens were reassembled into standardized textual documents following tokenization, stemming, and stop-word removal. Only significant linguistic information is present in these cleaned papers, making them appropriate for numerical feature extraction methods like word embedding and TF-IDF.

The preprocessing phase is essential to enhancing the suggested false news detection framework's overall efficacy. Preprocessing improves the quality of retrieved features and greatly boosts the prediction accuracy of both conventional machine learning algorithms and deep learning models by lowering textual noise, decreasing word redundancy, and standardizing language patterns.

3.4 Feature Extraction:

Machine learning algorithms cannot directly process textual information because they require numerical input for computation. Therefore, after completing the preprocessing stage, the cleaned textual data were transformed into numerical feature vectors using appropriate feature extraction techniques. Feature extraction is one of the most critical stages in Natural Language Processing because the quality of extracted features directly influences the performance of classification models.

In this study, two feature representation techniques were employed depending on the type of classification algorithm.

3.4.1 Inverse Document Frequency Term Frequency

The Term Frequency Inverse Document Frequency technique was used by traditional machine learning models, such as Logistic Regression, Naïve Bayesian, Decision tree, Random forest, Support vector machine, and XG Boost, to transform textual documents into numerical vectors.

Traditional machine learning models, including Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Support vector machine, and XG Boost, converted textual texts into numerical vectors using the Term Frequency Inverse Document Frequency approach.

Here, The following formula is used to calculate the

TF-IDF value: $TF\text{-}IDF=TF(t,d)\times IDF(t, D)$

where:

- TF (Term Frequency) shows how commonly a term appears in a document.
- IDF (Inverse Document Frequency) quantifies a term's uniqueness throughout the full collection of documents.

The advantages of TF-IDF include reduced influence of common words, efficient handling of sparse textual data, lower computational complexity, and strong compatibility with traditional machine learning classifiers.

3.4.2 Word Embedding:

Dense vector representations that maintain semantic and contextual links between words are necessary for deep learning architectures, in contrast to conventional machine learning models. Consequently, word embedding was used by the Long Short-Term Memory model to encode text based data. Word embedding convert each word into a continuous numerical vector, with words that are semantically comparable being closer together in the vector space. Instead of just taking word frequency into account, this representation allows the model to capture contextual meaning.

For instance, because they share semantic ties, words like "doctor," "hospital," and "patient" are represented by vectors that are close to one another. In political news pieces, terms like "government," "election," and "parliament" also show contextual resemblance.

Word embedding are more successful for sequential deep learning models like LSTM than TF-IDF because they retain both syntactic and semantic information. The subsequent machine learning and deep learning classification models used the extracted numerical characteristics as their input.

3.5 Model Development and Training:

Several supervised machine learning and deep learning algorithms were used after feature extraction to divide news articles into two groups: fake and real. The best model for detecting false news is found by using a variety of classification algorithms, which allow for a thorough comparison of their predicting skills.

Prior to model creation, the labelled dataset was split into training and testing subsets. While the testing data was set aside to assess model performance on previously unknown cases, the training data allowed the algorithms to learn language patterns associated with both fake and real news stories.

This approach reduces the risk of overfitting and aids in evaluating each classifier's capacity for generalization.

The following classification model were implemented.

3.5.1 Logistic Regression:

One of the most used supervised learning methods for binary text classification is logistic regression. It uses a logistic (sigmoid) function to determine the likelihood that a news story falls into the actual or phony category.

Logistic regression is highly interpretable, computationally efficient, and offers exceptional classification accuracy when paired with TF-IDF data. On high dimensional sparse datasets that are frequently used in natural language processing, the technique excels.

3.5.2 Naive Bayesian

According to the Bayes theorem, Naive Bayesian is a probabilistic classifier. Given the class label, it is assumed that textual features are statistically independent.

Naive Bayes is still broadly used because of its ease of use, quick training speed, and efficiency in document categorization, even though this assumption is rarely true for natural language. For comparing more sophisticated machine learning algorithms, it acts as a crucial baseline model.

3.5.3 Decision Tree:

A rule-based supervised learning system called Decision Tree divides the dataset recursively based on informative textual qualities in order to classify documents.

The resulting tree structure is simple to see and understand. However, when trained on big textual datasets, Decision Trees frequently experience overfitting, which limits their prediction effectiveness when compared to ensemble learning techniques.

3.5.4 Random Forest

Random Forest also called as ensemble learning aggregates the forecasts of several Decision trees to increase the precision of classification.

In order to lower model variance and enhance generalization performance, each tree is trained using randomly chosen samples and feature subsets. Compared to individual Decision Trees, Random Forest offers more robustness and efficiently manages noisy textual input.

3.5.5 Support Vector Machine:

Because Support Vector Machine can distinguish classes within high-dimensional feature spaces, it is one of the best methods for text classification.

SVM finds the best hyperplane to maximize the difference between the classes of bogus and true news. SVM performs remarkably well and has continuously shown greater accuracy in a variety of Natural Language Processing applications because TF-IDF representations generate hundreds of sparse features.

3.5.6 Extreme Gradient Boosting (XG Boost)

According to the ideas of gradient boosting, XG Boost is sophisticated ensemble learning

system. Unlike conventional boosting algorithms, XG Boost significantly improves prediction accuracy while preventing overfitting by utilizing effective optimization strategies, parallel processing, and regularization approaches.

XG Boost stands out as exceptionally is powerful classical machine learning algorithms because of its capacity to model intricate nonlinear relationships, which allows it to recognize tiny linguistic distinctions between real and fraudulent news articles.

3.5.7 Long short term memory(LSTM):

A specialized Recurrent Neural Network architecture for linear data analysis is called Long Short-Term Memory.

LSTM automatically learns contextual representations from embedded word sequences, in contrast to conventional machine learning models that depend on explicitly designed features. Memory cells and gating mechanisms that selectively keep or discard information across lengthy text sequences are part of the design.

These techniques allow the model to capture semantic links, phrase structure, and long-range contextual dependencies that traditional algorithms frequently miss. As a result, LSTM performs better at identifying fake news, especially when examining long news stories with intricate language patterns.

The labelled training dataset was used to iteratively tune each classifier's parameters during model training until prediction errors were reduced. In order to maximize classification performance while reducing overfitting, hyper parameters such learning rate, maximum tree depth, number of estimators, hidden units, batch size, and number of training epochs were suitably set.

3.6 Model Evaluation:

Following training, an independent testing dataset comprising news articles not used during model training was used to assess the produced models. An objective assessment of the models' capacity to accurately categorize new articles is obtained by testing them on unobserved data.

- To obtain a comprehensive assessment of model performance, four widely accepted evaluation metrics were employed: Accuracy, Precision, Recall, and F1-Score. Additionally, a Confusion Matrix was used to analyse correct and incorrect predictions.
- Accuracy:

This measures the overall proportion of correctly classified news articles commonly found in all predictions.

$$ACCURACY = \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

When datasets are unbalanced, accuracy may not accurately reflect classifier efficacy even when it offers an overall performance measure.

Precision:

This measures the proportion of news articles predicted as fake that are actually false.

$$\boldsymbol{PRECISION = \frac{TP}{Tp+FP} \times 100}$$

Recall:

This evaluates the model's ability to correctly identify actual digital fake news articles.

$$\boldsymbol{Recall = \frac{TP}{TP \times FP} \times 100}$$

Higher recall indicates that fewer fake news articles are incorrectly classified as genuine.

F1-score:

This usually combines Precision and Recall into a single performance metric by computing their harmonic mean.

$$\boldsymbol{F1 = \frac{2 \times PRECISION \times RECALL}{PRECISION + RECALL}}$$

This metric is particularly valuable when balancing false positives and false negatives is equally important.

Confusion Matrix:

The Matrix provides a detailed summary of model predictions by comparing predicted labels with actual class labels.

Actual Class	Fake predicted	Real predicted
Fake	True Positive (TP)	False Negative (FN)
Real	False Positive (FP)	True Negative (TN)

The confusion matrix makes it possible to analyze classification errors in great detail and assists in determining the advantages and disadvantages of each algorithm. The suggested study offers a thorough comparison of machine learning and deep learning models to identify the highest efficient method regarding automated fake news identification by incorporating all assessment metrics.

3.7 Prediction (Deployment):

The best-performing classification model is used to forecast the veracity of previously viewed news articles following the training and evaluation phases. During the distribution step, the trained model is transformed into a useful application that can help users, media outlets, and fact checking groups spot bogus news instantly. To guarantee consistency and preserve prediction accuracy, the deployed system adheres to the same preprocessing and feature extraction pipeline used during model training.

When a user provides a news headline or an entire news story, the input text first goes through preprocessing steps like stemming, tokenization, stop word removal, punctuation and special character removal, and lowercasing. These procedures standardize the input text and eliminate extraneous details that can interfere with the prediction process.

The same feature extraction method used during model training is used to transform the cleaned

text into numerical feature vectors after preprocessing. While the LSTM model uses word embedding to capture semantic and contextual information, typical machine learning techniques use TF IDF to produce sparse numerical representations. The model will obtain consistent input representations if the same feature extraction techniques are used for both training and prediction.

The trained classification model receives the processed feature vectors, analyzes the learnt linguistic patterns, and determines the likelihood that the news article falls into the Real or Fake category. The class with the highest likelihood score is used to construct the final forecast. The system can include a confidence score that indicates how certain the forecast is in addition to the anticipated class.

The deployed fake news detection system offers several practical advantages. It enables rapid verification of online news content, supports journalists and fact checking organizations in identifying misleading information, assists social media platforms in reducing the spread of misinformation, and promotes greater public awareness regarding the credibility of digital information.

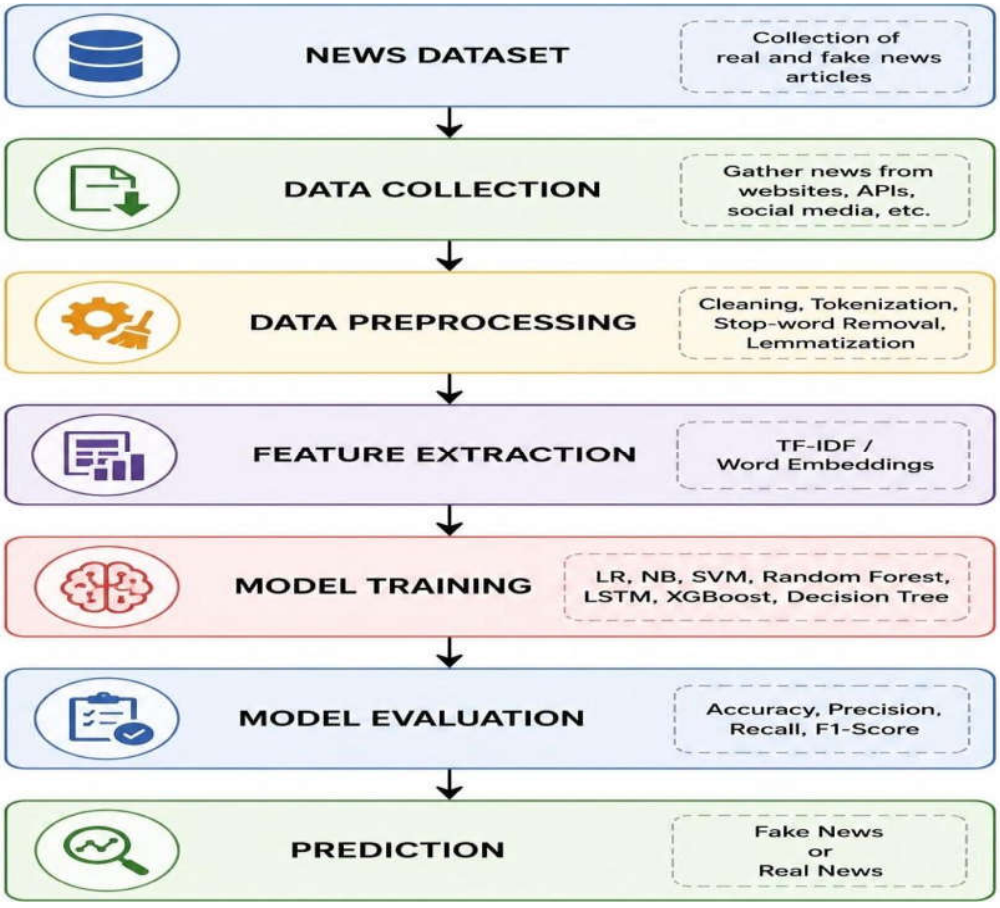

FIGURE 2: FAKE NEWS DETECTION WORK FLOW

The suggested architecture for detecting fake news employs a sequential procedure that methodically converts unprocessed textual data into a precise classification outcome. The first step of the procedure is gathering labeled news articles from publicly accessible benchmark datasets that include both real and fake news. Deep learning and supervised machine learning models are trained using these datasets.

The gathered data is subjected to extensive preprocessing, which eliminates extraneous components including stop words, HTML tags, punctuation, special symbols, and numerical values. To create clear, consistent textual data that can be analyzed, the text is further standardized using tokenization, stemming, and lowercasing.

Following preprocessing, the cleaned text is converted into numerical feature representations using word embedding for the LSTM model and TF-IDF for conventional machine learning algorithms. Important language features that allow the classifiers to discern between bogus and real news are preserved in these numerical representations.

A number of classification methods, such as Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine, XG Boost, and Long Short Term Memory, are then trained using the retrieved features. Each model gains discriminative patterns linked to authentic and fraudulent news stories throughout training.

An independent testing dataset is used to assess the classifiers after model training. Performance is assessed using the Confusion Matrix, F1score, Accuracy, Precision, and Recall. The model that performs the best is chosen for deployment based on these evaluation metrics.

Lastly, new news articles are fed into the trained model, which uses the same preprocessing and feature extraction workflow to determine if the material is real or fake news. This process guarantees an automated fake news detection system that is dependable, scalable, and effective enough to enable practical applications.

4. Results:

Using the same preprocessed dataset and identical experimental settings, six conventional machine learning algorithms and one deep learning model were compared in order to assess the suggested false news detection framework. Accuracy, precision, recall, and F1-score were used to evaluate the model's performance.

The findings demonstrated that while there were discernible variations in classification accuracy, all models performed satisfactorily. Due to the assumption of feature independence, Naïve Bayes recorded the lowest accuracy (90.1%), whereas Decision Tree reached 91.3% but was impacted by overfitting. The accuracy was increased to 95.1% via logistic regression, indicating that TF-IDF features are useful for linear classification. By using ensemble learning to reduce overfitting, Random Forest further improved the accuracy to 96.2%. Support Vector Machine (SVM) demonstrated its capacity to effectively categorize high-dimensional sparse text data with an accuracy of 97.0%.

Because of its gradient boosting and regularization strategies, XG Boost outperformed the other conventional models, achieving 97.9% accuracy, 98.0% precision, 97.6% recall, and 97.8% F1-score. By successfully learning contextual and sequential information, the Long Short-Term Memory model produced the best overall performance, with 98.3% accuracy, 98.4% precision, 98.1% recall, and 98.2% F1-score. These results show that XG Boost and

SVM give competitive accuracy with reduced processing requirements, however LSTM offers the best predictive performance.

4. Conclusion:

This overall study represents a comprehensive framework for digital fake news detection using machine Learning, Natural Language Processing, machine and deep learning techniques to automatically classify digital news articles as real or fake. The proposed methodology included data collection, text preprocessing, feature extraction, model training, performance evaluation, and real time prediction, providing a systematic approach for detecting false information.

Several machine learning algorithms and a deep learning model were evaluated to compare their effectiveness. The experimental results showed that all the above implemented models were capable of detecting digital fake news effectively. Among the machine learning techniques, XG Boost demonstrated the strongest overall performance, followed by Support Vector Machine, Random Forest, and Logistic Regression. Decision Tree and Naïve Bayes also produced satisfactory results but were comparatively not that much effective in capturing complex language patterns.

The Long Short-Term Memory model achieved the best overall performance due to its ability to capture contextual meaning, dependencies, and long term relationships within textual data. This enabled more accurate classification of genuine and fake news compared to conventional machine learning approaches.

In general, the findings indicate that integrating advanced NLP techniques with modern machine learning and deep learning models provides an effective solution for automated digital fake news detection. Such systems have significant applications in digital journalism, social media monitoring, online news verification, and fact verification platforms, helping to reduce the spread of false information and support the dissemination of genuine information.

REFERENCES:

1. Ahire, P. S., Lokhande, N. M., & Gadekar, M. J. (2026). Fake news detection using NLP and deep learning. Iconic Research and Engineering Journals, 9(11), 4366–4370.
2. Horne, Benjamin & Adali, Sibel. (2017). This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. 10.48550/arXiv.1703.09398.
3. Puri, Ridham. (2024). Fake News Detection: A Comprehensive Study on Analysis and Modern Classification Techniques. 10.13140/RG.2.2.23779.16164.
4. Randhe, K., Vyavahare, N., Vishwakarma, R., & Kudale, S. (2025). Fake news detection using natural language processing. International Journal of Scientific Research & Engineering Trends, 11(2), 2614–2619. Reddy, O. (2024). AI-based fake news detection using NLP.
5. Sasirekha, C.,& Bhuvaneshwari, B. (2025). Fake news detection using NLP. International Research Journal of Engineering and Technology (IRJET), 12(11), 282-285.
6. Shahwan younis ali, Dr. Ibrahim Mahmood, (2025). The Role of NLP In Fake News Detection and Misinformation Mitigation. Engineering and Technology Journal,

10(05), -. https://europub.co.uk/articles/-A-767315

7. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22 - 36.
8. Wang, W. Y. (2017). “Liar, liar pants on fire”: A new benchmark dataset for fake news detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017) (pp. 422–426).
9. Yadav, K., & Singh, J. B. (2026). Fake news detection using natural language processing and machine learning. International Research Journal of Engineering and Technology (IRJET), 13(4), 1395–1405.