INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

A Machine Learning Models for Classifying Fake and Real News Articles

Shivangi Shelke*, Dipali Jawale

Department of Computer Science, Dr. D. Y. Patil Arts, Commerce and Science College Pimpri- 18, Pune, Maharashtra, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP019

Received: 26 June 2025; Accepted: 30 June 2025; Published: 23 October 2025

Abstract — The era where misinformation spreads rapidly across digital platforms, ability to distinguish between authentic and fabricated news has become a critical societal challenge. This project presents a machine learning-based approach to fake and real news detection using natural language processing techniques. Utilizing a labelled dataset comprising 6,335 news articles, the model analyzes both the title and content of each entry to accurately classify them as either "FAKE" or "REAL." Pre-processing steps, including tokenization, vectorization, and noise removal, was applied to enhance text clarity. Multiple machine learning algorithms were evaluated, with performance measured through accuracy, precision, recall, and F1-score. The results underscore the efficacy of supervised learning techniques in automating the verification of news content, offering a scalable solution to combat the proliferation of misinformation in online media.

Keywords – Machine learning, Supervised learning, Natural Language Processing (NLP), Text Classification

I. Introduction

In today's digitally connected world, online platforms have become the primary source of news consumption. However, the ease of publishing and sharing content has also made it easier for false or misleading information to spread rapidly. The circulation of fake news poses significant threats to public awareness, trust, and decision-making, particularly in areas such as politics, health, and finance. As traditional fact-checking methods are often time-consuming and limited in scalability, there is a growing demand for automated solutions that can identify and filter deceptive news in real time.

Advancements in machine learning and natural language processing (NLP) have enabled the development of intelligent systems capable of analyzing textual data and detecting patterns that distinguish between real and fake news. This project leverages such technologies to build a predictive model that classifies news articles based on their authenticity. Using supervised learning methods, the system is trained on labelled news content, allowing it to learn from linguistic features and contextual cues present in the text.

In recent years, machine learning (ML), particularly in combination with natural language processing (NLP), has emerged as a powerful tool for analyzing and classifying text data at scale. Implemented in a Jupyter Notebook environment, the project follows a structured workflow: it begins with text preprocessing (such as tokenization, stopword removal, and vectorization), followed by the training and testing of classification algorithms. Evaluation metrics like accuracy, precision, recall, and F1-score are used to assess the model's effectiveness.

The primary objective of this study is to explore the feasibility and performance of machine learning techniques in fake news detection, and to demonstrate how such systems can contribute to reducing the impact of misinformation across digital platforms. Beyond technical implementation, this work also reflects on the ethical and societal impact of deploying automated fake news detection systems. While such models offer scalability and speed, they must be designed with caution to avoid bias, ensure fairness, and maintain transparency. As misinformation continues to evolve, so must the tools we use to combat it—making this research not only timely but also critical to the future of digital media integrity.

II. Literature Review

This research explored the use of machine learning algorithms such as Passive-Aggressive Classifier, Naive Bayes, and SVM for detecting fake news. They applied TF-IDF for feature extraction and compared the performance of these models. The study emphasizes the efficiency of traditional classifiers in handling text classification tasks [1].

The exponential growth of online platforms has led to a surge in misinformation, making fake news detection a critical research area. Early approaches such as the Naive Bayes classifier focused on word-frequency-based probability models to classify news as deceptive or not. These models, while simple, provided the foundation for more advanced techniques [4].

Subsequent studies explored the use of Support Vector Machines (SVM) and Logistic Regression, which proved effective for high-dimensional feature spaces typical in textual data. These models benefit from their robustness and interpretability in binary classification tasks such as real vs. fake news [2].

With the advent of deep learning, models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) have been employed to capture the sequential nature of language and improve contextual understanding. However, deep models

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

often require extensive training data and computational resources, making them less practical in low-resource environments [6].

Researchers also emphasize the role of textual features such as TF-IDF, n-grams, and sentiment polarity in enhancing classification accuracy. Standardized datasets, such as the one provided by kaggle, have further facilitated the comparison of different models under consistent evaluation protocols [7].

This study builds on these foundations by applying classical machine learning algorithms—Logistic Regression, Random Forest, and Decision Trees—on the fake and real news dataset, with feature extraction based on TF-IDF and unigram models. The models are evaluated using key metrics including accuracy, precision, and confusion matrix analysis, aligning with the methodologies outlined in previous research [5].

III. Methodology

3.1 Natural Language Processing

Natural Language Processing (NLP) is a specialized branch of artificial intelligence (AI) that focuses on enabling computers to understand, interpret, and generate human language in a meaningful way. It serves as the bridge between human communication and machine understanding, allowing systems to process and respond to text or speech inputs just like a human would.

In recent years, NLP has become a critical technology behind many modern applications—ranging from virtual assistants like Siri and Alexa to spam detection in emails, machine translation, and sentiment analysis on social media. Its ability to extract insights from unstructured text data has made it especially valuable in fields such as journalism, law, healthcare, and cyber security. The key challenge in NLP lies in the complexity and ambiguity of natural language. Words can have multiple meanings depending on context, grammar rules are often inconsistent, and human communication is full of nuances like sarcasm or emotion.

In the context of fake news detection, NLP allows the system to analyze the content of news articles—such as sentence structure, word frequency, and semantic relationships—and convert them into structured data. By applying NLP techniques like tokenization, text normalization, and vectorization (such as TF-IDF), the system can differentiate between deceptive and factual language patterns.

Thus, NLP is the foundation that enables the transformation of raw, messy text data into intelligent insights, making it a powerful tool for text-based classification tasks such as identifying fake news.

In fake news detection, several core NLP techniques have been applied to transform raw news articles into meaningful numerical data suitable for machine learning. These techniques play a vital role in analyzing and understanding textual content.

1. Text Cleaning and Normalization

Before using text for model training, it's important to clean and standardize it. The following steps are typically included:

Lowercasing: All characters are converted to lowercase to avoid treating words like "News" and "news" as different.

Punctuation Removal: Special characters and punctuation are removed to reduce noise in the data.

Whitespace Handling: Extra spaces and tab characters are removed.

Stop Word Removal (if applied): Common words like "the", "is", and "in" may be removed, as they don't add much meaning for classification tasks.

2. Tokenization

Tokenization is the process of splitting a large block of text into individual components (typically words or terms). While code does not explicitly tokenize with tools like nltk or spaCy, TF-IDF Vectorization internally performs this step by identifying terms from each document.

3. TF-IDF Vectorization

Key NLP techniques in implementation are TF-IDF (Term Frequency–Inverse Document Frequency). This method transforms the raw text into a matrix of numerical values that reflect:

Inverse Document Frequency (IDF) – How rare or unique the word is across the entire dataset.

TF-IDF helps the model by giving more weight to important and meaningful words and reducing the impact of very common words.

3.2 Dataset Description:

The dataset used in this implementation is a labeled collection of news articles categorized as either real or fake. It consists of multiple text fields, primarily focusing on the title and text content of each article. Each row in the dataset represents one news item, and it is assigned a binary label — where 1 denotes a real news article and 0 indicates a fake one. This dataset is commonly

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

used in binary text classification tasks and is suitable for training machine learning models to detect misinformation or deceptive content. Total Records are 6,335 news articles. 4 columns are present.

3.3 Workflow and Model Development:

1. Data Loading and Exploration: The dataset, consisting of labeled news articles, is loaded using the Pandas library. An initial analysis is done to check the shape, class distribution, and structure of the data to ensure it is balanced and suitable for binary classification.

2. Preprocessing: The textual data is cleaned to prepare it for analysis. This includes converting all text to lowercase, removing punctuation, special characters, and normalizing whitespace. These steps reduce noise in the dataset and help improve the quality of features extracted later. Stop word removal and tokenization may also be applied to retain only meaningful content.

3. Feature Extraction using TF-IDF: To convert raw text into numerical format, the TF-IDF (Term Frequency–Inverse Document Frequency) vectorizer is used. It transforms each news article into a vector that reflects the importance of each word relative to the entire corpus. This technique helps in emphasizing unique terms while down-weighting commonly occurring ones, making it effective for text classification.

4. Splitting the Dataset: The dataset is divided into training and testing sets using a 70:30 ratio with train_test_split. This ensures that the model is trained on a majority portion of the data and tested on a separate subset to evaluate how well it generalizes to unseen data. The random state is set for reproducibility of results.

5. Model Selection and Training: Four machine learning models—Logistic Regression, Navie Bayes Classifier, and Random Forest Classifier, Support Vector Machine —are implemented to perform the classification task. Each model is trained on the TF-IDF-transformed feature vectors. These models are chosen to compare the performance of both linear and ensemble-based approaches in detecting fake news.

6. Model Evaluation: After training, the models are evaluated using several performance metrics. These include accuracy, which measures overall correctness, and a classification report containing precision, recall, and F1-score, which provide more detailed insights into model performance. A confusion matrix is also plotted to visualize the number of true and false predictions.

3.3 Machine Learning Model:

1. Logistic Regression: Logistic Regression is a linear model frequently applied in tasks involving two-class classification. It predicts the probability of input data belonging to one of the two categories, making it well-suited for problems like fake news detection. It estimates the probability that a given input belongs to a particular class using the logistic (sigmoid) function. In this project, it is applied to classify news articles as either fake or real by analyzing patterns in the text data transformed into numerical features using TF-IDF. The simplicity and efficiency of Logistic Regression make it a strong baseline model for text classification tasks.

2. Random Forest Classifier: Random Forest Classifier is an ensemble-based algorithm that builds several decision trees using different portions of the dataset and feature sets. It averages the predictions of these individual decision trees to produce more accurate results. This approach helps in minimizing overfitting and improves the model's ability to generalize to new data. In this project, Random Forest is used to analyze TF-IDF features from the news articles and has shown strong performance due to its robustness and ability to handle complex feature interactions.

3. Support Vector Machine (SVM): Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification tasks. It identifies the most suitable boundary that effectively distinguishes between different classes within a high-dimensional feature space. In text classification problems like fake news detection, SVM is especially effective because it performs well with sparse and high-dimensional data, which is typical when using TF-IDF or Bag-of-Words features. SVM tries to maximize the margin between the two classes (e.g., fake vs. real news), leading to better generalization on unseen data. Variants like LinearSVC are commonly used for large text datasets due to their efficiency.

4. Navie Bayes Classifier: Naive Bayes is a probabilistic classifier based on Bayes' Theorem, assuming that all features are conditionally independent given the class label — hence the name "naive." Despite this strong assumption, it works surprisingly well in many practical text classification tasks. The Naive Bayes classifier is a simple yet powerful probabilistic machine learning algorithm based on Bayes' Theorem. It is particularly popular for text classification tasks such as spam filtering, sentiment analysis, and fake news detection.

IV. Result

Evaluation results of four machine learning models—Logistic Regression, Naive Bayes, Support Vector Machine (SVM Linear), and Random Forest—used for classifying fake and real news.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Model	Accuracy	Precision (avg)	Recall (avg)	F1-Score (avg)
Logistic Regression	0.9219	0.92	0.92	0.92
Naive Bayes	0.8927	0.89	0.89	0.89
SVM (Linear)	0.9369	0.94	0.94	0.94
Random Forest	0.9203	0.92	0.92	0.92

Table. 4.1 Model Performance Comparison

The comparison table presents the evaluation results of four machine learning models—Logistic Regression, Naive Bayes, Support Vector Machine (SVM Linear), and Random Forest—used for classifying fake and real news. Among all models, the SVM (Linear) classifier achieved the best overall performance with an accuracy of 93.69%, and the highest average scores across precision (0.94), recall (0.94), and F1-score (0.94). This indicates that the SVM model is both accurate and consistent in identifying fake and real news correctly. Logistic Regression and Random Forest performed similarly, with accuracies of 92.19% and 92.03%, respectively. Both models also showed balanced precision and recall values (0.92), making them reliable alternatives. Naive Bayes, while faster and simpler, achieved the lowest performance, with an accuracy of 89.27% and F1-score of 0.89. This suggests it is slightly less effective for this dataset, though still viable for rapid text classification.

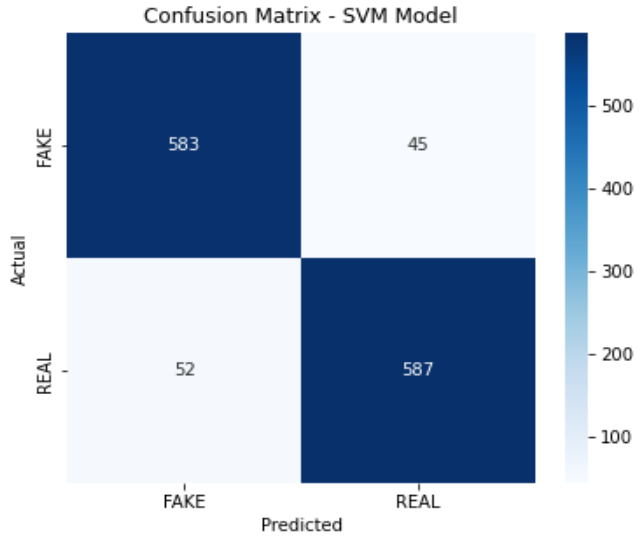

Fig 4.1 Confusion Matrix of Support Vector Machine

The confusion matrix for the Support Vector Machine (SVM) model provides a clear view of its classification performance on the fake news detection task. Out of all fake news samples, the model correctly identified 583 as fake, while misclassifying 45 as real. Similarly, it accurately classified 587 real news articles and incorrectly labeled 52 of them as fake. This distribution indicates that the model is highly effective at distinguishing between fake and real news, with only a small number of misclassifications. The balanced number of true positives and true negatives, along with relatively low false positives and false negatives, supports the model's strong performance in terms of precision, recall, and overall reliability.

Overall, the results demonstrate that SVM is the most effective model for this fake news detection task in terms of both accuracy and overall classification quality.

V. Conclusion

This implementation successfully demonstrates the use of natural language processing and machine learning techniques to classify news articles as real or fake. Through data preprocessing, TF-IDF feature extraction, and the application of multiple classification algorithms, the Support Vector Machine (SVM) emerged as the best-performing model with the highest accuracy, precision, recall, and F1-score. The evaluation metrics and confusion matrix further confirm the reliability and effectiveness of

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

the SVM model in identifying misinformation. Overall, this study highlights how machine learning can serve as a powerful tool in combating the spread of fake news and improving the trustworthiness of online information.

References

1. Jain, A., & Upadhyay, A. (2020). *Fake News Detection using Machine Learning Algorithms*. International Journal of Engineering Research & Technology (IJERT), 9(05), 623–627.
2. Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy, 1*(1), e9. https://doi.org/10.1002/spy2.9
3. Kaggle. (2020). *Fake and Real News Dataset*. https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset
4. Mihalcea, R., & Strapparava, C. (2009). The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 Conference* (pp. 309–312).
5. 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.
6. 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)* (pp. 422–426).
7. Zhou, X., & Zafarani, R. (2018). Fake News: A Survey of Research, Detection Methods, and Opportunities. *arXiv preprint arXiv:1812.00315*.
8. Ajao, O., Bhowmik, D., & Zargari, S. (2018). *Fake News Identification on Twitter with Hybrid CNN and RNN Models*. In *Proceedings of the 9th International Conference on social media and Society* (pp. 279–287). ACM.
9. Chakraborty, A. (2020). *Fake News Detection using Sentiment Analysis and GloVe Embeddings.*
10. Ajao, O., Bhowmik, D., & Zargari, S. (2018). *Fake News Identification on Twitter with Hybrid CNN and RNN Models*. In *Proceedings of the 9th International Conference on social media and Society*(pp.279–287).