IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, May 2026
Impact Factor: 8.2

Fake News Detection Using NLP

Prafull Thorat[1], Tejas Tidame[1], Tejeshwari Rasal[2]
[1] Student, Computer Science, C.H.M.E. Society's Bhonsala Military College, Nashik, Maharashtra, India
[2] Assistant Professor, Department of Computer Science,
C.H.M.E. Society's Bhonsala Military College, Nashik, Maharashtra, India

Abstract: *Social media, online news portals, and messaging apps have become the main sources of information in the digital age. However, the unchecked spread of fake news has serious social, political, and economic consequences, such as influencing public opinion, causing panic, damaging one's reputation, and interfering with elections. Human fact-checkers are used in conventional verification, but they are insufficient to combat the volume and speed of false information. In order to automate real-time authenticity assessment, this paper presents an intelligent Fake News Detection System that uses machine learning and natural language processing (NLP). The system improves the accuracy of classifying articles as authentic or fraudulent by analyzing textual content, extracting linguistic features, and modeling contextual patterns. Our strategy tackles the main drawbacks of manual techniques, providing a scalable, effective tool to stop the spread of false information and strengthen trustworthy news verification.*

Keywords: Natural Language Processing, KNN, machine learning, BERT

I. INTRODUCTION

Our primary information sources in the current digital age are messaging apps, social media sites, and online news portals. The problem is that false information spreads more quickly than the truth, which can have detrimental effects. It distorts public opinion, causes needless fear, destroys reputations, and even influences elections. Verifying if that article is legit? It's a nightmare, time-consuming, inefficient, and sometimes impossible for the average person. The traditional approach relies on human verification, but they're just not efficient enough to keep up with the speed and volume of false information spreading online every single minute. That's where we come in. This paper addresses an imperative need: creating an intelligent system for Fake News Detection using Natural Language Processing (NLP) and Machine Learning (ML) technique. The objective is to improve detection efficiency, stem the tide of false information, and put the power of real-time verification at the hands of users. In the following sections, we will dwell into the methodology, experiments, results, and how it contributes to the larger scheme of combating digital deception.

The information age has been characterized by the dissemination of false information to manipulate people's perceptions. Although today's systems use basic NLP and ML algorithms such as Logistic Regression and Random Forest, there is a demand for advanced algorithms to manage the intricate nature of the texts used in spreading misinformation.

Background

The digital revolution has completely changed the face of information consumption. Twitter (now X), Facebook, online news portals such as CNN or BBC, or WhatsApp have now become a part of our daily lives, bringing news to our fingertips. However, there is a darker side to this revolution: the explosion of fake news and misinformation. Fake news is not just annoying; it is destructive. It can spread panic, such as COVID-19-related fake news leading to dangerous cures. It can destroy reputations, such as fake news ruining people's lives overnight. And it can win or lose elections, compromising the very essence of democracy around the world. Research has found that fake news can spread 6 times faster than actual news on Twitter.[13] The traditional verification process? It involves fact-checkers from organizations such as Snopes or FactCheck.org. But they are already overwhelmed by the sheer number of fake news

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, May 2026
Impact Factor: 8.2

stories that need to be verified in a very short time. A fake news story spreads to millions in just minutes, but the fact-checker might take hours or even days to verify the claims made in the fake news story.
The need for the above process to change has given rise to the need for the automation process. And the answer to the above is Natural Language Processing and Machine Learning, which are particularly good at spotting text patterns that might have otherwise gone unnoticed by the fact-checker.

Motivation

This critical gap underscores the urgent need for automated, intelligent systems capable of accurately detecting and classifying fake news at large scale. Specifically, this project proposes the development of a sophisticated Fake News Detection System leveraging Natural Language Processing and Machine Learning techniques to analyze textual content and discern veracity. This system aims to extract meaningful features from news articles, employing advanced linguistic analysis and contextual understanding to identify patterns indicative of authentic or fabricated information, thereby improving detection accuracy and reducing the impact of misinformation.

Problem statement

"The latest models in Deep Learning need large volumes of annotated data to perform well. For cases when data is sparse or skewed, the SoftMax classifier in BERT is likely to face challenges in separating decision boundaries between the less-frequent categories. The current work develops a novel technique that involves integrating BERT with KNN to enhance the process of classification with limited availability of data. It utilizes the capability of KNN to estimate the local densities in the BERT-embeddings space."

Scope and objective of the research

1. Scope of the Research

Model Architectures: This research is going to consider types of architectures, namely, BERT (for pre-trained transformers using self-attention), K-NN (used for classification). And also focuses on a sequential hybrid pipeline where a pre-trained BERT model (e.g., Bert-base-uncased) serves as a feature extractor, and a k-NN algorithm replaces the standard SoftMax classification head.
Data Domain: The domain of data includes only texts (news headlines and full texts of the news articles).
Evaluation Criteria: Comparison of the models in terms of their classification accuracy in terms of metrics: Precision, Accuracy and F1-Score

2. Research Objectives

1. Integration of Hybrid Architecture

To develop a hybrid classification model where the classical SoftMax classifier is substituted by a k-Nearest Neighbour (KNN) algorithm, enabling the model to base its decision-making on the distance of the testing sample from existing examples within the latent vector space.

2. Verification of Model Performance and Reliability

To verify the efficiency of the developed model in detecting OOD data and achieving improved classification accuracy for the Kaggle dataset, particularly assessing whether the use of KNN and the underlying distance-based logic help minimize overconfidence issues in BERT models.

3.Comparison of Standalone model and hybrid model

To verify the accuracy and efficiency of standalone model and hybrid model for results.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, May 2026
Impact Factor: 8.2

Theoretical Fundamentals

1. Data Acquisition

This study uses an open-source data set provided by Kaggle containing news articles that are categorized into two Datasets true (21417 true news article) and fake (23502 fake news article), where the methodology will be tested using this data set and built upon the same.

2. Data Preprocessing

data preprocessing involves Word Piece Tokenization, where text is divided into subword units to handle vocabulary efficiently. Input sequences are formatted by adding special tokens such as [CLS] (classification) and [SEP] (sentence separation). The process involves adding structural tokens like [CLS] for classification and [SEP] for sentence separation, followed by converting these tokens into numerical input IDs. To ensure consistent input dimensions, sequences are padded or truncated to a fixed length, accompanied by an attention mask that directs the model to ignore padding during computation. [2] Unlike traditional NLP, this approach avoids heavy cleaning like stemming or stop word removal, preserving the full linguistic context the transformer needs to operate.

3. Feature Extraction

In this approach, BERT is used as a feature extractor to obtain dense semantic embeddings, which are then fed into KNN for classification. The KNN algorithm computes distances (e.g., Euclidean or cosine) between embedding vectors to classify new samples based on the nearest neighbours. This combination leverages BERT's contextual representation capability with KNN's simplicity and effectiveness in similarity-based tasks.

4.Model Training

The pre-trained BERT model is employed for extracting the contextual meaning of the news article text. It remains static throughout the entire process and serves as a feature extractor. The output of the BERT model is further processed via a shallow neural network comprising two layers of fully connected nodes with ReLU activation and dropout regularization. The last layer outputs probabilities of classification into fake and real news categories.

Simultaneously, the KNN classifier model is trained on the BERT embedding vectors. It classifies the sample based on proximity measured via cosine distance with 7 neighbours being considered. The final classification is achieved by aggregating the two classifiers via a soft voting strategy with weights 70% for BERT and 30% for KNN.

5. Classification

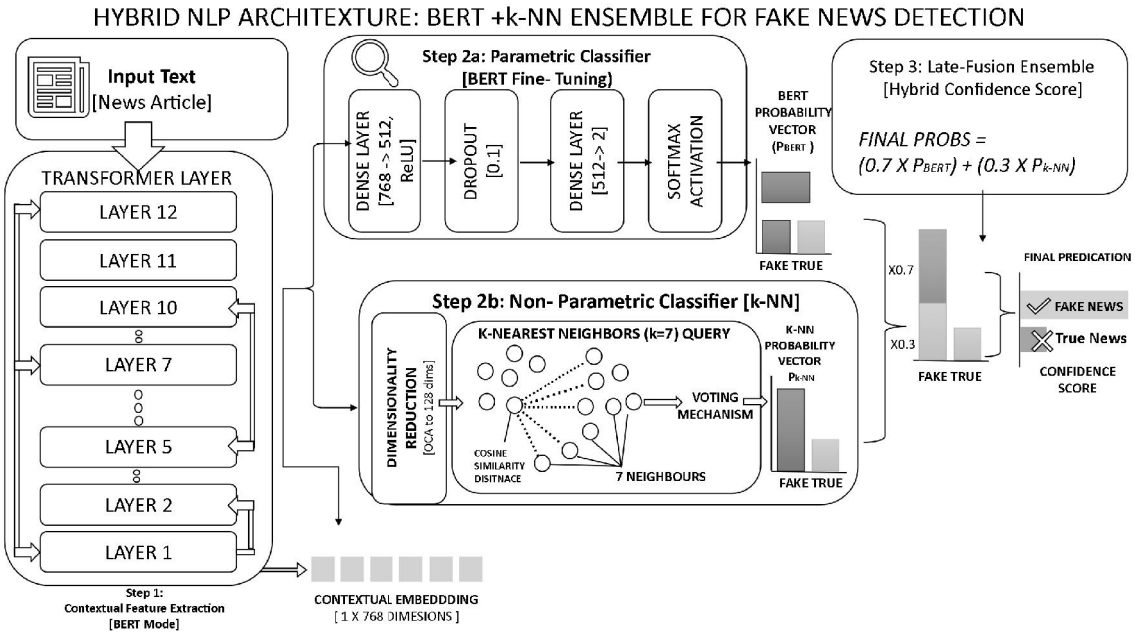
BERT-only classifier: BERT produces probabilistic outputs using a SoftMax function, selecting the class with maximum probability.

KNN-only classifier: KNN assigns labels based on the majority class of nearest neighbours in the feature space

Hybrid BERT + KNN ensemble: The final prediction is calculated using a weighted average of the probability scores:

$$final\_probs = (0.7 * bert\_probs) + (0.3 * knn\_probs)$$

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, May 2026
Impact Factor: 8.2



Technology Stack

• Language: Python.
• NLP: NLTK, spaCy, and BERT.
• Machine Learning: Scikit-learn, TensorFlow, or PyTorch.
• Data Handling: Pandas and NumPy.

Scope and Applications in It Industry:

• Social Media Monitoring: Detecting fake news on platforms like Twitter, Facebook, and Instagram to reduce misinformation spread.
• News Verification Systems: Assisting news agencies in automating the verification of user-generated content.
• Financial & Stock Market Analysis: Identifying false financial news that may manipulate stock prices.
• Public Health & Crisis Management: Detecting false health-related news during pandemics or emergencies to prevent panic and misinformation.
• IT Industry Integration: Integration with chatbots, recommendation systems, and content moderation tools for real-time verification.

Present and Future Scope

Social Media Platforms

In the present landscape (2020–2025), social media efforts are largely reactive, focusing on the real-time flagging and demotion of known text-based misinformation and the implementation of basic bot detection systems. Looking beyond 2025, the scope shifts toward a more proactive stance. This includes the sophisticated detection of synthetic media, such as Deepfakes, and the ability to track the movement of news stories across different platforms to identify coordinated manipulation.

Media & Journalism

Currently, AI serves as an assistant to human fact-checkers by scoring and prioritizing suspicious content to streamline their workflow. The future of this domain lies in the integration of Explainable AI (XAI). These tools will not only

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, May 2026
Impact Factor: 8.2

detect falsehoods but will also provide transparent justifications for fact-checking decisions, which is essential for building and maintaining public trust in journalistic institutions.

Government & Public Health

The present focus for government agencies involves identifying health-related rumours and managing "infodemics," such as those seen during COVID-19, to ensure clear public communication. In the future, the goal is the establishment of "Truth-as-a-Service" APIs. these specialized interfaces will allow governmental bodies to verify information instantly during crises and provide a robust defence against complex, state-sponsored disinformation campaigns.

Performance Evaluation & Benchmarking

```
========== BERT ONLY ==========
Accuracy: 0.890423162583519
...                precision    recall  f1-score   support

          True       0.88      0.89      0.89      3212
          Fake       0.90      0.89      0.89      3523

      accuracy                           0.89      6735
     macro avg       0.89      0.89      0.89      6735
  weighted avg       0.89      0.89      0.89      6735
```

```
========== KNN ONLY ==========
Accuracy: 0.8957683741648107
              precision    recall  f1-score   support

        True       0.89      0.89      0.89      3212
        Fake       0.90      0.90      0.90      3523

    accuracy                           0.90      6735
   macro avg       0.90      0.90      0.90      6735
weighted avg       0.90      0.90      0.90      6735
```

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, May 2026
Impact Factor: 8.2

```
========== BERT + KNN HYBRID ==========
Accuracy: 0.9063103192279138
              precision    recall  f1-score   support

        True       0.90      0.90      0.90      3212
        Fake       0.91      0.91      0.91      3523

    accuracy                           0.91      6735
   macro avg       0.91      0.91      0.91      6735
weighted avg       0.91      0.91      0.91      6735


========== ACCURACY COMPARISON ==========
                Model  Accuracy
0          BERT Only  0.890423
1           KNN Only  0.895768
2  BERT + KNN Hybrid  0.906310
```

Limitations and challenges

Challenges

High-Dimensionality (Curse of Dimensionality): Embeddings produced by BERT (Bidirectional Encoder Representations from Transformers) have 768 dimensions, which negatively impacts distance-based algorithms, such as K-Nearest Neighbour (KNN), since there will be fewer distinctive distances among data points.

Computational Latency: KNN compares new data to all training instances when predicting, making its execution time slower than conventional deep learning approaches based on forward pass.

Memory Usage: Keeping all training embeddings for use in KNN algorithm causes higher memory consumption than using deep learning algorithms.

Limitations

Fixed Feature Space: Once the BERT network was frozen, the feature space became constant and cannot adjust to changes in language or fake news.

Model Dependency on Hyperparameters: The choice of certain parameters, such as the value of k and ensemble weights, affects the model's performance. Such dependency makes it less universal.

II. CONCLUSION

The BERT standalone method is advantageous due to its ability to capture deep contextual features, it still suffers from being overconfident and not fully interpretable. On the other hand, the hybrid model successfully leverages both the representational power of BERT and the interpretability of KNN.

REFERENCES

[1] Study: False news spreads faster than the truth
https://mitsloan.mit.edu/ideas-made-to-matter/study-false-news-spreads-faster-truth
[2]https://mbrenndoerfer.com/writing/wordpiece-tokenization-bert-subword-algorithm
[3] R. Oshikawa, J. Qian, and W. Y. Wang, “A Survey on Natural Language Processing for Fake News Detection,” in Proceedings of the 12th Language Resources and Evaluation Conference (LREC), 2020.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, May 2026 Impact Factor: 8.2

[4] W. Y. Wang, "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.
[5] G. K. V. Maroju et al., "Comparative Analysis of LSTM, GRU, and BERT Models for Fake News Detection," 2025.
[6] "Detection of Fake News Using Machine Learning and Natural Language Processing," Journal of Advances in Information Technology (JAIT), vol. 13, no. 6.
[7] Jacob Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT*, 2019.
[8] Nicolas Papernot and P. McDaniel, "Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning," in *Proceedings of ICML*,
[9] Chuan Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in *Proceedings of ICML*, 2017.
[10] Rama Moorthy, H., Avinash, N. J., Krishnaraj Rao, N. S., Raghunandan, K. R., Dodmane, R., Blum, J. J., & Gabralla, L. A. (2025). Dual stream graph augmented transformer model integrating BERT and GNNs for context aware fake news detection. Scientific reports, 15(1), 25436. https://doi.org/10.1038/s41598-025-05586-w
[11] Khairunnisa, & Khairunnas, Khairunnas & Sutriawan, Sutriawan. (2026). A HYBRID BERT–GNN FOR DETECTING HOAXES AND NEGATIVE CONTENT IN INDONESIAN SOCIAL MEDIA. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer). 11. 627-639. 10.33480/jitk.v11i3.7330.
[12] Liu, Q., Xiao, K. & Qian, Z. A hybrid re-fusion model for text classification. Sci Rep 15, 9333 (2025). https://doi.org/10.1038/s41598-025-90864-w