Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

# Fake News Detection Using Machine Learning and Natural Language Processing

**Kapil Dilawar[1], Arshad Hussain[2]**

[1]Student, School of Computer Application & Technology, Career Point University, Kota (Raj.)

[2]Professor, School of Computer Application & Technology, Career Point University, Kota (Raj.)

**Abstract**

The spread of fake news on social media and online platforms has become a serious problem in today's digital world. Fake news can mislead people, create panic, and even influence political or social opinions. Traditional methods of manual fact-checking are too slow and inefficient, especially given the huge volume of online content.

This research focuses on the use of Machine Learning (ML) and Natural Language Processing (NLP) techniques to automatically detect fake news. Different algorithms such as Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, and deep learning models like Long Short-Term Memory (LSTM) are explored.

The paper also discusses feature extraction methods such as TF-IDF, Bag-of-Words, and Word Embedding's. The expected outcome of this research is to design an accurate, efficient, and real-time fake news detection system that can be applied to social media monitoring, news websites and digital media platforms.

**Keywords:** Fake News Detection, Machine Learning, Natural Language Processing, Text Classification, Social Media Analysis, Misinformation Detection.

## Introduction

In today's digital era, information is generated and shared at an unprecedented speed through social media platforms, news websites, blogs, and online communication channels. While this rapid flow of information has made access to knowledge easier, it has also given rise to a major global challenge — the widespread circulation of **fake news**. Fake news refers to intentionally fabricated or misleading content that is designed to deceive readers, manipulate public opinion, create social disruption, or promote political and financial agendas. Its impact has been observed in various domains such as elections, public health, financial markets, and emergency situations, where misinformation can lead to fear, panic, and even violence.

CAREER POINT
INTERNATIONAL JOURNAL OF RESEARCH
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

The traditional approach to identifying fake news primarily relied on **manual verification by journalists and fact-checkers**. However, with millions of posts, articles, and messages being generated daily, manual fact-checking has become **slow, inefficient, and nearly impossible at scale**. This has created a strong need for **automated, intelligent, and scalable fake news detection systems** capable of analyzing vast amounts of data in real time.

Recent advancements in **Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP)** offer powerful computational techniques that can analyze linguistic patterns, writing styles, semantic structures, and contextual meanings present in news articles. By learning from large datasets of real and fake news, these models can detect subtle patterns such as emotional tone, exaggeration, biased language, or misleading narrative structures. ML algorithms like Naïve Bayes, Logistic Regression, SVM, and Random Forest have shown promising results for fake news classification using statistical features such as **TF-IDF and Bag-of-Words**.

On the other hand, deep learning models such as **LSTM and Transformer-based architectures (BERT)** demonstrate a superior ability to understand long-term dependencies, sentence context, and semantic relationships within text. These models can interpret the deeper meaning behind words and phrases, making them highly effective for identifying deceptive language and misinformation.

Furthermore, the widespread use of the internet and social media platforms has increased the complexity of misinformation. Fake news today comes in multiple forms — manipulated text, sensationalized headlines, misleading statements, edited images, and fabricated evidence. This makes it essential to develop **hybrid models** that combine the strengths of both machine learning and deep learning to improve detection performance, accuracy, and robustness.

Therefore, this research aims to design and evaluate an automated Fake News Detection System that utilizes a hybrid integration of ML, DL, and NLP techniques. The system classifies news articles as **real or fake** by examining linguistic patterns, contextual clues, and semantic features extracted from both short political statements and long-form news content. The goal of the study is not only to achieve high accuracy but also to contribute to building safer digital ecosystems, reducing misinformation, and helping government agencies, media organizations, and online communities combat the spread of fake information.

©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

## Literature Review

1. Rubin et al. (2016) conducted one of the earliest structured studies on fake news detection by focusing on linguistic patterns in deceptive articles. They examined writing style, syntactic structure, and word usage to identify deception cues present in fabricated news. Their research highlighted that deceptive texts often contain exaggerated vocabulary, emotional expressions, and inconsistent sentence flow. Using classical text classification approaches such as SVM and decision trees, they achieved around 76% accuracy, demonstrating that linguistic features alone can provide substantial insight into distinguishing fake content from real news.

2. Ahmed et al. (2017) expanded the machine learning landscape for fake news detection by comparing multiple supervised algorithms, including Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression. Their experiments revealed that SVM delivered the best performance due to its capability to handle high-dimensional text data efficiently. The study emphasized that traditional machine learning models, when supported with feature extraction techniques like TF-IDF and Bag-of-Words, can still perform competitively in classifying deceptive and legitimate news articles. Their results became a benchmark for later research in the domain.

3. Wang (2017) introduced the renowned **LIAR dataset**, which contains more than 12,000 short political statements labeled across six truthfulness categories such as "true," "mostly false," and "pants-on-fire." This dataset addressed the need for more granular truth-level classification. Wang used deep learning approaches like CNNs and RNNs and demonstrated that these models significantly improve semantic understanding compared to traditional ML. His work provided a foundational dataset and methodology for future research on fine-grained credibility assessment.

4. Sharma et al. (2019) proposed a hybrid deep learning model combining TF-IDF feature extraction with LSTM networks to capture long-term dependencies in text. Their approach utilized word embeddings to convert textual data into meaningful vector representations, which enhanced the model's ability to understand context and detect patterns typical in fake news. The combination of TF-IDF (for statistical significance) and LSTM (for semantic depth) resulted in improved accuracy over standalone traditional or deep learning models, making it a strong hybrid strategy.

Career Point International Journal of Research (CPIJR)
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

5. Singh and Kumar (2021) introduced a hybrid ensemble model that integrated NLP preprocessing techniques with multiple machine learning algorithms to improve classification reliability. Their workflow included text cleaning, tokenization, stop-word removal, and feature generation, followed by ensemble techniques like Random Forest and Gradient Boosting. They found that combining multiple classifiers boosted overall accuracy and reduced misclassification, especially for social media datasets where texts are shorter, noisier, and more informal.

6. Zhou et al. (2020) presented a deep learning approach based on Convolutional Neural Networks (CNNs) to capture both shallow and deep semantic features in fake news articles. CNNs helped extract contextual n-gram patterns that are often overlooked by traditional models. Their experiments demonstrated that convolutional layers can effectively detect subtle deceptive traits such as exaggerated claims and manipulated phrases. This study highlighted the role of context-aware neural architectures in improving fake news classification performance.

7. Kumar and Gupta (2021) emphasized the importance of **feature engineering** in fake news identification. They incorporated sentiment analysis, metadata features (such as author, date, source), and contextual patterns into their dataset. Their hybrid model used a combination of ML classifiers along with handcrafted features, showcasing that external metadata can significantly strengthen prediction accuracy. Their research demonstrated that fake news cannot be detected from text alone; external contextual cues also play a vital role.

8. Popat et al. (2018) contributed an explainable framework for fake news detection by integrating evidence-based reasoning. Their model not only classified news articles as real or fake but also generated justifications for each decision by identifying supporting or contradicting evidence from trusted sources. Their method improved both accuracy and interpretability, addressing a major limitation of traditional black-box models. This study was significant because it shifted the field toward transparency and trustworthiness in automated fake news detection systems.

9. Ruchansky et al. (2017) proposed the "CSI Model," a comprehensive deep learning system combining three major components: **C**apture (text features), **S**core (user features), and Intervene (temporal propagation patterns). Their approach acknowledged that fake news detection should consider not just the content, but also the users who spread it and how it

CAREER POINT
INTERNATIONAL JOURNAL OF RESEARCH
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

propagates over time. This multi-source architecture significantly outperformed models that relied solely on textual features, demonstrating the importance of behavioral cues.

10. Shu et al. (2020) introduced a graph-based fake news detection method using social network propagation data. Their approach modeled news dissemination as a graph structure, analyzing how clusters of users react, share, and comment on specific articles. Fake news often spreads in a rapid, clustered manner, and their graph neural network (GNN) captured these propagation dynamics effectively. Their work established that social context and propagation structure are crucial components for achieving robust fake news detection.

**Research Methodology**

The methodology of this research focuses on creating an effective and reliable framework to detect fake news using a combination of Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP). This section explains each major component of the system — starting from data collection, text preprocessing, feature extraction, and model training, to the evaluation of classifiers. The overall objective of the methodology is to develop a hybrid system capable of learning both statistical word patterns and deeper contextual meaning, resulting in highly accurate classification of fake and real news.

**Overview of the Proposed System**

The proposed system uses a hybrid approach that combines classical ML algorithms with LSTM-based deep learning. The workflow begins with data collection from multiple authentic sources, which ensures diversity of news topics. After collection, the data undergoes preprocessing steps such as cleaning, tokenization, lemmatization, and normalization to remove noise and create standardized text. Next, feature extraction methods like Bag-of-Words, TF-IDF, and Word Embeddings are applied to convert textual data into meaningful numerical vectors. Classical ML models such as Naïve Bayes, Logistic Regression, SVM, and Random Forest are trained using TF-IDF features, whereas deep learning models, particularly LSTM, are trained using word embeddings to capture deeper context. At the final stage, both ML and DL predictions are combined using a hybrid fusion method to achieve better accuracy and robustness.

**Data Collection**

Data collection forms the foundation of the fake news detection system, as the performance of the model depends on the quality and diversity of available texts. In this research, two

**Career Point International Journal of Research (CPIJR)**
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
**October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395**

widely used benchmark datasets — the LIAR dataset and the Kaggle Fake News dataset — were used. These datasets provide a rich variety of political, social, and general news statements, ensuring that the model learns broad linguistic patterns rather than topic-specific features. To improve reliability, data from both sources was cleaned, standardized, and merged into a unified dataset.

**LIAR Dataset**

The LIAR dataset is one of the most comprehensive fake news resources, containing over 12,800 short political statements collected from PolitiFact.com. Each statement is manually labeled across six truthfulness levels ranging from "True" to "Pants-on-Fire," allowing detailed credibility analysis. The dataset also includes metadata such as speaker information, context of the statement, subject, and fact-check justification. This multidimensional structure helps the model understand not just the text but also the context, making it suitable for training hybrid models. For this research, the labels were converted to a binary format (True/Fake) to support classification tasks.

**Kaggle Fake News Dataset**

The Kaggle Fake News dataset contains thousands of full-length news articles labeled as Fake or Real. It includes fields such as title, author, and the complete news text, covering multiple domains like politics, sports, health, and technology. This domain diversity helps the model learn general deception patterns rather than overfitting to a specific topic. Unlike the short statements in the LIAR dataset, the Kaggle dataset provides long-form textual content, enabling deep learning models to capture extended sentence patterns and semantic relationships.

**Data Integration**

To boost dataset size and model robustness, both datasets were integrated after converting them into a uniform structure. Fields such as label names, text columns, and missing values were standardized to maintain consistency. Duplicate entries were removed, incomplete rows were discarded, and the final merged dataset was stored in CSV format for further preprocessing. This integration ensures that the model is trained on a richer and more balanced dataset.

**Importance of Data Diversity**

CAREER POINT
INTERNATIONAL JOURNAL OF RESEARCH
Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

Data diversity is essential because fake news patterns differ across categories. Models trained only on political datasets may fail when dealing with health-related fake news. To avoid this bias, the combined dataset was balanced across categories and label distribution. Oversampling techniques such as SMOTE were applied to handle class imbalance, ensuring that the model learns meaningful patterns from both fake and real news samples.

**Experimental Results and Analysis**

This section provides a detailed evaluation of the proposed Fake News Detection System, which integrates Machine Learning (ML) and Deep Learning (DL) techniques to classify news articles as fake or real. The goal of this experiment was to compare the performance of different algorithms, analyze their accuracy, and determine the most efficient model for real-world deployment.

All experiments were conducted using Python programming language with powerful open-source libraries such as Scikit-learn, TensorFlow, Keras, Pandas, NumPy, and Matplotlib.

**Experimental Setup**

The experimental setup was carefully designed to ensure consistency, reliability, and reproducibility. All models were trained and tested using the same datasets and computational environment.

| **Parameter** | **Description** |
|---|---|
| Programming Language | Python 3.10 |
| Development Environment | Google Colab (GPU Runtime) |
| Libraries Used | Scikit-learn, TensorFlow, Keras, Pandas, NumPy, Matplotlib |
| Dataset Used | LIAR Dataset, Kaggle Fake News Dataset |
| Training/Validation/Test Split | 80% / 10% / 10% |
| Hardware Configuration | 12 GB RAM, NVIDIA Tesla T4 GPU |
| Performance Metrics | Accuracy, Precision, Recall, F1-Score, Confusion Matrix, ROC Curve |

Each algorithm was trained independently on the same dataset for fairness. Hyperparameters such as learning rate, number of estimators, and kernel functions were tuned using Grid Search Optimization to achieve optimal results.

©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

**Data Description**

The combined dataset contained over 40,000 news articles from multiple domains, such as:

- Politics
- Economy
- Health
- Sports
- Entertainment

Each article included fields like headline, body text, author, and label (Fake or Real). The data was preprocessed by removing punctuation, stop words, and duplicate entries. After cleaning, nearly 38,000 balanced samples were used for training and evaluation.

Oversampling techniques such as SMOTE (Synthetic Minority Oversampling Technique) were applied to ensure an equal number of fake and real samples, preventing class imbalance issues.

**Evaluation Metrics**

To accurately assess model performance, multiple evaluation metrics were used:

- Accuracy: The ratio of correctly classified articles to total predictions.
- Precision: How many predicted fake news articles were actually fake.
- Recall: The model's ability to correctly identify fake articles.
- F1-Score: The harmonic mean of Precision and Recall, balancing false positives and negatives.
- Confusion Matrix: Visual representation of correct and incorrect classifications.
- ROC Curve: Plots True Positive Rate (TPR) vs False Positive Rate (FPR); the higher the AUC (Area Under Curve), the better the classifier.

**Experimental Results of Machine Learning Models**

Machine learning models were trained using TF-IDF (Term Frequency–Inverse Document Frequency) features extracted from the preprocessed text. Each algorithm underwent fine-tuning for best hyperparameter configuration.

**(a) Naïve Bayes Classifier**

- Accuracy: 88.9%

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

- Precision: 86.5%
- Recall: 84.7%
- F1-Score: 85.6%

**Observation:**

Naïve Bayes showed good performance on short news snippets but failed to capture contextual depth in long or complex articles. Still, it was fast and computationally efficient, making it a good baseline model.

**(b) Logistic Regression**

- Accuracy: 92.1%
- Precision: 91.0%
- Recall: 90.4%
- F1-Score: 90.7%

**Observation:**

This model provided stable and interpretable results, efficiently identifying fake articles using word frequency and tone. However, its linear nature limited deeper contextual understanding.

**(c) Support Vector Machine (SVM)**

- Accuracy: 94.2%
- Precision: 93.5%
- Recall: 94.0%
- F1-Score: 93.7%

**Observation:**

SVM delivered the best performance among classical ML algorithms due to its ability to handle high-dimensional data. Although training took longer, its results were the most reliable and consistent.

**(d) Random Forest Classifier**

- Accuracy: 93.1%
- Precision: 91.8%

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

- Recall: 92.5%
- F1-Score: 92.1%

**Observation:**

Random Forest combined multiple decision trees to enhance prediction stability and reduce overfitting. It performed slightly below SVM but offered good robustness.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Naïve Bayes | 88.9 | 86.5 | 84.7 | 85.6 |
| Logistic Regression | 92.1 | 91.0 | 90.4 | 90.7 |
| SVM | 94.2 | 93.5 | 94.0 | 93.7 |
| Random Forest | 93.1 | 91.8 | 92.5 | 92.1 |

**Deep Learning Models**

Deep learning models were trained using Word Embeddings instead of TF-IDF to capture semantic relationships.

**(a) LSTM (Long Short-Term Memory) Model**

- Accuracy: 95.3%
- Precision: 94.8%
- Recall: 95.1%
- F1-Score: 94.9%

**Observation:**

The LSTM network captured long-term linguistic dependencies, effectively recognizing patterns like emotional language or repetition. Training required more time but resulted in better contextual understanding than traditional ML models.

**(b) BERT + LSTM Model**

- Accuracy: 96.0%
- Precision: 95.5%
- Recall: 95.7%

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

- F1-Score: 95.6%

**Observation:**

Combining BERT embeddings with LSTM further improved contextual understanding. BERT's bidirectional transformer architecture allowed better interpretation of words based on their surrounding context, improving the detection of sarcastic or ambiguous content.

**Hybrid Model (SVM + LSTM)**

To combine the advantages of both ML and DL models, a hybrid SVM + LSTM system was implemented. It used SVM to capture lexical-level features and LSTM to extract semantic-level patterns.

**Performance:**

- Accuracy: 96.4%
- Precision: 95.9%
- Recall: 96.2%
- F1-Score: 96.0%
- AUC: 0.97

**Observation:**

The hybrid approach minimized both false positives and false negatives, providing a balanced, highly accurate system suitable for real-world use.

**Discussion of Results**

- **Traditional ML models** (e.g., SVM, Logistic Regression) were simple, interpretable, and fast but lacked deep contextual understanding.
- **Deep Learning models** (LSTM, BERT) captured semantics and emotional tone effectively, yielding higher accuracy.
- The **Hybrid model** (SVM + LSTM) offered the best balance of accuracy, precision, and recall, showing strong potential for deployment in social media monitoring systems.

**Key Insights:**

- Fake news articles often contain emotionally charged words like *"shocking," "exclusive,"* and *"urgent."*
- Combining lexical (TF-IDF) and contextual (BERT embeddings) features improves classification.
- Cross-validation confirmed result stability with ±0.5% variance across folds.

**Visualization Summary:**

- **Accuracy Curve:** Training accuracy stabilized at ~96%.
- **Loss Curve:** Validation loss decreased steadily, indicating no overfitting.
- **ROC Curve:** AUC = 0.97, showing high sensitivity and specificity.

**Error Analysis**

Despite strong performance, a few limitations were observed:

1. **Ambiguous News:** Articles mixing true and false claims were sometimes misclassified.
2. **Satirical Content:** Parody or humorous news was occasionally flagged as fake.
3. **Domain Bias:** Underrepresented domains like technology had slightly lower accuracy.
4. **Short Headlines:** Limited context led to misclassification.

**Future Improvements:**

- Multi-label classification (Fake, Partly True, Satire, Real)
- Integration of **RoBERTa** or **GPT-based** transformers
- Use of **image and metadata** features for multimodal analysis

**Comparative Summary**

| Approach | Accuracy | Strengths | Weaknesses |
|---|---|---|---|
| Naïve Bayes | 88.9% | Fast and simple | Poor context understanding |
| Logistic Regression | 92.1% | Interpretable, | Linear limitation |

Career Point International Journal of Research (CPIJR)
©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

| | | | |
|---|---|---|---|
| | | consistent | |
| SVM | 94.2% | Strong with TF-IDF | High training cost |
| Random Forest | 93.1% | Non-linear, robust | Slower prediction |
| LSTM | 95.3% | Deep semantic learning | Long training time |
| BERT + LSTM | 96.0% | Context-aware | Requires large memory |
| Hybrid (SVM + LSTM) | 96.4% | Best performance, balanced approach | Slightly complex architecture |

**Summary of Experimental Findings**

- Integration of ML and DL models significantly improved fake news classification.
- BERT embeddings provided richer contextual meaning compared to TF-IDF.
- LSTM effectively modeled word dependencies and sequential context.
- Hybrid SVM + LSTM achieved the highest accuracy of 96.4%, proving to be the most reliable approach.
- The proposed system demonstrated strong scalability, robustness, and real-world applicability for detecting misinformation on online platforms and social media.

**Conclusion**

In the present era of digital communication, the spread of information through social media and online platforms has become extremely rapid. However, the reliability of such information has come under serious question due to the widespread presence of fake and misleading news. This research focused on developing an automated and intelligent system capable of detecting fake news using the combined power of Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP).

The primary goal of this research was to design a model that could analyze and classify news articles as fake or real with a high level of accuracy, efficiency, and interpretability. The system aimed to support fact-checking organizations, journalists, and social media platforms in identifying misinformation before it spreads to the public. The research also sought to

©2022 CPIJR | Volume 4 | Issue 4 | ISSN : 2583-1895
October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395

contribute academically to the fields of artificial intelligence and natural language understanding by exploring the effectiveness of hybrid learning models.

The research began by collecting and preparing datasets from reliable and publicly available sources, namely the LIAR dataset and the Kaggle Fake News dataset. These datasets provided a diverse collection of real and fake news articles covering various domains such as politics, economy, health, and entertainment. The preprocessing phase included data cleaning, tokenization, stop-word removal, lemmatization, and text normalization to ensure the textual information was consistent and suitable for machine learning analysis.

For the detection process, several feature extraction techniques were employed. Statistical approaches such as Bag-of-Words (BoW) and TF-IDF (Term Frequency–Inverse Document Frequency) were used to represent text numerically, while Word Embeddings (Word2Vec, GloVe, and BERT) were implemented to capture the semantic meaning and contextual relationships between words. These features helped the models understand both the surface structure and the deep context of news articles.

Multiple Machine Learning algorithms were trained and tested, including Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), and Random Forest. Among these, SVM produced the best results with an accuracy of 94%, proving that it is highly effective in handling high-dimensional textual data. Logistic Regression and Random Forest also performed well, while Naïve Bayes served as a fast and simple baseline model for comparison.

**References**

1. 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. https://doi.org/10.1145/3137597.3137600
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. Zhou, X., & Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. ACM SIGKDD Explorations Newsletter, 21(2), 48–60. https://doi.org/10.1145/3373464.3373479
4. 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). Association for Computational Linguistics.

5. Singh, R., Gupta, A., & Kumar, N. (2020). Combining NLP and machine learning for fake news detection. International Journal of Information Technology and Computer Science, 12(4), 14–25.
6. Gupta, S., & Kumar, R. (2021). A hybrid model for fake news detection using BERT and SVM. Journal of Information and Computational Science, 11(9), 250–260.
7. Sharma, D., Jain, P., & Arora, R. (2022). Fake news detection using ensemble machine learning models. International Journal of Advanced Computer Science and Applications, 13(1), 245–257.
8. Patel, M., Joshi, V., & Reddy, S. (2022). Cross-lingual fake news detection using multilingual NLP models. International Journal of Artificial Intelligence Research, 6(2), 89–102.
9. Raj, P., & Mehta, A. (2023). Evaluating the role of sentiment analysis in fake news detection. International Journal of Data Science and Analytics, 9(3), 213–229.
10. Ali, T., Khan, F., & Ahmed, Z. (2023). Real-time fake news detection using hybrid ML and deep learning approaches. Journal of Information Technology and Digital Transformation, 5(2), 55–68.
11. Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). *Fake news detection: A deep learning approach.* SMU Data Science Review, 1(3), 10–18.
12. Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). *Detecting fake news with capsule neural networks.* Applied Soft Computing, 101, 107050. https://doi.org/10.1016/j.asoc.2020.107050
13. Bozarth, L., & Budak, C. (2020). *Toward a better performance evaluation framework for fake news detection.* Information Systems Frontiers, 22(5), 1039–1055.
14. Zhou, X., & Zafarani, R. (2020). *A survey of fake news: Fundamental theories, detection methods, and opportunities.* ACM Computing Surveys (CSUR), 53(5), 1–40. https://doi.org/10.1145/3395046
15. Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2021). *FakeBERT: Fake news detection in social media with a BERT-based deep learning approach.* Multimedia Tools and Applications, 80(8), 11765–11788.

**Career Point International Journal of Research (CPIJR)**
©2022 CPIJR ǀ Volume 4 ǀ Issue 4 ǀ ISSN : 2583-1895
**October-December-2025 | DOI: https://doi.org/10.5281/zenodo.18645395**

16. Cui, L., Lee, D., & Sun, H. (2020). *CoAID: COVID-19 healthcare misinformation dataset.* In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 875–884).
17. Silva, A., & Santos, J. (2022). *Explainable AI approaches for fake news detection.* Journal of Artificial Intelligence Research and Applications, 8(2), 44–60.
18. Ghosh, S., & Jalal, A. (2021). *Detecting fake news using sentiment analysis and natural language processing.* International Journal of Computer Applications, 183(45), 1–8.
19. Raza, S., & Ding, C. (2023). *Multimodal fake news detection: Combining text and image features using deep learning.* Neural Computing and Applications, 35(11), 8103–8122.
20. Dwivedi, P., & Kumar, M. (2023). *Adaptive learning models for real-time fake news detection in multilingual datasets.* Journal of Machine Learning and Intelligent Systems, 9(4), 120–136.