

## NLP-Based Model for Identifying Misinformation

**Dr. Rakesh Kumar Khare<sup>1</sup>, Nipun Rao Boyana<sup>2</sup>, Simrat Kaur<sup>3</sup>, Sanchita Sharma<sup>4</sup>**

*Computer Science and Engineering,SSIPMT, Raipur, India*

[rk.khare@SSIPMT.com](mailto:rk.khare@SSIPMT.com), [nipun.boyana@SSIPMT.com](mailto:nipun.boyana@SSIPMT.com), [simrat.kaur@SSIPMT.com](mailto:simrat.kaur@SSIPMT.com), [sanchita@SSIPMT.com](mailto:sanchita@SSIPMT.com)

### Abstract

The massive proliferation of digital media has significantly increased the transmission of fake news and disinformation, which affects the confidence of the population and decision-making. The vast quantity of online content can not be managed through manual methods of fact-checking and requires automated answers. This paper introduces an NLP-based machine learning system that can be used to identify fake information in news stories. The system uses text pre-processing, TF-IDF feature extraction and conventional machine learning classifiers such as Support Vector Machine (SVM) and Logistic Regression. The experimental analysis of the proposed model shows that the model can have an accuracy around 91 per cent, which demonstrates that interpretable and lightweight machine learning methods can effectively detect and identify fake information in academic and real-time prototype settings.

**Keywords:** Identification of False News, Disinformation, Natural Language Processing, Machine Learning, and TF-IDF.

**Citation:** Dr. Rakesh Kumar Khare, Nipun Rao Boyana, Simrat Kaur, Sanchita Sharma. 2025. NLP-Based Model for Identifying Misinformation. *FishTaxa* 36(1s): 460-466

### Introduction

Modern day digital platforms have revolutionized the creation and provision of information and have led to instantaneous worldwide communication. Due to this, accessibility has improved, and false information has also spread at a faster rate. Fake news is news that is purposely factually incorrect or false, or that is intentionally misleading information that is presented as true with often the aim of influencing a people, or to secure political or economic benefits.

False information can fast spread and affect elections, health choices of people, and even social peace. Automated detection methods are also under study by researchers since manual verification processes cannot be scaled to this problem. Linguistic patterns can be useful in detecting the authenticity and the fraudulent nature of the news articles with the help of the machine learning (ML) and natural language processing (NLP) procedures.

### Objective:

This study aims at creating and evaluating a machine learning platform using natural language processing (NLP) that can easily and effectively identify false information on its own.

### Literature Review

#### 2.1 LITERATURE SURVEY

With the increasing grounds of digital media platforms, the false information and fake news has acquired a lot of coverage. Although this kind of fake news dissemination in the form of propaganda and rumors is as old as history, the dissemination of so-called fake news has been significantly increased by social networks and online publication platforms, increasing in magnitude, speed, and effect. To hit upon this challenge, different scientific fields have contemplated on behavioral, linguistic and computational methods. Vosoughi et al. (2018) carried out one of the most relevant studies in terms of the spread of fake and real news on the Internet. They discovered that the fake news has higher probabilities to spread quicker, deeper and wider than the real news, first of all, due to novelty and emotional resonance. This observation demonstrated the pressing necessity in the automated systems that could identify and sieve off fake information at a mass scale.

Shu et al. (2017) provided a comprehensive insight into data mining in fake news detection, as it emphasizes the significance of content-, social, and user-based features. Their work has served as the basis of application of machine learning techniques in detecting authenticity of news and analysis of the patterns of texts. It was also noted by the authors that linguistic indications, in particular those that are used together with statistical learning models, can be fairly useful as predictors of fake information.

The article by Castillo et al. (2011) concentrated on truthfulness of information in the social media, which in this case is Twitter.

They examined such characteristics as the propagation behavior, credibility of the source and writing style. This experiment suggested that the credibility evaluation might be partially automated with the methods of learning that are supervised and, therefore, it provokes a new round of research on the text-based classification algorithms.

Wang (2017) presented the LIAR dataset, a benchmark dataset to identify fake news and a set of short political statements that are marked by various degrees of truthfulness. This data set helped in hastening the research on automated misinformation detection and also contributed to the possibility of standard evaluation of machine learning models. This study had shown that despite the basic characteristics of lingo, good classification results could be achieved provided the features were constructed in this manner.

## Role of Machine Learning

To build a resistant force against the growing problem of disinformation, machine learning (ML) is required to allow analyzing textual content automatically, at scale, and based on data. In contrast to the classical rule-based systems, the ML algorithms acquire the patterns directly due to the data, which enables them to follow new language trends and various unethical methods applied in crafting fake news. Language, semantics, and statistical features can effectively be used along with the Natural Language

Processing (NLP) in order to identify news articles as genuine or fake according to the machine learning model.

### 3.1 Data Acquisition and Understanding

The most important aspect of a machine learning system is high quality data. Some of the most recurrent data used in the identification of fake news are typically connected with news pieces and categorized as either real or fake. The data features are contained in each of the data fields in the form of article title, article content, and authenticity label. These labeled datasets make it possible to do supervised learning and make the models learn to associate textual patterns with target class labels.

Such information can be collected in filtered archives of fake news, reputable news creators, and publicly available information. In order to reduce bias and maximize likelihood of generalizing the results to other population, it is advisable to make the information diverse and balanced. Unbalanced or mislabeled data may have a negative impact on the accuracy and equity in prediction.

### 3.2 Data Preprocessing and Preparation

The noise present in raw textual data includes punctuations, stop words and symbols, URLs and disproportionate formatting. The processing, i.e. the preprocessing, is also among the most significant steps to process the text so that the machines learning can be analyzing it. Most of the pre-processing phases are of the following nature:

Be like it, make it small.

Elimination of special characters, URLs, digits and punctuations.

Blocking of terminologies that are not semantically relevant term words. Disaggregation of the text in the form of a word through tokenization.

lemmatization or stemming Words are distilled down to the minimum. The actions reduce the sizes and optimize the learning behaviour of the model and the capability of the model to acquire significant linguistic patterns.

### 3.3 Feature Engineering Using NLP Techniques

Because the direct processing of the raw text by the machine learning framework is not possible, feature engineering is needed to put the textual values into the form of the numbers. NLP-based feature extraction techniques enable models to isolate the significance of words, frequency and contextual significance.

#### 3.3.1 TF-IDF Vectorization

Term Frequency-Inverse Document Frequency (TF-IDF) is one of the most used feature extraction techniques to determine fake news. Uncommon words in the corpus which are common in a specific document are assigned more weight. It becomes simpler to identify words and phrases that are commonly associated with misinformation, such as hyperbole or emotionally charged words and phrases.

TF-IDF is interpretable, computationally-efficient and can be used with traditional machine learning models. It is a critical component of the proposed system due to its ability to be used in representing textual data.

## 3.2. BENEFITS OF MACHINE LEARNING

- 1.Automation - Does away with manual verification.
- 2.Scalability - Operates in large amounts of data.
- 3.Adaptability - acquires changing patterns of language.
- 4.Consistency - Makes consistent classification decisions.
- 5.Efficiency -

Works in almost real-time conditions.

## Methodology

### 4.1 Data Collection

The information, which this research will be based on, is the information of the labelled news articles which can be considered real or fake. The articles data records will include the title of the articles, the main information and a label of a specific classification which will ascertain the authenticity of the articles. The data would enable supervised training as the machine learning models would get to train using the previous, labelled examples.

To attain the quality of data the following steps were adopted:

- Cutting of unnecessary news items.
- The deletion of redundant or partial records.
- Very short articles and those that have poor information are sieved.

Such measures helped to increase the precision of the datasets and ensured that the model was not trained on irrelevant and non-representative data. data.

### 4.2 Data Preprocessing

Online sources of data tend to give a lot of noise and inconsistencies that may adversely influence the model performance. Hence, cleaning and standardizing of the text was done prior to feature extraction.

#### 4.2.1 Text Cleaning

The raw text was cleaned by:

- The characters were all done in small letters to prevent discrepancy.
- Doing away with periods and number, web addresses and symbols.
- Cutting out of unnecessary spaces and formatting characters. This increases consistency of documents and minimizes repetitions.

#### 4.2.2 Tokenization

The act of breaking up text into smaller units known as tokens that are normally words is referred to as tokenization. As an example, the sentence is divided into separate words. The use of tokenization allows the model to examine word-level features and compute word frequencies that are used to extract feature.

#### 4.2.3 Stop-Word Removal

Stop-words such as the, is, and, and of are prevalent in the text but lack a lot of meaning with them. Removing stop-words helps:

- Reduction in the dimensions of the features.
- Improve the calculating skills.
- Eliminate the irrelevant patterns of the model.

#### 4.2.4 Stemming and Lemmatization

Stemming and lemmatization cut words till their root and base. An example may be specified of running to running. This guarantees that the various grammatical variations of a word share a common property and in effect this guarantees that there is less generalization and sparsity of the feature space.

### 4.3 Feature Extraction

Machine learning algorithms use only numerical data; hence, non-numeric data should be transformed to numerical data. One of the most important steps is feature extraction which produces linguistic properties used in the process of detecting misinformation.

#### 4.3.1 TF-IDF Vectorization

Term Frequency-Inverse Document Frequency (TF-IDF) would be used to numerically represent text in such a research. The words are weighted by TF-IDF based on:

- Word frequency in a document.
- The frequency of the word in the entire data.

The more frequently used words in the fake news in comparison to the real news are given greater weight. TF-IDF is chosen because of the ease of computation, interpretability and text classification.

## 4.4 Machine Learning Models

Any form of classical machine learning algorithms have been contrasted in terms of performance and reliability.

### 4.4.1 Logistic Regression

The most popular known binary classifications algorithms that are utilized in making approximations about the likelihood that a piece of news is actually authentic and not fake is the logistic Regression with weighted features. It can be read and simple in high dimensional, text, information in conjunctions with TF-IDF characteristics.

### 4.4.2 Naive Bayes Classifier

Naive bayes classifier is a probabilistic model which is founded on the Bayes rule. Although it is independent of features, it does a good job in text classification because the word distributions are statistical. It is computationally efficient, and it is appropriate with large datasets.

### 4.4.3 Support Vector Machine (SVM)

Support Vector machine develops an optimum decision frontier that classifies fake and actual news articles with a greater margin. SVM works effectively in sparse and high-dimensional feature space which makes it a powerful misinformation detector of text-based misinformation.

## 4.5 Model Training and Testing

The dataset was divided into:

- The training models are trained on 70% of training data.
- Performance evaluation in terms of 30% testing data.

Cross-validation was used to enhance the stability of the model and parameter tuning was used to minimize overfitting. The classifiers were all trained and tested under equal conditions in order to compare it fairly.

## 4.6 Performance Evaluation

The trained models were tested on the basis of the conventional performance measures: Accuracy - the general accuracy of predictions.

Precision - consistency of fake news forecasting. Recall - capacity to recognize fake news.

F1-score - accuracy versus the recall. Confusion matrix The correct and the incorrect.

The measures will be an overall analysis of model performance

## 4.7 Prediction Process

A system: in the case of a new news story, the system: Preprocesses the input text,

Generates TF-IDF features, Learned this classifier.

Foretells sincerity or falseness of the article.

This will allow the facilities of real time news checking to be involved.

## Result and Discussion

### 5.1 Experimental Setup

The experiments were performed using labelled data of genuine and fake news stories. Following the process of preprocessing and feature extraction with the help of TF-IDF, the data were split to training and testing ratio at 70:30. They used several machine learning classifiers (Logistic Regression, Naive Bayes, and Support Vector Machine (SVM)) and trained them using the training data and tested them using the test data.

The whole of the experiment was conducted using common computing equipment to confirm the possibility of implementing the proposed system both in academic and real-time prototype setting.

### 5.2 Overall Model Performance

The standard classification measures that were applied in order to measure the performance of the models included the accuracy, precision, recall, F1-score and the confusion matrix. The Support Vector Machine and the Logistic Regression were the most reliable of the classifiers that were tested. The model with the highest overall classification accuracy slightly over 91 suggests that the proposed method can be employed in the process of separating the real and fake news pieces with the only help of text properties.

**5.3 Classification Report Analysis**

The performance of the models in every class is outlined in the classification report:  
 Class 0 - Real News Class 1 - Fake News

The results have been summarized as follows:

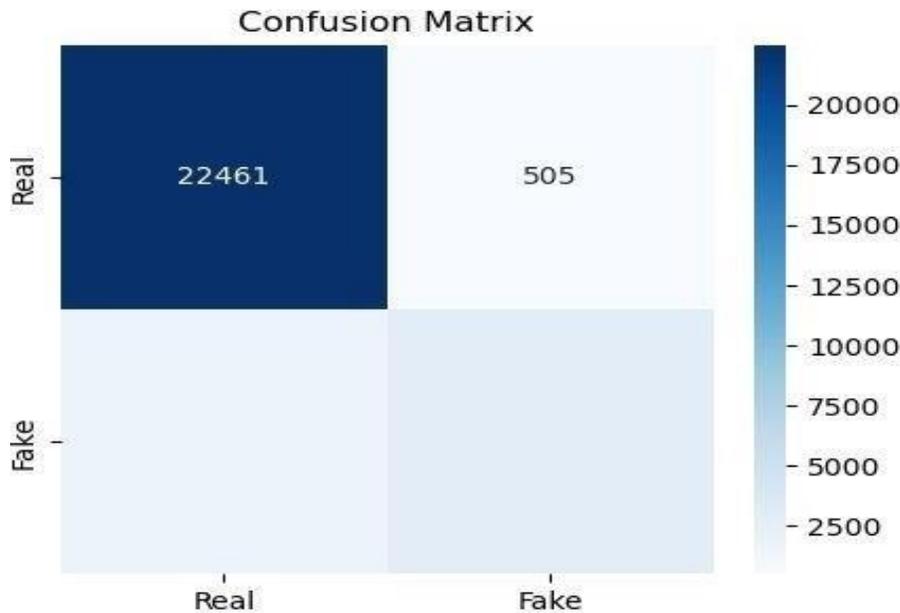
Metric	Real News (Class 0)	Fake News (Class 1)
Precision	0.92	0.85
Recall	0.98	0.60
F1-Score	0.95	0.70

**Interpretation:**

The error rate of fake news is high, the average among the 0.85s (0.85) represents the fact that the system does not predict an article as fake, and the prediction is correct in most cases. The recall of real news is very high (0.98) which indicates that the model is very effective in recalling real news articles correctly hence minimizing the chances of mislabeling real news articles. The less intense recall of fake news (0.60) indicates that deterrence of fake news actually is a misdetection of real new articles, which is usually very difficult in the context of misinformation detection because the writing style of fake news is not very misleading.

**5.4 Confusion Matrix Analysis**

The confusion matrix offers a visual level of the prediction results of the model, which shows correct and incorrect classification.



**Figure 1 Confusion Matrix**

**Future Scope**

Despite the fact that the suggested NLP-based misinformation detection system has excellent performance and real-life application, there are still some future improvement and development opportunities. Since

misinformation is currently more intricate and larger than ever before, the research efforts in the future must be aimed at enhancing accuracy, adaptability, and robustness without reducing computational efficiency.

**6.1 Deep Learning Models Training**

The new deep learning structures such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and transformer-based models such as BERT or RoBERTa can be used to develop work in the future. This type of model can encode a more fine contextual and semantic association between text which would, perhaps, enhance the recall of fake news detection. Transformer models can be explored to be computationally-intensive, optimized or fine-tuned to achieve improved performance. Although the proposed NLP-based misinformation detection system demonstrates very good outcomes and rather realistic

applicability, it also has several opportunities of further improvements and advancement. The misinformation continues to evolve in its complexity, scale, and power; hence, the future study should aim at aiming to increase accuracy, flexibility, and resiliency and still be computationally efficient.

## 6.2 Multi-modal false information Detection

The existing system itself is based on text only. The multi-modal data to be incorporated in future systems should include:

Pictures and videos related to news content.

Metadata like source of publication, time and the author information.

Interaction information of users such as shares, comments and pattern of interaction.

Text and non-textual information can be combined to achieve a high level of accuracy of detection and decrease false negativity.

## 6.3 Online Learning Systems

Patterns of misinformation change too fast and the fixed models become ineffective with time. Further studies can also consider the study of online and incremental methods of learning where the models keep on updating

themselves with the new labelled data. This would allow the system to change the approach to fake news as it appears and remain effective over time.

## 6.4 IntelligibleAI (XAI) Interaction.

Without the transparency and the trust, more sensitive applications of the system of detection of the misinformation use it. The subsequent uses of the applications might imply the Explainable AI (XAI) approaches to:

Name key words or phrases that are used in telling whether things are classified.

There are confidence scores and cause of prediction scores. Improve the system making it more reliable and readable.

The aspect of explainability is mostly significant on the circumstance of academic, legal and media situations relating to it.

## 6.5 Interlingual and Multilingual Development

Most detectors of fake news use language training data and domain specific training data to train them. Future work can focus on:

- Inter-rater cross-validation on dataset.
- The counterfeit news in the multilingual formula to be introduced in the global systems.
- Code-mixed language analysis and regional language analysis.

This would make the systems more generalized and extrapolated to other cultures and languages.

## Conclusion

The evolution that has overwhelmed digital media and online news platforms has resulted in the expansion of misinformation at a higher rate, which has presented critical challenges to the credibility of information and the level of trust, and informed decision making by people. The manual verification methods cannot support the enormous amount and rapid rate of online content anymore and automated detection systems are needed.

This study discussed this issue by presenting and analyzing an NLP-based machine learning model of detecting misinformation in news articles.

The suggested structure used a structured pipeline comprising of data collection, text preprocessing, extraction of features using TF-IDF, and classification that used classical machine learning models like the Logistic regression, Naive Bayes and Support Vector Machines. The creation of a solution, with the focus on the accuracy, computational efficiency, and interpretability was stressed to make sure that it would be appropriate in academic, educational, and prototype real-time contexts.

The experimental data showed that the suggested method had a total accuracy of about 91, which implies that it is very powerful in the differentiation of real news and fake news based on textual characteristics only. The model presented high levels of reliability when it came to determining true news articles and relatively acceptable levels of effectiveness in the detection of fake news. The results demonstrate that, even with the development of sophisticated deep learning models, classical machine learning models, in combination with effective NLP models, are still highly applicable and competitive.

Although the system is functional, there were some limitations observed such as low recall on fake news, use of text only data, and possible bias in the data. Such obstacles demonstrate the natural complexity of identifying the advanced misinformation that is highly resembling the legitimate journalism. However, the findings can be used as a solid basis of additional enhancements and expansions.

Summing up, this study establishes that machine learning methods based on NLP are lightweight, interpretable, and scalable and can have a meaningful impact on misinformation fighting. The proposed system can play a significant role in the area of fake news detection by providing an effective alternative to deep learning models that require a lot of resources. As the technology becomes more advanced in future, with the support of multi-modal analysis, explainable AI, and real-time deployment, these systems could facilitate fact-checking activities and enhance the consistency of information, as well as alleviate the effects of misinformation in the digital age at the societal level.

## References

1. Peng, X., Xu, Q., Feng, Z., Zhao, H., Tan, L., Zhou, Y., ... & Zheng, Y. (2024). Automatic news generation and fact-checking system based on language processing. arXiv preprint arXiv:2405.10492.
2. Khan, J., Ahmad, K., Jagatheesaperumal, S. K., Ahmad, N., & Sohn, K. A. (2024). Social Media Informatics for Sustainable Cities and Societies: An Overview of the Applications, associated Challenges, and Potential Solutions. arXiv preprint arXiv:2412.03600.
3. Abubakar, H. D., Umar, M., & Bakale, M. A. (2022). Sentiment classification: Review of text vectorization methods: Bag of words, Tf-Idf, Word2vec and Doc2vec. *SLU Journal of Science and Technology*, 4(1), 27- 33.
4. Gurrām, V. V. V. R., & Nalam, J. M. (2024). Automated Detection of Fake News in Natural Language Processing: A Comparative Study of TF- IDF and Lexical-Based Stance Detection with Logistic Regression.
5. Chavez, A. (2021). TF-IDF classification based Multinomial Naïve Bayes model for spam filtering (Doctoral dissertation, Dublin, National College of Ireland)