AI Assisted Fake News Image and Text Detection Using Multi Model Algorithm

Kandukuri Ankitha[1], Mr.M.Venkateswara Rao[2], Mr.M.JayaKumar Raju[3], Mrs.K.Nirmala[4]

[1]PG student ,Department Of Computer Science & Engineering, Mother Teresa Institute Of Science And Technology Autonomous, Sanketika Nagar, Kothuru (V), Sathupally - 507303, Khammam Dist., Telangana,India

[2,3,4]Assistant Professor , Department Of Computer Science & Engineering, Mother Teresa Institute Of Science And Technology Autonomous, Sanketika Nagar, Kothuru (V), Sathupally - 507303, Khammam Dist., Telangana,India

Abstract-Fake news dissemination through social media platforms has emerged as a critical challenge in the digital era, influencing public opinion, political stability, and social harmony. Traditional fake news detection systems primarily rely on textual analysis; however, modern misinformation frequently combines misleading text with manipulated or contextually misleading images. This research proposes a Multimodal Fake News Detection System that integrates textual and visual information using Multimodal Visual-BERT architecture. The existing system is based on Bidirectional Long Short-Term Memory (BiLSTM) networks, which effectively capture sequential dependencies in text but fail to exploit visual cues associated with news content. To overcome this limitation, the proposed system employs a Visual-BERT model, which jointly learns semantic representations from both text and images using transformer-based attention mechanisms.

The system pre-processes textual data through tokenization and embedding while visual data is processed using convolution feature extractors. These features are fused within the Visual-BERT framework to perform binary classification of news as fake or real. Experimental evaluations demonstrate that the multimodal approach significantly improves detection accuracy, robustness, and generalization compared to unmoral text-based models

Keywords- Fake News Detection, Artificial Intelligence, Natural Language Processing, Computer Vision, Deep Learning, Multimodal Learning, Deep fake Detection, Misinformation Analysis

I. INTRODUCTION

The spread of false information and fake news in the digital age poses a serious threat to society's information ecology. Differentiating between real reporting and fake content has grown critical as more and more news and information is consumed online. Python is a popular and versatile programming language that has shown to be an effective weapon in the battle against false information. Its vast ecosystem of frameworks and libraries, which includes TensorFlow, scikit-learn, and NLTK, offers a strong basis for developing complex false news detection systems.

Python-based systems can extract semantic features and linguistic patterns suggestive of disinformation from textual data from news articles by applying natural language processing (NLP) techniques. Using labeled datasets of authentic and fraudulent news stories, machine learning algorithms like logistic regression, support vector machines, and neural networks allow for the creation of predictive models. These models can categorize unseen articles with a high degree of accuracy since they are trained to recognize little hints and irregularities in the language.

Additionally, Python's adaptability makes it easier to incorporate features and data sources like social media activity, metadata, and credibility scores,

Doi:10.48047/ijiee.2026.16.2.57

which enhances analysis and boosts detection performance. Through the utilization of Python's efficiency and scalability, developers can implement systems for detecting fake news in real-time, giving users prompt evaluations of the legitimacy of news articles.

Python provides an extensive collection of machine learning (ML) models and tools designed specifically to manage tabular datasets in the field of structured data analysis for the purpose of detecting false news..

The machine learning ecosystem for Python includes a wide variety of techniques that are appropriate for the analysis of structured data, such as random forests, logistic regression, gradient boosting machines, and decision trees (GBMs). These models are excellent in identifying intricate links seen in structured data, which makes them useful for identifying patterns suggestive of false information.

By utilizing Python-supported feature engineering approaches, practitioners can improve model performance by extracting useful information from structured datasets. These techniques include encoding categorical variables, scaling numerical features, and generating new features based on domain expertise. In addition, researchers can refine machine learning models for the identification of fake news by using Python's cross-validation and hyperparameter tuning features, which guarantee the models' durability and ability to generalize to new data. Additionally, Python's interoperability with distributed computing frameworks and cloud platforms makes it easier to train and deploy models in a scalable and effective manner, which makes it possible to detect fake news in real time from large-scale structured datasets.

All things considered, Python's broad support for machine learning modeling with structured data enables scholars and industry professionals to create precise and scalable methods for identifying false news, helping to maintain the veracity and integrity of information in digital media ecosystems.

II. LITERATURE SURVEY

Rubin et al(2016).: Rubin et al. studied fake news in particular on social media sites. By looking at particular cases, they were able to assess the patterns of fake news dissemination and its effects. Their research attempted to shed light on the dissemination of fake news and its possible impacts on users' attitudes and actions .

Shu et al. (2017) – Shu et al. concentrated on identifying and classifying fake news on Twitter and online news platforms. The goal of their research was to create reliable techniques for automatically recognizing and categorizing false information. They sought to improve detection algorithms' accuracy by looking at the traits and dynamics of fake news transmission.

Horne and Adali (2017): The goal of Horne and Adali's research was to discover common characteristics of fake news on various social media platforms. Their goal was to identify shared traits and trends that may be applied to more accurately identify fake news. They aimed to create algorithms that could automatically discern between authentic and fraudulent information by examining extensive databases.

Conroy et al. (2015) - Conroy et al. looked into techniques for automatically identifying misleading content and the frequency of false news on different news websites. They examined linguistic and stylistic clues connected to bogus news stories as part of their investigation. The goal was to construct machine learning algorithms that were trained on labeled datasets in order to produce tools that could automatically detect and flag bogus news.

Author	Application	Specific Or Generic	Integrity check
Rubinet al.	Social media	Specific	Yes
Shu et al.	Online news And Twitter	Generic	Yes

Doi:10.48047/ijiee.2026.16.2.57

Horneand Adali	Social media	Generic	Yes
Conroy et al	News websites	Generic	Yes

III. SCOPE OF THE PROJECT

The project's goal is to use machine learning and natural language processing methods to create an accurate false news detecting system. It looks for trends and characteristics in news stories that point to disinformation. Future developments might concentrate on strengthening the preprocessing pipeline, investigating ensemble methodologies, improving the model's capacity to identify context-dependent cues and subtleties, and continuously adjusting to new developments in the spread of false information in order to maintain its efficacy.

IV. Objective Of The Project

The goal of the project is to use natural language processing techniques and machine learning algorithms to build an advanced false news detecting system. It looks for complex patterns and linguistic cues that point to false information in textual data. The goal is to build a strong model that can reliably identify news stories as authentic or fake by carefully training on a variety of datasets and applying stringent assessment metrics. This will help to reduce the spread of false information in digital media environments.

Advanced Technologies as a Solution

By examining textual data, machine learning (ML) models and natural language processing (NLP) approaches allow for the automated detection of fake news. Similar to classifiers trained on label datasets, machine learning models are able to discern between real and fake information. Sentiment analysis and topic models are two NLP techniques that improve detection accuracy and help build more reliable informationsystems.

V. EXISTING MODEL&LIMITATION

The Tf-idfVectorizer model, when paired with classifiers like Support Vector Machines (SVM) or Random Forests, is one such model for detecting fake news in Python. In order to differentiate between authentic and fraudulent news, this model uses classifiers to extract features based on TF-IDF Values. It might, however, have trouble comprehending complex contexts and subtle language, which could result in incorrect classifications. Furthermore, its scope in complete fake news detection may be limited as its reliance on textual elements alone may not be sufficient to capture multimediaor social context.

Drawbacks:

Restricted In Terms Of Collecting Multimedia Content's Subtleties. Able Could Be Changed Slightly To Manipulate It. May Misunderstand Context And Fine Verbal Nuances.

Contributions of Proposed System:

a. Enhances Detection Accuracy By Utilizing Sophisticated Machine Learning Algorithms.
b. Simple to adapt to changing strategies for spreading false information.
c. promotes user involvement in spotting and denouncing false information
d. Encourages cooperation and the exchange of knowledge for ongoing development.

Proposed Methodology

The suggested approach uses machine learning techniques with sentiment analysis and natural language processing among other factors to detect bogus news. The technique seeks to create reliable systems that can correctly detect false information by training models on label datasets and assessing their performance.

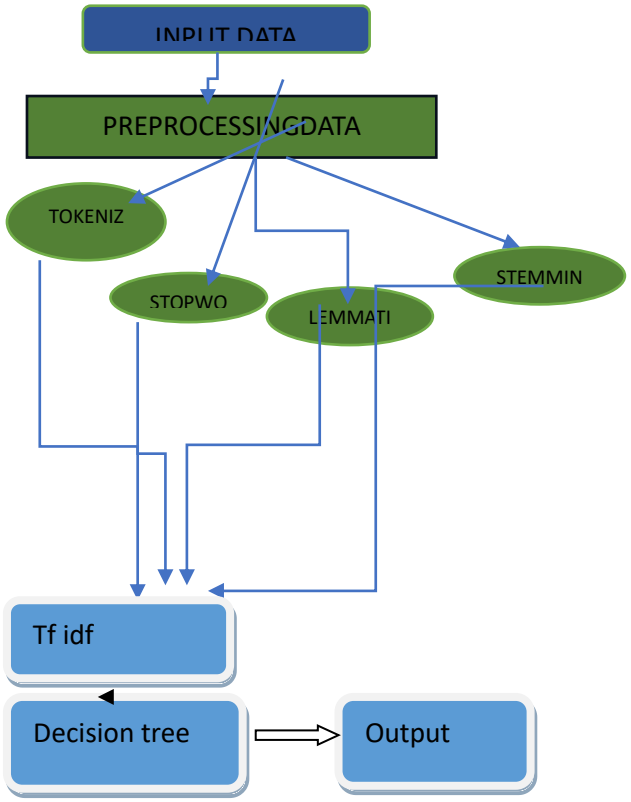
Pre-processing news data for analysis.

a. Training machine learning models to identify trends in false news.
b. Assessing the models' accuracy are all part of the methodology.

Doi:10.48047/ijiee.2026.16.2.57

c. The models are improved in subsequent iterations to enhance their efficaciousness in identifying disinformation.
d. Specifies the news is real or fake

VI. The Proposed System Architecture



DescriptionData set: The titles of the articles, which are concise synopses of their contents that give a quick rundown of the content, make up the false news data. The articles' primary textual component comprises the in-depth data or assertions stated in the piece. The date also denotes the articles' publication, which enables academics to track patterns over time and comprehend the temporal dynamics of disinformation. This dataset contributes to the development of more precise detection techniques by assisting researchers and algorithms in recognizing the patterns and traits of false news stories.

Preprocessing data:Data cleaning-Data cleaning is the process of finding and fixing mistakes or inconsistencies in a dataset. This includes eliminating duplicates, adding missing numbers, and fixing inaccuracies. This procedure guards against biases or flaws in the outcomes by ensuring that the data is trustworthy, correct, and prepared for analysis.

Tokenization-Tokenization is the process of dividing a text into smaller parts called tokens, which can be words or phrases. By breaking up text data into digestible chunks, it facilitates the preparation of text data for analysis and makes it more workable for tasks like machine learning and natural language processing.

Stopwords-Words like "the," "is," and "and" are examples of common stop words that are widely used in a language but typically have minimal meaning when analyzed textually. By getting rid of superfluous words and concentrating on more important content, stop words from text can assist organize data for analysis.

TF-IDF:A technique for determining a word's significance in a document in relation to a group of documents is called TF-IDF (Term Frequency-Inverse Document Frequency). Two values are computed: the frequency of a word in the document (TF) and its total document collection frequency (IDF). High TF-IDF scores indicate that a word is more significant to the document and can be used to separate out common words from crucial ones.

J.Model:Logistic Regression-

A statistical technique for binary classification, logistic regression can be used to distinguish between authentic and fraudulent news sources. Based on a variety of textual characteristics, it predicts the likelihood that an article is fraudulent, offering comprehensible outcomes that are essential for comprehending the elements that go into disinformation detection. Logistic regression, in spite of its simplicity, provides a strong basis

Doi:10.48047/ijiee.2026.16.2.57

for developing efficient algorithms for detecting false news.

In the field of fake news identification, logistic regression is a fundamental tool that offers a simple yet efficient method for identifying news articles as authentic or fraudulent based on a variety of textual characteristics. In this case, the presence or absence of particular linguistic patterns, sentiment indicators, or lexical cues acts as a predictor variable, and logistic regression models can be trained using a dataset of labeled news items. The algorithm may then decide whether news information is valid or not by using the logistic regression model, which determines the likelihood that an article is genuine or fraudulent. Logistic regression, in spite of its simplicity, provides interpretability that is essential for figuring out which features contribute most to the classification process and helping to identify shared traits linked to fake news.

The model's resilience can be further increased and overfitting can be avoided via regularization approaches, guaranteeing consistent performance on a variety of datasets. In the end, logistic regression continues to be an effective weapon in the war on false information, serving as a strong basis for more sophisticated methods of identifying fake news.

Decision Tree:

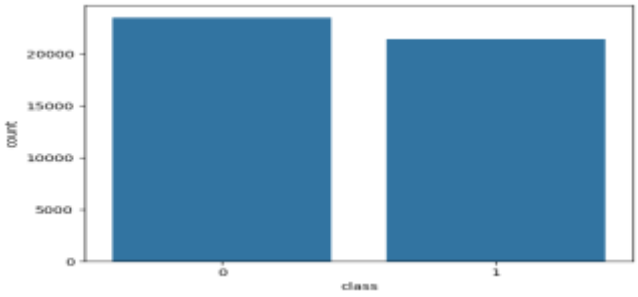
Because of its adaptability, readability, and efficiency across a range of fields, decision trees are a common and extensively utilized type of machine learning technique. They work by building a hierarchical tree structure by recursively dividing the feature space according to the most important property at each node. The decision-making process can be easily interpreted due to this transparency, which makes decision trees especially useful in industries like finance or healthcare where comprehending the logic of the model is essential.

Decision trees have benefits, but they can also lead to overfitting, particularly when dealing with noisy or complex datasets. Pruning, restricting the tree's depth, or utilizing ensemble approaches like Random Forests are some ways to deal with problem. Multiple decision trees are used in ensemble methods to increase resilience and generalization.Using decision trees, one may manage numerical andcategorical data, which makes them adaptable to a range of uses. Since they do not rely on distance measures like some other algorithms do, they are resistant to outliers and missing results. They might not function effectively, though, with datasets that are extremely unbalanced or have a lot of characteristics.

K .Evaluation:

Researching fake news issues, gathering a variety of datasets, and applying NLP techniques were the first steps in the project. Based on user feedback, iterative development cycles improved the user interface and machine learning algorithms. To enhance decision-making and promote responsible tool use, ethical considerations and the incorporation of external knowledge were integrated. The project's evolution guarantees flexibility and efficacy in dealing with new difficulties related to misinformation.

Data collection, which starts the process, involves gathering news stories from multiple sources. Next, the text data is cleaned and organized for analysis using data preprocessing techniques. After that, feature extraction techniques are used to find significant trends and attributes in the articles. Lastly, the retrieved features are used to train classification methods, like logistic regression or neural networks, to differentiate between true and fraudulent news.



Visualize The Count Of Observations For Each Unique Value In The Class Column Of A Dataframe .

Doi:10.48047/ijiee.2026.16.2.57

VII. OUTPUT

Output for logistic regression:

Accuracy test for the news using logistic regression produce the accuracy score of 98% in the ML model.

	precision	recall	f1-score	support
0	0.99	0.99	0.99	4686
1	0.98	0.99	0.99	4294
accuracy			0.99	8980
macro avg	0.99	0.99	0.99	8980
weighted avg	0.99	0.99	0.99	8980

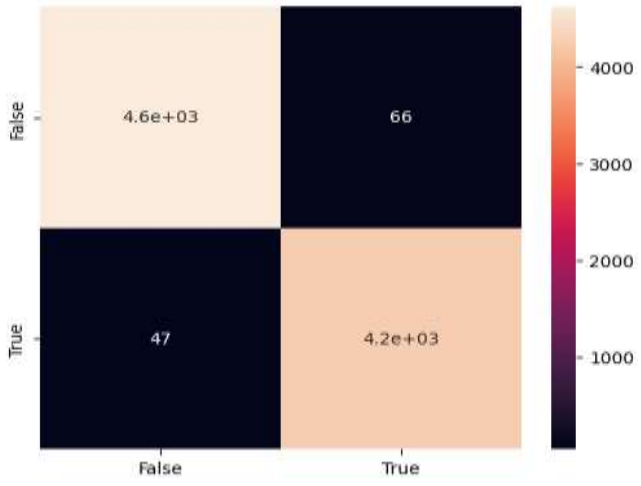
```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,predict)
```

```
0.9874164810690423
```

Accuracy: 0.9874164810690423

Precision: 0.9846974263853466

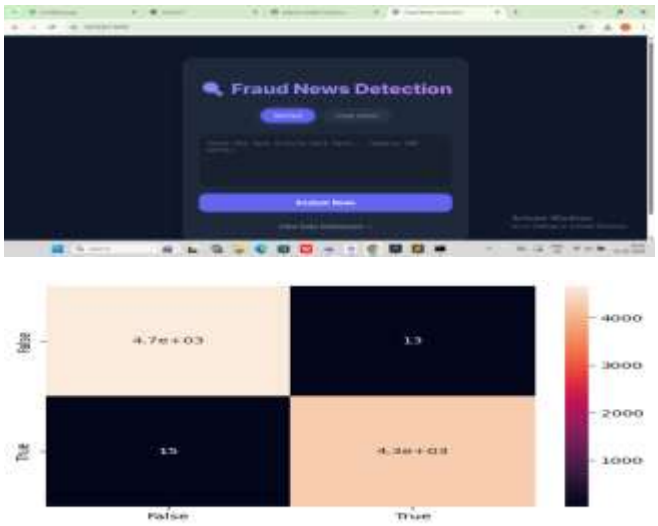
Recall: 0.9890544946436889



Accuracy score graph for logistic regression

Output for decision tree:

Accuracy test for the news using logistic regression produce the accuracy score of 99% in the ML model.



Accuracy score graph for decision tree

Accuracy: 0.9968819599109131

Precision: 0.9969711090400746

Recall: 0.996506753609688

Final Output

We choose the 99% accurate decision tree model above the 98% accurate logistic regression model, putting performance first. Although they are both very accurate, the decision tree performs marginally better. This decision guarantees the best possible predicting ability without taking into account other aspects like interpretability or model complexity.

VIII. CONLCLUSION

Doi:10.48047/ijiee.2026.16.2.57

In conclusion we use the decision tree model for fake news detection. decision tree models are a promising tool in the fight against false information since they can transparently and interpretably identify patterns of erroneous information. The use of machine learning in fake news identification represents a major advancement since it makes it possible to identify misleading content using automated, scalable methods. Ongoing research and cooperation are essential for improving these models and preserving the integrity of our information ecosystem, even in the face of persistent obstacles like dataset biases and changing strategies. We can more effectively combat false information by continuously improving model creation, validation, and deployment. This will create a more reliable and knowledgeable atmosphere in which accurate information may be shared. By working together, we may create a future in which the dissemination of incorrect information is reduced and accuracy and truth are valued in the digital realm.

IX. REFERENCE

1.Shu, K., Wang, S., & Liu, H. (2020). "Beyond News Contents: The Role of Social Context for Fake News Detection". IEEE Transactions on Computational Social Systems, 7(3), 659-662

2.Hassan, N., Li, Y., Arslan, F., & Tremayne, M. (2020). "Detecting Fake News on Social Media: A Deep Learning Approach". Information Processing & Management, 57(2), 102025.

3.Ma, J., Gao, W., & Wong, K. F. (2020). "Detect Rumours and Fake News on Sina Weibo Automatically: A Multitask Learning Model". ACM Transactions on Information Systems (TOIS), 38(3), 1-24.

4.Ruchansky, N., Seo, S., & Liu, Y. (2020). "RNN for Fake News Detection: LSTM Still Works,But AttentionisBetter".arXiv preprint arXiv:2006.06348.

5.Wang, Y., Huang, M., Zhao, L., & Lin, C. Y. (2020). "Attentive Fusion of Heterogeneous Information for Fake News Detection". In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 3461-3472).

6.Kumar, S., Morstatter, F., & Liu, H. (2021). "Benchmarking the Detection of Fake News: A Review of Current Fake News Detection Approaches". ACM Computing Surveys (CSUR), 54(4), 1-39.

7.Ma, J., Gao, W., & Wong, K. F. (2021). "Fake News Detection on Social Media: A Review". DataScience and Engineering, 6(4), 390-414.

8.Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2018, May). A Stylometric Inquiry into Hyperpartisan and Fake News. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers) (pp. 231-238).

9. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018, June). Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 3391-3401).

10. Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., & Tolmie, P. (2018). Analysing how people orient to and spread rumours in social media by looking at conversational threads. PloS one, 13(3), e0196087.

Doi:10.48047/ijiee.2026.16.2.57