

ClariNews-Fake News Detection using ML

Sharayu Sangekar¹, Jahanvi Tiwari², Uday Deshmukh³, Umang Gotmare⁴, Parth Deshpande⁵ and Om Bhate⁶

¹⁻⁶Yeshwantrao Chavan College of Engineering, Nagpur, India
Email: sharayu138@gmail.com

Abstract— The online environment now faces its most severe threat because people create false reports and articles. Social media platforms and online platforms together create a fast-growing platform which enables misinformation to spread and shape public opinions at a speed that exceeds human detection abilities. People need automated solutions because manual news verification becomes impossible when dealing with overwhelming news content. ClariNews stands as the solution that enables news verification through its news verification system which was developed as a news verification system. The detection of fake news stories has been addressed by the combined utilization of machine learning and natural language processing (NLP). It was designed to handle both text content and visual elements which enables it to identify even the most complex visual changes. It enables user testing through its ability to examine both article content and accompanying images which allows the system to confirm whether an image remains untouched or has been modified. ClariNews allows users to pick specific fields which include politics and sports and health to attain their most exact outcomes. It strengthens itself to multiple languages, including English, Hindi and Marathi, which will increase its reach among a wider public. The detection system delivers results to users while showing them how to identify fake news stories. The organization ClariNews functions as a basic tool which enables society to achieve advanced social awareness together with improved educational results. The project demonstrates that digital media awareness needs to be understood through multiple languages and image verification methods. The research combines learning-based methods with natural language processing to enhance its ability to detect fake news.

Index Terms— Detection of Fake News, SSL, Cross-Lingual Text Analysis, Text Classification, Media Literacy.

I. INTRODUCTION

The present times show that digital content distribution systems deliver their content to users in all regions within extremely quick timeframes. Digital content distribution systems currently face their most urgent challenge because misinformation now spreads through digital channels at the same speed as legitimate information. People worldwide depend on digital communication platforms and online resources as their primary means of obtaining instant updates which results in the unwitting dissemination of false information that disrupts public perception and governmental credibility and social group dynamics [2]. The manual fact-checking system needs to develop advanced computational systems because it fails to handle today's data distribution speed. ClariNews functions as an advanced multimodal detection system which analyzes textual data through natural language processing and evaluates visual content through image recognition technology.

The system operates on more than text-based input which traditional systems use because it processes both written material and visual elements. The system uses this method to grasp complete information context while it detects false information through more precise and fast methods used in current digital environments. The system uses three methods from its technology stack which includes Logistic Regression and Support Vector Machines and Convolutional Neural Networks to identify false content and study variations between different media platforms. The system uses a comprehensive training dataset which includes data from Kaggle and BuzzFeed and Weibo and multiple social network platforms to maintain its performance in actual situations. The model's primary strength lies in its ability to adjust to new languages and various subjects without facing major challenges. The three languages which represent the primary languages spoken by Indian people are used by Clarinews to test their cases in Hindi Marathi and English.

II. KEY OBJECTIVES OF THE CLARINEWS SYSTEM

The ClariNews system is designed with three main components that collectively outline its core objectives and overall functionality. The system uses a model which can identify false information through its multi-modal detection method which combines NLP. The model transforms text data into TF-IDF-based feature vectors which it uses to test fake news detection through Logistic Regression and Support Vector Machine algorithms.

The system provides an advanced verification service which supports English, Hindi, and Marathi languages and can detect misinformation in both regional and low-resource language settings. The system uses live news APIs to monitor current news and trending news which helps maintain accurate information for readers. The output would provide explanations which users could easily understand because it would include confidence scores and contextual information that would help achieve transparent digital media education.

II. LITERATURE REVIEW

Your training experience extends from data which existed until the cutoff month of October in the year 2023[1]. The study conducted in 2024 identified false information through machine learning methods which researchers used to analyze articles. Research study introduced an automated classification system which functions through web-based technology but failed to provide information about the datasets used in the research. Their exceptional work output achieved through their assessments of higher performance standards [2,3,4]. The development of fake news detection through textual analysis can achieve effective results when using traditional ML models. The proposed system has unified capabilities which restrict its performance because it fails to process multimodal content which uses visual and audio material that people commonly use in contemporary deceptive news. The dataset suffers from non-transparency which creates obstacles for researchers who want to reproduce the research. The research presented in [5,6] worked on fake news detection through the development of an FDHN model which combines fuzzy logic and deep learning to improve its classification accuracy. This combination helps the system handle ambiguity in data more effectively while also improving its ability to predict false information accurately. The researchers tested the model using LIAR1 and LIAR2 datasets, which resulted in exceptional performance. Researchers performed ablation studies to confirm their results, which showed that the model outperformed multiple baseline models. The method demonstrates its main benefit through exceptional accuracy combined with superior ability to interpret results when compared to standard deep learning techniques [7,8]. In [9,10] The 2024 study established a new fake news detection method through its hybrid model which combined CNN with LSTM and FastText and transformer systems that included BERT XLNet and RoBERTa. The system achieved outstanding test results on WELFake and Fake News Net datasets by reaching an accuracy rate of approximately 0.99 and showing strong F1 performance. The results show that deep contextual understanding combined with explainable AI better enables fake news detection. The machine learning model which detects unreliable information was developed using data from 2023. The system applies Decision Tree analysis together with NLP-based feature extraction and vectorization methods to assess the authenticity of article content [10]. Twitter posts served as the primary source of social media data. The model achieved an accuracy of about 90 percent which shows that traditional machine learning methods still succeed at detecting fake content. The model comes with certain restrictions that limit its effectiveness. The system cannot process different types of information because it cannot handle visual material which requires authentication. The system needs additional data in order to function properly because its current design only supports one language and small datasets yet it fails to operate effectively in real-world conditions [11,12,13]. The research review conducted in 2022 [14,15] focused on comparing several machine learning models which included SVM, CNN, Naive Bayes, Decision Trees, and LSTM for detecting false information. The researchers used LIAR,

Twitter, Weibo, and BuzzFeed datasets for their experimental work. The assessment incorporated four performance measures which are accuracy and precision and recall and F1-score. System with highest accuracy LSTM achieved 82.29% accuracy, while SVM and Linear SVM produced the lowest accuracy results with their respective performance of 75% and 92%. Research team tested various algorithms through their study of Random Forest and Logistic Regression and KNN and SVC which they combined with NLP methods to analyze a Kaggle dataset for detecting false information. The assessment of performance used both accuracy and classification report measures as evaluation metrics [16]. The researchers used TF-IDF as their method for extracting text features which they applied to pre-processing tasks that produced positive results for classifier performance during classification. The researchers showed that traditional machine learning methods would deliver both reliable results and accuracy equivalent to modern methods. The 2022 paper Fake News Detection in social media. A review using Machine Learning techniques which appeared in the journal [17,18] demonstrates comprehensive research on various machine learning techniques that researchers assessed through testing on different data sets which included Facebook Twitter and Instagram social media platforms to identify questionable material. The discussion revealed that the different models had various performance paths which resulted in higher accuracy through the combination of content writing and imaging data handling. The research failed to examine current deep learning systems which use transformers because these systems represent a major improvement over existing models [19]. The models did well at extracting features, but sucked at handling the fake content. Research study of 2021 [19] investigates the methods which detect false information through its selected features. The main research goal of this study used three classifiers which included KNN and GEFes and Quantum KNN to select features for false information detection. The researchers conducted experiments using data from BuzzFeed and Facebook platforms. The accuracy values for KNN and QKNN were 91.3% and 84.4%, respectively. Results demonstrate that proper feature selection boosts model performance while simultaneously decreasing computing requirements [20]. The system has two main drawbacks because it uses small datasets and it lacks deep learning models which prevents the system from handling multiple large data sources effectively.

III. METHODOLOGY

The ClariNews system identifies false news through its entire system which integrates natural language processing with machine learning and computer vision methods. The complete system operates through three main components which start with text analysis and move to image analysis before delivering the final results. The method is as follows:

A. Collection of Data Sources

The researchers constructed a dataset by gathering information from credible sources, including Kaggle, BuzzFeed, the LIAR dataset, and news APIs. The dataset included text material which matched with image content that received true and false designations. The study used data from different platforms to achieve better generalization while reducing bias from specific sources.

B. Text Preparation and Engineering

The NLP process starts with text segmentation into smaller components, which are followed by the removal of stop words to maintain only relevant content. The process ends with lemmatization, which converts words to their fundamental dictionary forms. The text processed through the preprocessing step undergoes numerical mapping through TF-IDF vectorization which creates word [12,13].

C. Text Categorization Approach

Textual Data Classification employs two traditional machine learning techniques. The first method uses Logistic Regression (LR) which serves as a baseline classifier because of its straightforwardness and its ability to deliver reliable results [14]. The second method selects Support Vector Machine (SVM) because of its capability to handle high-dimensional sparse text data. The models use a supervised learning approach which divides the data into training and testing sets in an 80:20 ratio. The classifiers will indicate if the given news text is true or false.

D. Language-Adaptive Analysis Module

It can process articles from various languages which include English and Hindi and Marathi. The system achieves better accuracy through its process of content normalization and preprocessing which extracts essential features from all content.

E. Real-Time Model Deployment

It needs live news publications to obtain article content without any delay. The system uses Flask as its main backend framework and Flutter as its frontend solution to provide web and mobile access to users.

F. Model Analysis

The system's performance evaluation uses international standard classification criteria. The assessment process uses four measurement standards which assess correct classification rates, track predicted positive case relevance, actual positive case identification and measure the harmonic mean between precision and recall. The confusion matrix displays results through a visual system that shows results and calculates prediction accuracy while it detects real news and false news cases.

IV. FROM DATA TO DETECTION

The flow is shown as follows:

- Place Information from the Article: When the chief's mouth types in the address of an article on the GUI or into the URL text field via the web.
- Data-driven web platform: Web scraping techniques and News APIs provide essential support for the analysis of news articles from online sources. The system collects contextual information together with essential reference data from chosen news content which improves its analytical accuracy and reliability.
- Engineering of Data: To make text data apt for analysis, the process of cleaning may be a great task. The entire work may be exhausting but fruitful at the end with the help of following techniques:
 - The process of text analysis begins with the first step of Tokenizing which separates the text into its basic components of individual units [2-7].
 - The process of lemmatization transforms words into their fundamental base forms [2-7].
 - The process of TF-IDF computation provides a method to create text features through its computation [2-7].
 - Omitting occasional words that have scant importance in meaning [2-7].
 - Key elements needed for excellent accuracy in modeling are inferred [2-7].
 - Different ways exist to identify and get rid of overlapping or redundant news articles [2-7].

D. Topic-Wise Content Mapping

The system organizes data into three predetermined categories which include Entertainment, Sports, and Finance.

The domain classification system enables simultaneous extraction of distinct features from each domain which enhances verification accuracy.

E. False Information Verification Framework

The news detection framework employs natural language processing methods together with supervised learning techniques to examine articles which have been classified as false or true while using logistic regression and SVM models to assess how accurate they are.

F. Outcomes Determination

According to the model, the system will then judge whether the information is true or false.

G. Findings and Actionable Insights

System will both confirm the news and show authentic sources which readers can use to check the information. The system identifies the news as Fake through its warning system which needs proof from authentic verified sources.

H. Output Display

The user obtains final verification results and believability assessment methods together with workflow completion instructions through the interface without needing an interpreter.

I. Explainability of Results

The system provides brief explanations or key factors influencing the prediction, helping users understand why a piece of news is classified as true or false.

Fig.1 shows the system pipeline of ClariNews

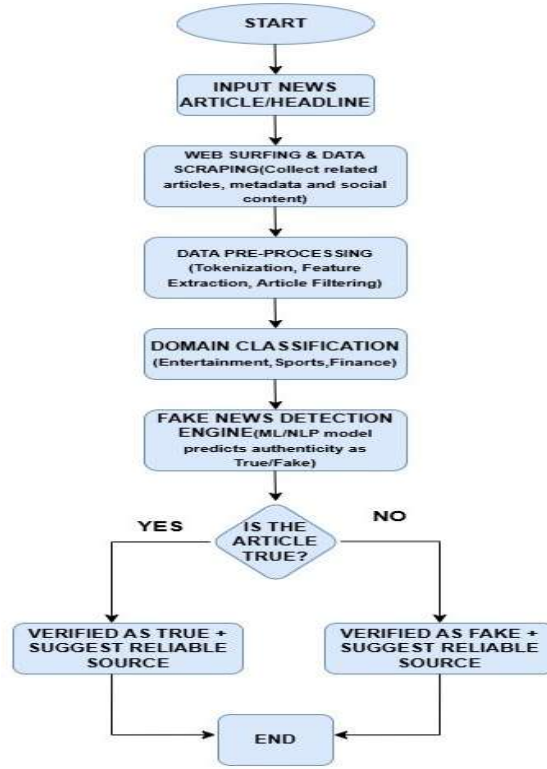


Fig.1.From Input to Insight

V. BROWSER-BASED APPLICATION

The Clarinews system functions as an advanced AI platform which detects fraudulent news articles while also supporting web and mobile applications and performing text and image examination. The system combines both machine learning and deep learning approaches, which include Logistic Regression, Support Vector Machines, and Convolutional Neural Networks as its models [9,10].

Application allows users to speak different languages because it supports three languages which include English and Hindi and Marathi. The system creates training materials through data collection from multiple sources including Kaggle and Buzzfeed and Weibo and various social media platforms which enable users to conduct instant result verification. Clarinews uses Natural Language Processing technology and computer vision systems to verify the authenticity of news articles. Table I shows the key features and functionalities of ClariNews.

TABLE I. KEY FEATURES AND FUNCTIONALITIES

Sr. No	Key Features	Functionalities
1.	Multimodal Identification	The multimodal detection system with NLP and CNN technologies is supposedly capable of combining both text as well as image analysis implements.
2.	Multi-language analysis	Social media users can increase their national reach by adding validation in six languages (English, Hindi, Marathi, Malayalam, and Punjabi).
3.	Machine Learning Integration	The system requires improvement through SVM and LR and CNN because this combination will reduce variance and prevent overfitting.
4.	Category-Based Classification	The software classifies news content into three categories, i.e. Entertainment, Sports, and Finance.
5.	Real-Time Verification	The system checks news instantly using live data to provide quick and reliable verification results.
6.	Contextual Explanation	It helps them gain trust and helps them better understand concepts.
7.	User Interface	Flutter as its front-end framework (better user experience) while Flask serves as the back-end (computation and task management).
8.	Scalable Training	Sources include Kaggle, Weibo and social networks, which provide extensive user reach combined with strong model performance capabilities

VI. SYSTEM PERFORMANCE REVIEW

After close observation throughout the evaluation system, we took a moment to evaluate the ClariNews (fake news detection model) with general metrics to establish its accuracy as described in the text. The following evaluation metrics are defined in the context of this study:

- TP (TRUE POSITIVE): The volume of times when a label has appropriately been given to fake news [2,11,16].
- TN (TRUE NEGATIVE): Cases in which genuine news content is accurately recognized as real. [2,11,16]
- FP (FALSE POSITIVE): Examples illuminate the way of how fake news is wrongly identified to be real news. [2, 11, 16]
- FN (FALSE NEGATIVE): Examples illuminate the way of how fake news is wrongly identified to be real news. [2, 11, 16]

A. Assessing Detection Quality

The system demonstrates its advanced capabilities through its successful identification of false reports which uses its natural language processing and its ability to analyze multiple languages. Research demonstrates that transformer-based models achieve accurate results yet their computation requirements remain excessive, making them unsuitable for applications which require real-time processing. Popular method of verification because it strikes the best balance among three key factors which include verification accuracy and system performance and real-time usability in the verification process of text-based news content [20]. Table II shows the performance comparison of the machine learning models in the system:

TABLE II: MODEL PERFORMANCE

MODEL	A (Accuracy)	P (Precision)	R (Recall)	F1 (F1-Score)
LR	82%	78%	73%	75%
SVM	85%	80%	75%	78%

B. Model Outputs

The overall performance metrics of the model are summarized in Table III as shown below:

TABLE III. EVALUATION RESULTS OF THE SYSTEM

METRIC	VALUE
Accuracy (A)	85%
Precision (P)	80%
Recall (R)	75%
F1-Score (F1)	78%

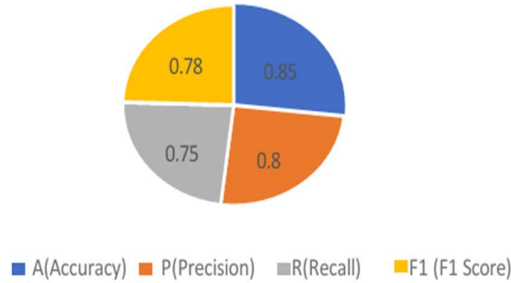


Figure 2. Comparative View of Model Accuracy Measures

VII. SYSTEM OUTPUT ANALYSIS

The results from ClariNews show accurate and reliable performance because the system marks TRUE NEWS and FAKE NEWS according to the testing results shown in figures 3 and 4:

The outputs are displayed within Fig. 3 and 4, which show how system processes both authentic news and fabricated content. The system starts from a base that uses NLP methods for cleaning text and extraction of features before it proceeds to train classifiers through machine learning techniques. The machine evaluates news content by using the linguistic and semantic features it has collected from the content, and then produces a confidence score for its prediction, which shows how much trust the machine places in the news content's

authenticity. The experimental results prove that the proposed method succeeds in detecting false news from digital news environments while achieving both high accuracy and operational efficiency.



Figure.3. True News



Figure.4. Fake News

The outputs are displayed within Fig. 3 and 4, which show how system processes both authentic news and fabricated content. The system starts from a base that uses NLP methods for cleaning text and extraction of features before it proceeds to train classifiers through machine learning techniques. The machine evaluates news content by using the linguistic and semantic features it has collected from the content, and then produces a confidence score for its prediction, which shows how much trust the machine places in the news content's authenticity. The experimental results prove that the proposed method succeeds in detecting false news from digital news environments while achieving both high accuracy and operational efficiency.

VIII. CONCLUSION

This paper presents ClariNews, a system which uses machine learning techniques to detect deception while employing natural language processing for fact verification. The system detects false information in real-time because the machine learning framework achieved 85 percent accuracy while maintaining reliable precision and recall and F1-score results. It which enables users to interpret results while classifying content in multiple languages has become more user-friendly while it assists users in developing digital media literacy skills. The research will focus on developing secure verification systems which can be easily scaled through the use of deep learning and transformer models and Indian language localization. Fig.5 illustrates the comparison of model performance metrics (overview).

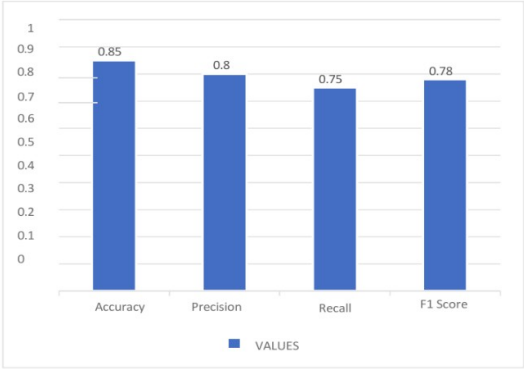


Fig.5.Overview of Metric Values

ACKNOWLEDGMENT

Numerous individuals provided essential assistance which enabled us to complete this project. The team wants to express their gratitude to the analysts and fact-checkers and researchers who provided valuable assistance in creating a system that produces reliable and precise outcomes. This platform development process received valuable input from academic experts and digital literacy advocates and policymakers who provided their expert

feedback. The testing and evaluation phase required specific suggestions from users which helped to create the complete design of ClariNews including its usability and system structure. This project started at Yeshwantrao Chavan College of Engineering in Nagpur with strong assistance from the Department of Computer Technology and its faculty members. Their continuous support and guidance proved essential during the entire research and development stages.

REFERENCES

- [1] N. F. Baarir and A. Djeflal, "Fake News detection Using Machine Learning" The 2020 2nd International Workshop on Human Centric Smart Environments for Health and Well-being (IHSH) will take place in February 2021.
- [2] J. Shaikh and R. Patil, "Fake News Detection using Machine learning," IEEE Xplore, Dec. 01, 2020.
- [3] S. I. Manzoor, J. Singla, and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," IEEE Xplore, Apr. 01, 2019.
- [4] Jain, A. Shaky, H. Khatter, and A. K. Gupta, "A smart System for Fake News Detection Using Machine Learning," 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Sep. 2019.
- [5] K. Hiramath and G. C. Deshpande, "Fake News Detection Using Deep Learning Techniques," 2019 1st International Conference on Advances in Information Technology (ICAIT), Jul. 2019.
- [6] M. F. Mridha, A. J. Keya, Md. A. Hamid, M. M. Monowar, and Md. S. Rahman, "A Comprehensive Review on Fake News Detection With Deep Learning."
- [7] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting Fake News using Machine Learning and Deep Learning Algorithms," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Jun. 2019
- [8] Abdulrahman and M. Baykara, "Fake News Detection Using Machine Learning and Deep Learning Algorithms," IEEE Xplore, Dec. 01, 2020.
- [9] R. K. Kaliyar, "Fake News Detection Using A Deep Neural Network," IEEE Xplore, Dec. 01, 2018.
- [10] A. Jain and A. Kasbe, "Fake News Detection," IEEE Xplore, Feb 01, 2018. 23[11] S. Kumar and B. Arora, "A Review of Fake News Detection Using Machine Learning Techniques," IEEE Xplore, Aug 01, 2021.
- [11] S. Gilda, "Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection," IEEE Xplore, Dec. 01, 2017.
- [12] S. A. Alameri and M. Mohd, "Comparison of Fake News Detection using Machine Learning and Deep Learning Techniques," IEEE Xplore, Jan. 01, 2021.
- [13] V. M. Kresnakova, M. Sarnovsky, and P. Butka, "Deep learning methods for Fake News detection," 2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE
- [14] International Conference on recent achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTIMACRo), Nov. 2019.
- [15] S. Aphiwongsophon and P. Chongstitvatana, "Detecting Fake News with Machine Learning Method," IEEE Xplore, Jul. 01, 2018
- [16] Fake News Detection Using Content-Based Features and Machine Learning | IEEE Conference Publication IEEE Xplore
- [17] S. B. S. Mugdha, S. M. Ferdous, and A. Fahmin, "Evaluating Machine Learning Algorithms Bengali Fake News detection," IEEE Xplore, Dec. 01, 2020.
- [18] 2022 4th International Conference on Advances in Computing Communication Control and Networking (ICAC3N)
- [19] 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI) 24
- [20] 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)