International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

Automated Fake News Detection System Using Machine Learning

[1] Mr.Farooqhussain Mohammad , [2] B.Varsha, [3] E. Sri Charan, [4] p.Sravan Kumar
[1]AssistantProfessor, [234]Students
Department of Computer Engineering(Internet Of Things)
Siddhartha Institute of Technology & Sciences, Narapally
farooqhussain.cse@siddhartha.co.in , 23tq1a6913@siddhartha.co.in, 23tq1a6903@siddhartha.co.in, 23tq1a6910@siddhartha.co.in

Abstract

In the modern digital era, the rapid growth of online platforms and social media has significantly increased the spread of information. However, this has also led to the widespread dissemination of fake news, which can mislead people, influence public opinion, and create social and political instability. Detecting fake news manually is a challenging and time-consuming task due to the large volume of data generated every day. Therefore, there is a need for an automated system that can accurately identify and classify fake news.

This project focuses on developing a Fake News Detection system using Machine Learning and Natural Language Processing (NLP) techniques. The system analyzes textual data from news articles and classifies them as real or fake based on their content. Initially, the dataset is preprocessed using techniques such as tokenization, stop-word removal, and text normalization. The processed text is then converted into numerical form using feature extraction methods like Term Frequency–Inverse Document Frequency (TF-IDF).

Machine learning algorithms such as Logistic Regression and Naive Bayes are implemented to train the classification model. These models learn patterns from the dataset and predict the authenticity of unseen news articles. The performance of the system is evaluated using metrics such as accuracy, precision, recall, and F1-score, which help in assessing the effectiveness of the model.

The results demonstrate that the proposed system can efficiently classify news articles with high accuracy and reliability. This project highlights the importance of machine learning techniques in combating misinformation and supports the development of intelligent systems for improving information authenticity. The proposed system can be further extended for real-time applications and integrated into social media platforms to reduce the spread of fake news.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

I. Introduction

In the current digital era, the rapid growth of the internet and social media platforms has led to an enormous increase in the dissemination of information. While this has improved access to news and knowledge, it has also resulted in the widespread circulation of fake news, which refers to false or misleading information presented as legitimate news. Fake news can influence public opinion, create confusion, and even lead to serious social, political, and economic consequences.

One of the major challenges faced by individuals and organizations is the difficulty in verifying the authenticity of news articles. Due to the large volume of data generated every day, manual verification of news content is impractical and time-consuming. As a result, fake news spreads rapidly across platforms such as social media, news websites, and messaging applications, making it difficult to control misinformation.

The main problem addressed in this project is to develop an automated system that can accurately classify news articles as real or fake based on their textual content. This is achieved by applying machine learning and natural language processing (NLP) techniques to analyze patterns, linguistic features, and contextual information present in the text.

In this project, a dataset containing real and fake news articles is used to train classification models. Techniques such as text preprocessing, feature extraction using TF-IDF, and machine learning algorithms like Logistic Regression and Naive Bayes are employed to build an effective prediction system. The system learns from historical data and predicts the authenticity of unseen news articles.

The objective of this project is to create a reliable and efficient fake news detection model that can assist users, media organizations, and online platforms in identifying misleading information.

II. Literature Survey

Fake news detection has emerged as a significant research area in the fields of machine learning, natural language processing (NLP), and data science due to the rapid growth of online information and social media platforms. Researchers have extensively studied various techniques to identify and classify fake news based on textual content and linguistic patterns.
Several studies have shown that machine learning algorithms such as Logistic Regression, Naive Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forest are widely used for text classification tasks, including fake news detection. These algorithms analyze features such as word frequency, sentence

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

structure, and contextual patterns to distinguish between real and fake news articles. Among these, Logistic Regression and Naive Bayes are commonly preferred due to their simplicity, efficiency, and effectiveness in handling large textual datasets. Natural Language Processing (NLP) techniques play a crucial role in fake news detection. Researchers have applied preprocessing methods such as tokenization, stop-word removal, stemming, and lemmatization to clean and prepare textual data. Feature extraction techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) and word embeddings are used to convert text into numerical form, enabling machine learning models to process and analyze the data effectively.

Recent studies have also explored the use of deep learning models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) for fake news detection. These models are capable of capturing complex patterns and contextual relationships in text, leading to improved prediction accuracy. However, they require larger datasets and higher computational resources compared to traditional machine learning models. Research findings indicate that the performance of fake news detection systems largely depends on data quality, feature selection, and model optimization.

III. System Analysis

System analysis focuses on understanding the challenges involved in detecting fake news and designing an efficient automated solution. With the rapid growth of digital platforms, a huge volume of news content is generated daily, making manual verification impractical. The system must be capable of processing large textual datasets and identifying patterns that distinguish real news from fake news. It requires the use of Natural Language Processing (NLP) techniques to clean and analyze textual data effectively. Feature extraction methods such as TF-IDF are essential to convert text into numerical form for machine learning models. The system also needs to support classification tasks using algorithms like Logistic Regression and Naive Bayes. Performance evaluation using metrics such as accuracy, precision, recall, and F1-score is crucial to ensure reliability. Additionally, the system should be scalable and efficient for real-time applications. Overall, system analysis ensures the development of a robust and accurate fake news detection system.

Existing System

The existing system for detecting fake news primarily relies on manual verification and traditional fact-checking methods. Journalists and fact-checking organizations analyze news content to determine its authenticity, which is time-consuming and labor-intensive. Some platforms use basic keyword-based filtering or rule-based systems, but these methods lack accuracy and fail to understand context. Existing systems are not capable of handling the massive volume of online data generated

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

every day. They also struggle to detect subtle misinformation, sarcasm, or misleading headlines. Additionally, these systems often depend on human intervention, which slows down the detection process. As a result, fake news spreads rapidly across social media platforms before it can be identified and controlled. Therefore, traditional systems are insufficient in addressing the growing problem of misinformation.

Disadvantages

· Time-consuming and requires manual effort
· Not scalable for large volumes of data
· Limited accuracy in detecting complex or misleading content
· Cannot understand context, sarcasm, or semantics
· Delayed detection leads to rapid spread of fake news
· Heavy dependence on human verification

Proposed System

The proposed system introduces a machine learning-based approach for automatic fake news detection using NLP techniques. It processes news articles by performing text preprocessing steps such as tokenization, stop-word removal, and normalization. The cleaned text is then transformed into numerical features using TF-IDF, enabling machine learning models to analyze textual data effectively. Classification algorithms such as Logistic Regression and Naive Bayes are trained on labeled datasets to learn patterns that distinguish real and fake news. The system evaluates model performance using metrics like accuracy, precision, recall, and F1-score to ensure reliability. It is designed to be automated, reducing the need for manual intervention. The system can classify unseen news articles in real-time, making it highly efficient.

Advantages of Proposed Systyem

· Automated detection with minimal human intervention
· High accuracy using machine learning models
· Capable of handling large-scale textual data
· Faster and real-time prediction capability
· Understands patterns and context better than rule-based systems
· Reduces the spread of misinformation

IV. Methodology

The methodology for the fake news detection system follows a structured machine learning and NLP pipeline. Initially, a dataset containing labeled news articles (real and fake) is collected from reliable sources. The raw textual data is then processed

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

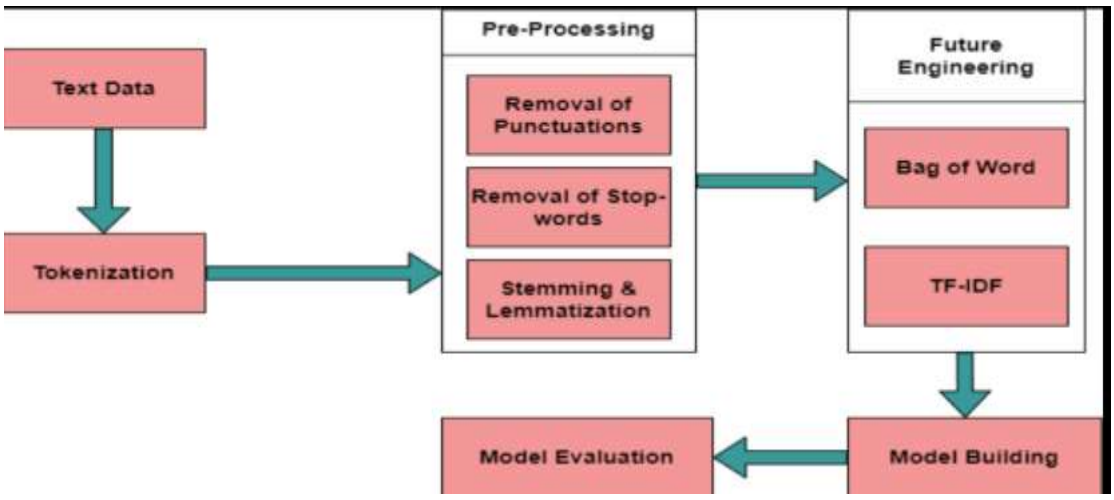
using Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and text normalization to clean and prepare the data for analysis.

After preprocessing, the text data is converted into numerical form using feature extraction techniques like TF-IDF (Term Frequency–Inverse Document Frequency), which helps in representing the importance of words in each document.

Next, the dataset is split into training and testing sets. Machine learning algorithms such as Logistic Regression and Naive Bayes are applied to train the classification model.

System Architecture

The system architecture for the fake news detection model is designed as a multi-stage pipeline that ensures efficient processing of textual data and accurate classification. It begins with the **data collection layer**, where news articles are gathered from datasets or online sources. This data is passed to the **data preprocessing layer**, where NLP techniques such as tokenization, stop-word removal, and normalization are applied to clean and structure the text. The processed data then moves to the **feature extraction layer**, where methods like TF-IDF convert textual information into numerical vectors suitable for machine learning models.



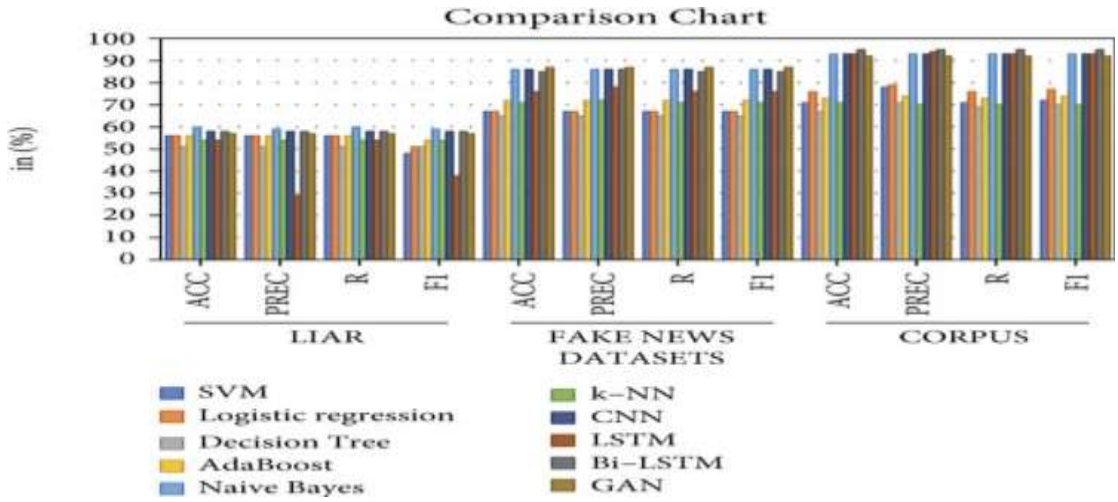
V. Result and Output

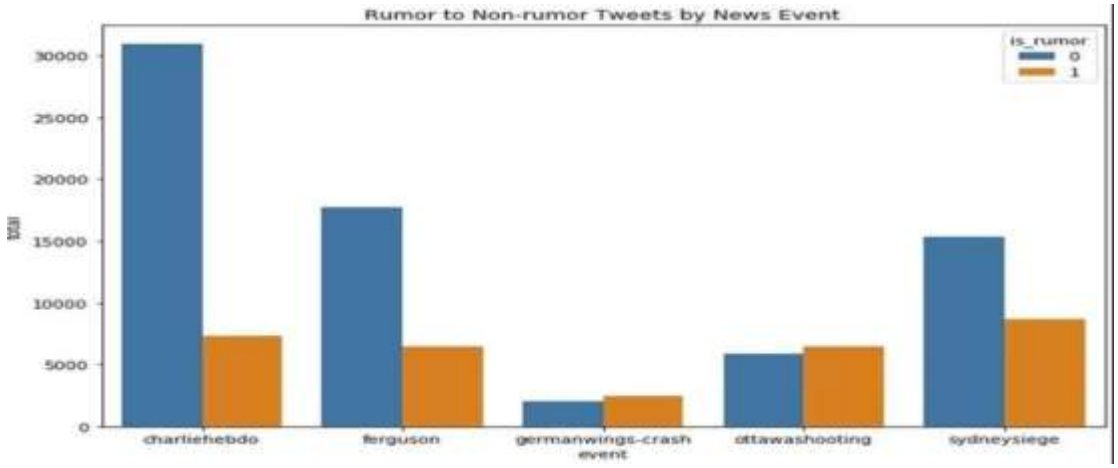
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



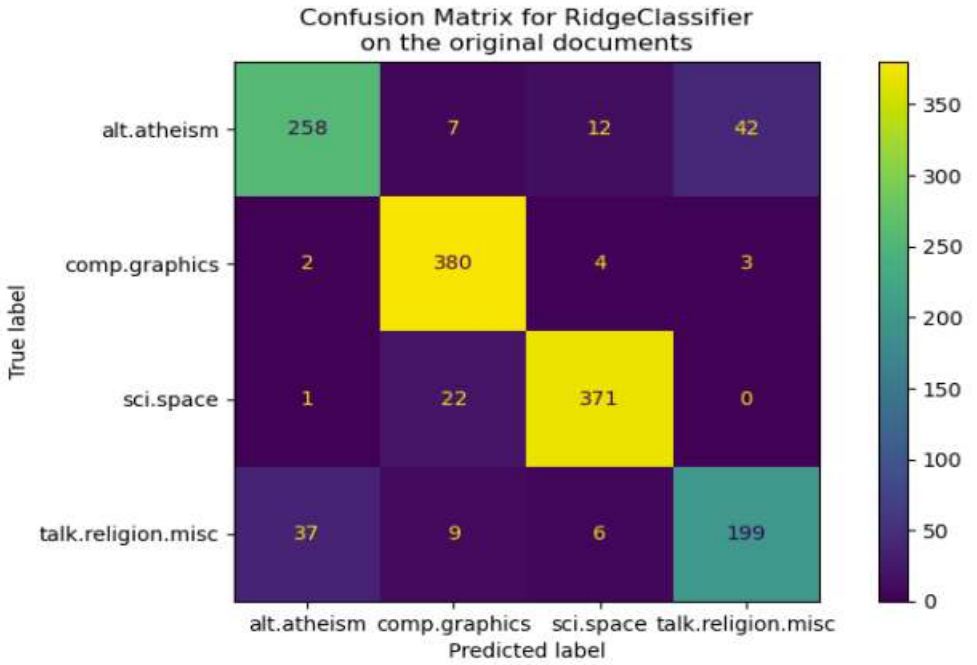


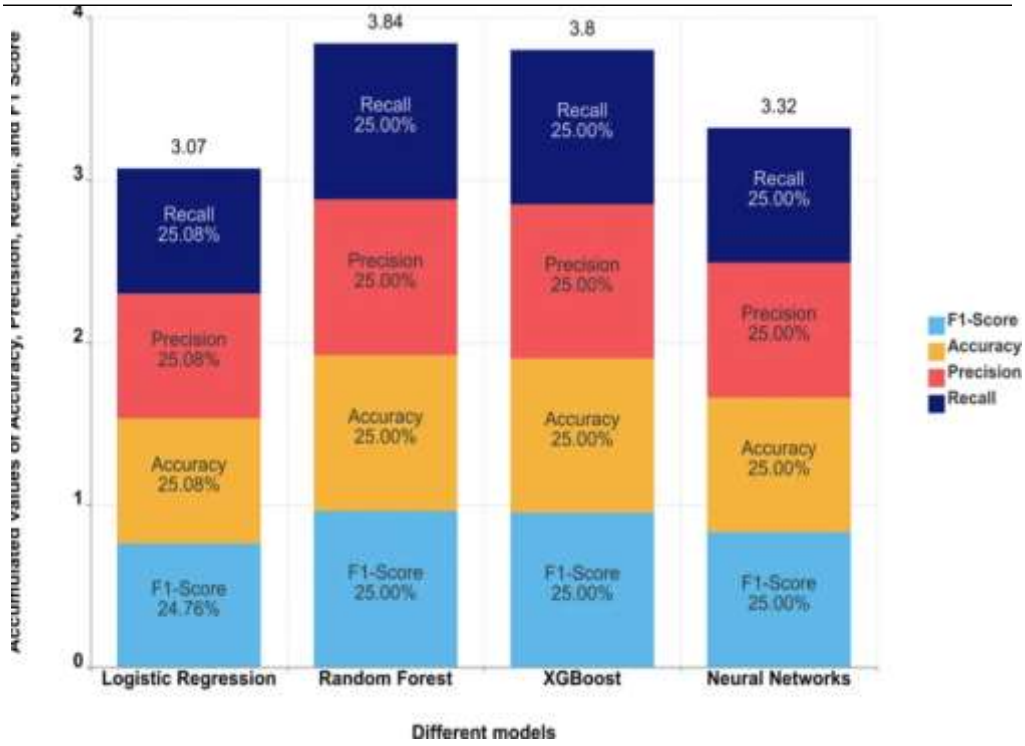
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



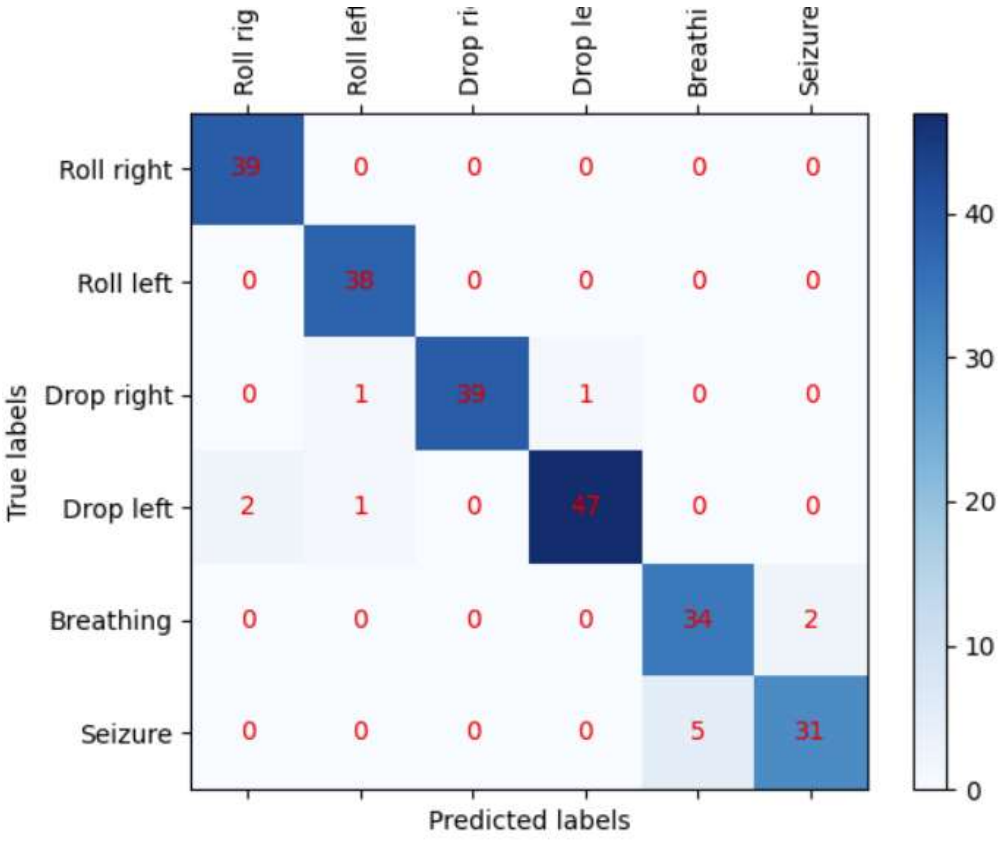


International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



VI. Conclusion

In conclusion, this project successfully developed a machine learning-based system for detecting fake news using Natural Language Processing techniques. By applying preprocessing methods and feature extraction techniques like TF-IDF, the system effectively transformed textual data into a format suitable for analysis. The implementation of machine learning algorithms such as Logistic Regression and Naive Bayes enabled accurate classification of news articles as real or fake.

The model demonstrated strong performance when evaluated using metrics such as accuracy, precision, recall, and F1-score, proving its reliability in identifying misinformation. The automated nature of the system reduces the need for manual verification and allows for faster detection of fake news.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

Although the system has some limitations, such as dependency on the quality of training data and challenges in understanding highly complex language patterns, it provides a solid foundation for further improvements. Future enhancements may include the use of advanced deep learning models and real-time integration with social media platforms.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.