International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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

STUDENT PERFORMANCE PREDICTION

[1] Sunanda Kondapalli, [2] Sindhura Devi, [3] Soumya Sri, [4] Akula Koushik

[1]AssistantProfessor, [234]Students

Department of Computer Engineering(Internet Of Things)

Siddhartha Institute of Technology & Sciences, Narapally

kondapallisunanda.cse@siddhartha.co.in , 23tq1a6914@siddhartha.co.in, 23tq1a6944@siddhartha.co.in, 23tq1a6953@siddhartha.co.in

Abstract

Student performance prediction is a challenging task due to the complex, non-linear, and dynamic nature of educational data. This project focuses on the application of machine learning techniques, particularly Linear Regression with time-series features, to predict student academic performance. By utilizing historical data such as attendance, study hours, previous scores, and assignment performance, the system identifies patterns that influence academic outcomes. Lag features are engineered to capture temporal dependencies and improve prediction accuracy. The developed model achieves a high R^2 score, indicating strong predictive capability. Interactive visualizations are used to present trends and compare actual versus predicted performance, making the results easy to interpret. Additionally, the system can predict future performance based on user input, demonstrating its practical usefulness. Overall, this project highlights how simple yet effective machine learning models can be used for accurate and efficient student performance forecasting, supporting data-driven decision-making in education.

I. Introduction

Academic performance is influenced by various factors such as study habits, attendance, previous academic results, and personal or socio-economic conditions. Educational institutions often face challenges in identifying students who are likely to struggle academically at an early stage. Without proper analysis of available data, it becomes difficult for teachers to provide timely support and improve student outcomes. Therefore, there is a need for an intelligent system that can analyze student-related data and predict academic performance in advance.

The main objective of this project is to develop a predictive model that can estimate students' academic performance using historical data from the Student Performance Dataset available in the UCI Machine Learning Repository. By applying machine learning algorithms such as Linear Regression and incorporating techniques like feature engineering and data preprocessing, the system can learn patterns from past data and make accurate predictions. This approach enables early identification of at-risk students and helps educators take preventive measures. Ultimately, the system

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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

aims to enhance decision-making, improve teaching strategies, and contribute to better academic success in modern education systems.

II. Literature Survey

Sentiment analysis has emerged as a significant research area in the field of Natural Language Processing (NLP) and machine learning. Many researchers have explored how computational techniques can be used to analyze opinions, emotions, and attitudes expressed in textual data. Early studies focused on lexicon-based approaches, where predefined dictionaries of positive and negative words were used to determine sentiment polarity. Later, machine learning algorithms such as Naïve Bayes, Decision Trees, Support Vector Machines, and Logistic Regression were widely applied to improve classification accuracy using labeled datasets. With the advancement of deep learning, models like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and transformer-based models such as BERT have shown significant improvements in capturing contextual meaning and handling complex language patterns. Research also highlights the importance of preprocessing techniques like tokenization, stop-word removal, and feature extraction methods such as TF-IDF and word embeddings. Several studies emphasize challenges such as sarcasm detection, domain dependency, and multilingual sentiment analysis. Overall, existing literature concludes that sentiment analysis plays a crucial role in understanding user opinions in domains like social media monitoring, product reviews, and customer feedback, making it a powerful tool for decision-making and business intelligence.

Several researchers have focused on improving sentiment analysis by incorporating advanced feature engineering and hybrid models. Studies show that combining lexicon-based methods with machine learning algorithms can enhance classification performance. Researchers have also explored n-grams, part-of-speech tagging, and dependency parsing to better capture linguistic patterns in text. These approaches help in identifying subtle sentiments and improving accuracy. Moreover, ensemble techniques such as combining multiple classifiers have been proven to produce more robust and reliable results compared to single models.

III. System Analysis

System analysis for student performance prediction focuses on understanding how academic and related data can be used to forecast students' future outcomes. The system aims to analyze factors such as attendance, study hours, previous grades, socio-economic background, and participation in activities. It involves collecting relevant data from educational institutions and preparing it for analysis through preprocessing techniques like cleaning, normalization, and handling missing values. The system must identify important features that significantly influence student

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

performance. Various machine learning algorithms are evaluated to select the most suitable model for prediction. Performance metrics such as accuracy, precision, and recall are used to measure effectiveness. The system should be scalable to handle large datasets from multiple students. It also needs to ensure data privacy and security. Overall, system analysis helps in designing an efficient model that supports early identification of at-risk students and improves academic outcomes.

Existing System

The existing system for student performance evaluation is mainly based on traditional methods such as manual analysis and basic statistical techniques. Teachers typically assess student performance through exams, assignments, and attendance records without using advanced analytical tools. These systems rely heavily on human judgment and may not consider multiple influencing factors simultaneously. Basic tools like spreadsheets are sometimes used, but they lack predictive capabilities. The existing approach often provides results only after exams, which delays intervention for struggling students. It does not utilize historical data effectively for forecasting future performance. Additionally, these systems may not handle large volumes of data efficiently. There is also a lack of automation and real-time monitoring. As a result, the existing system is limited in providing accurate and timely insights into student performance.

Disadvantages of Existing System

- No predictive capability for future performance
- Heavy dependence on manual evaluation
- Delayed identification of weak students
- Limited use of historical data
- Low accuracy in decision-making
- Time-consuming and inefficient

Proposed System

The proposed system utilizes machine learning techniques to predict student performance more accurately and efficiently. It collects and processes academic and behavioral data such as attendance, study time, previous grades, and other relevant factors. Data preprocessing techniques are applied to clean and transform the data for better analysis. Feature selection methods are used to identify the most important attributes affecting performance. Machine learning models such as Linear Regression, Decision Trees, Random Forest, and Support Vector Machines are trained on the dataset. The system evaluates model performance using metrics like accuracy, precision, and recall to select the best model. It enables early prediction of student performance, allowing timely intervention. The system can handle large datasets and

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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

provides scalable solutions. It also supports real-time analysis and continuous improvement through retraining. Overall, the proposed system enhances decision-making for educators and helps improve student success rates.

Advantages of Proposed System

· Provides early prediction of student performance
· Improves accuracy using machine learning models
· Reduces manual effort and saves time
· Enables timely intervention for weak students
· Utilizes historical and real-time data effectively
· Scalable for large datasets
· Supports better decision-making for educators
· Automated and efficient system

IV. Methodology

The methodology of the student performance prediction system follows a structured approach to ensure accurate and reliable results. Initially, data is collected from various sources such as student records, attendance logs, academic scores, and demographic information. The collected data is then preprocessed to handle missing values, remove inconsistencies, and normalize the dataset. Techniques like data cleaning, encoding categorical variables, and feature scaling are applied. After preprocessing, feature selection is performed to identify the most important factors influencing student performance, such as study time, attendance, and previous grades. The dataset is then divided into training and testing sets. Machine learning algorithms such as Linear Regression, Decision Trees, Random Forest, and Support Vector Machines are trained on the data. The models are evaluated using performance metrics like accuracy, precision, recall, and F1-score. The best-performing model is selected and used for prediction. Finally, the system is deployed to predict future student performance and assist educators in decision-making, with continuous updates to improve accuracy over time.

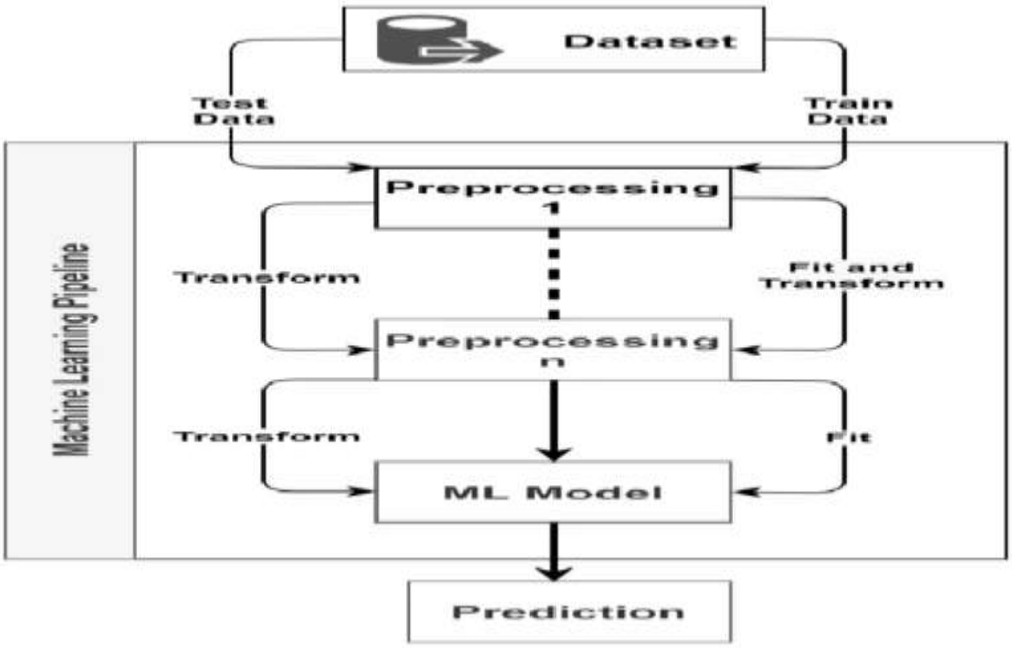
System Architecture

The system architecture of the student performance prediction system is designed as a multi-stage pipeline that processes student data efficiently from input to output. It begins with the data collection layer, where academic records, attendance, and other relevant student information are gathered from institutional databases or datasets. This data is then passed to the preprocessing layer, where it is cleaned, normalized, and transformed to handle missing or inconsistent values. The next stage is the feature selection and extraction layer, where important attributes influencing student

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

performance are identified. The processed data is then fed into the model training layer, where machine learning algorithms are applied to build predictive models. After training, the model evaluation layer assesses performance using metrics such as accuracy, precision, and recall to ensure reliability. Once validated, the system moves to the prediction layer, where new student data is analyzed to forecast performance outcomes. Finally, the deployment layer integrates the system into educational platforms, enabling real-time predictions and supporting teachers in making informed decisions to improve student success.

The evaluation module validates the model to ensure accuracy and reliability. The prediction module generates results for new student data inputs. These results are then displayed through a visualization or dashboard layer, allowing easy interpretation of predictions.



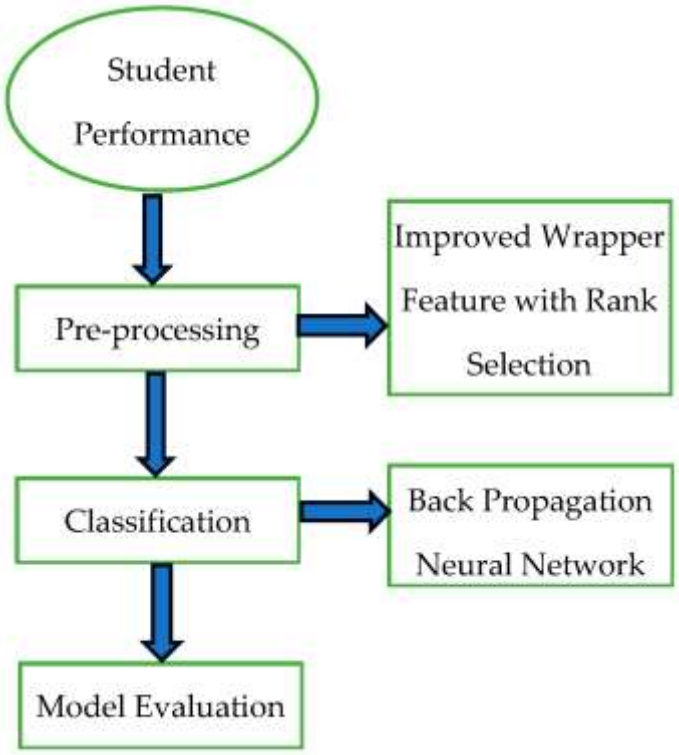
V. Result and Output

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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

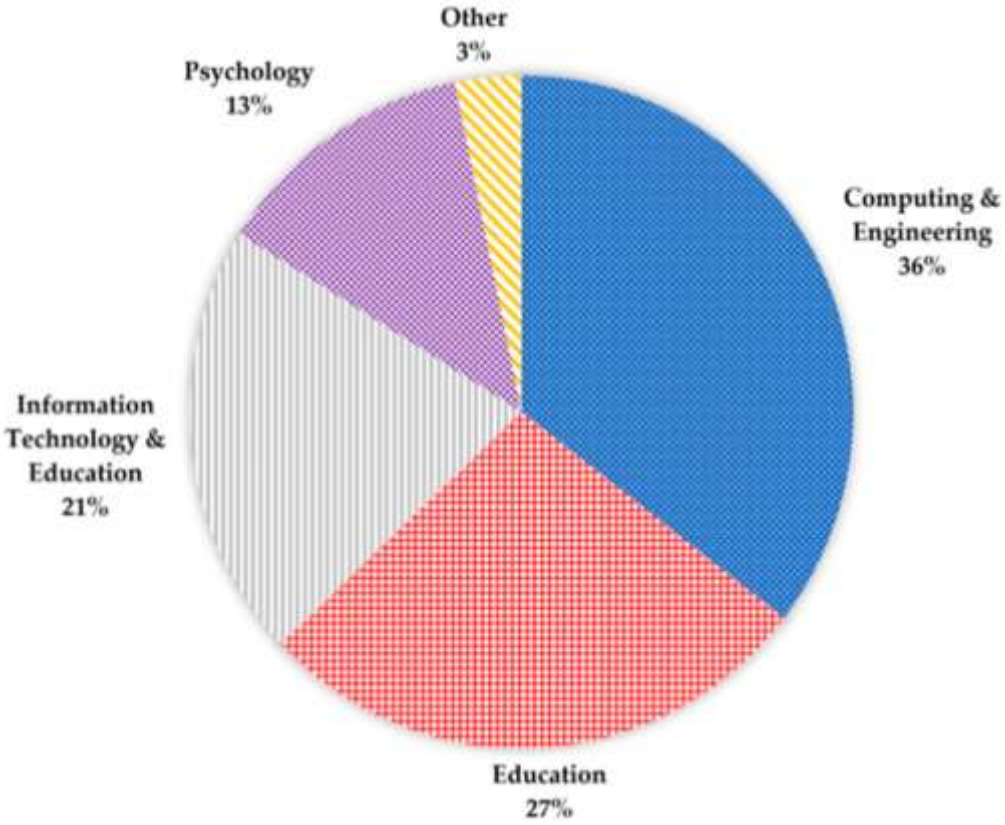


International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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

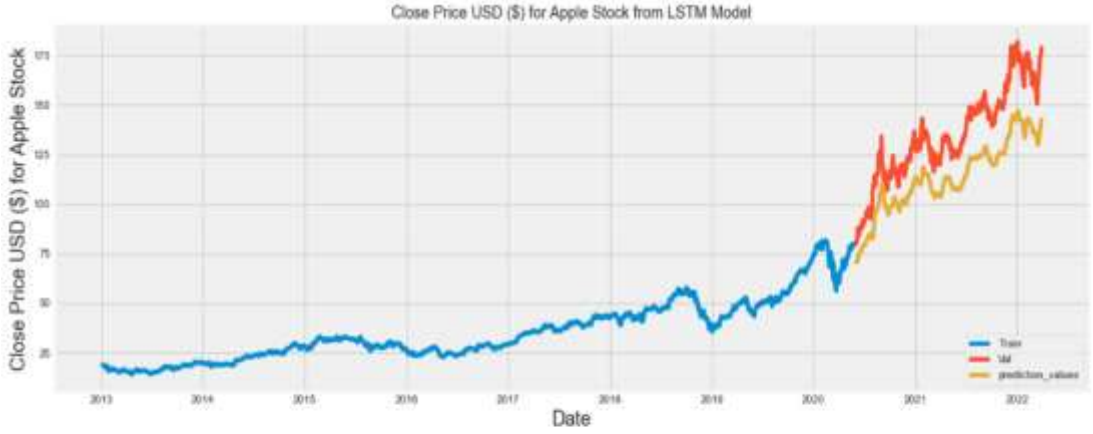


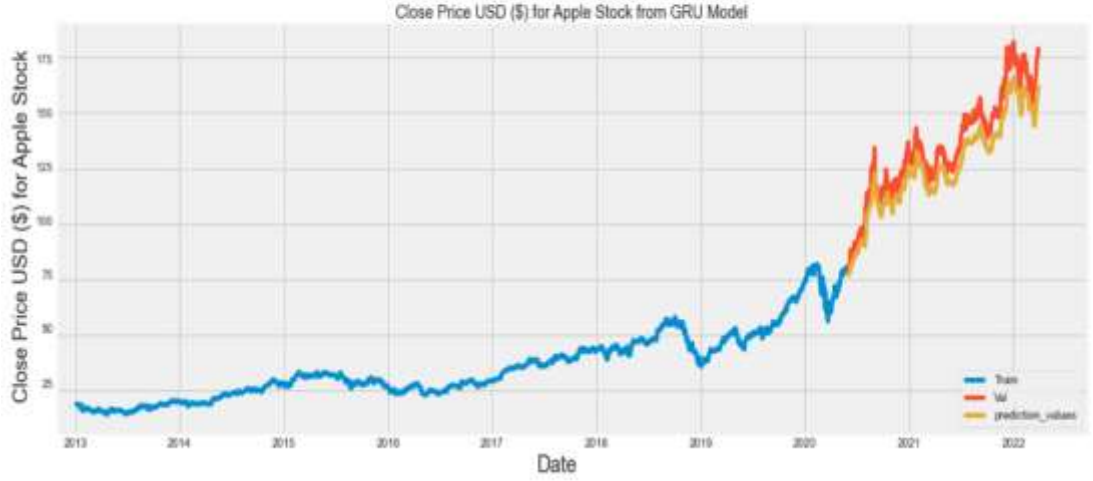
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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



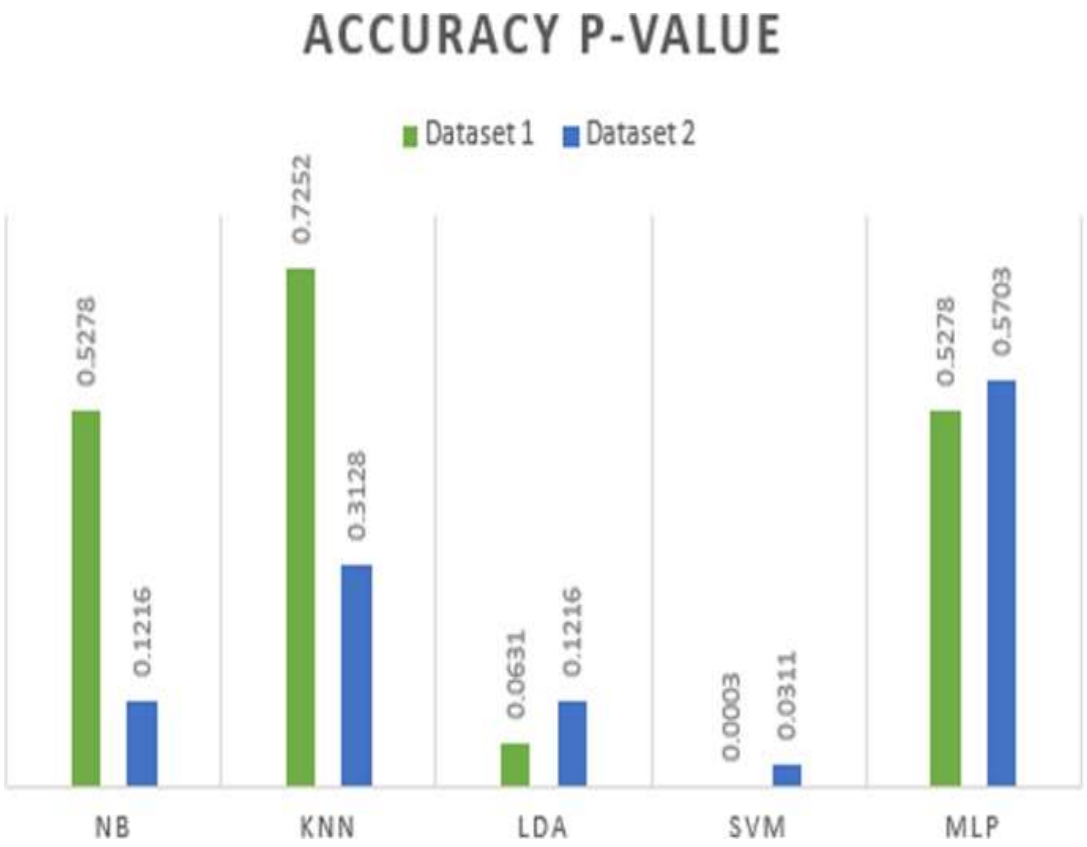


International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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



VI. Conclusion

In conclusion, the Student Performance Prediction project successfully demonstrates the application of machine learning techniques in analyzing academic data to forecast student outcomes. By considering key factors such as attendance, study hours, previous exam scores, and assignment performance, the system effectively identifies patterns that influence student success. The project emphasizes the importance of early detection of at-risk students, enabling educators to provide timely support and improve learning outcomes. It also highlights the growing role of data-driven decision-making in modern education systems. Furthermore, the use of machine learning models ensures improved accuracy, efficiency, and scalability compared to traditional methods. Overall, this project proves that predictive analytics can play a crucial role in enhancing academic performance, supporting teachers, and contributing to the development of smarter and more effective educational environments.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024).

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

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

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.

[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.