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

HOSPITAL 30 DAY PATIENT READMISSION RISK PREDICTION USING CLINICAL AND DEMOGRAPHIC FEATURES WITH SHAP EXPLAINABILITY

[1] R Shirisha, [2] A Srinidhi, [3] P Ashwini, [4] K Siri, [5] J Harsha Vardhan
[1]AssistantProfessor, [2345]Students
Department of CSE(Data Science)
Siddhartha Institute of Technology & Sciences, Narapally

shirisharangu.cse@siddhartha.co.in , 24TQ1A6704@siddhartha.co.in,
24TQ1A6737@siddhartha.co.in, 24TQ1A6725@siddhartha.co.in, 24TQ1A6723@siddhartha.co.in

Abstract

Hospital readmissions within 30 days represent a significant challenge in modern healthcare systems, as they increase operational costs and often indicate gaps in the quality of patient care. This project presents a predictive analytics framework aimed at identifying patients who are at high risk of readmission by utilizing both clinical and demographic data. The dataset includes important features such as patient age, medical history, diagnosis information, length of hospital stay, prior admissions, and treatment details, enabling a comprehensive assessment of patient health status.

To build an effective prediction system, multiple machine learning algorithms—including Logistic Regression, Random Forest, and Gradient Boosting—are implemented and compared. Data preprocessing techniques such as handling missing values, feature encoding, and normalization are applied to enhance data quality and model performance. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score to ensure reliable and robust predictions.

A key contribution of this project is the integration of explainable artificial intelligence using SHAP (SHapley Additive exPlanations). This approach provides clear insights into how different features influence the prediction of readmission risk, allowing healthcare professionals to understand and trust the model's decisions. By highlighting the most impactful factors, SHAP enhances transparency and supports informed clinical judgment.

I. Introduction

Hospital readmission within 30 days of discharge is widely recognized as a key indicator of healthcare quality and effectiveness of patient management. Frequent readmissions not only increase healthcare costs but also place additional pressure on hospital resources and staff. In many cases, readmissions may indicate incomplete treatment, inadequate discharge planning, or lack of proper follow-up care. Therefore, identifying patients who are at risk of readmission has become an essential task for healthcare providers to ensure better patient outcomes and efficient resource utilization.

With the rapid growth of healthcare data, predictive analytics and machine learning techniques have emerged as powerful tools for analyzing patient information and

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

uncovering patterns associated with readmission risk. This project focuses on leveraging clinical features such as diagnosis details, length of hospital stay, prior admissions, and treatment history, along with demographic factors like age and gender, to develop a reliable prediction model. By analyzing these variables, the system can accurately estimate the likelihood of a patient being readmitted within 30 days.

The implementation of such predictive models enables hospitals to proactively identify high-risk patients before discharge and take preventive measures. These measures may include scheduling follow-up visits, ensuring proper medication adherence, and designing personalized care plans. Additionally, the integration of explainable artificial intelligence techniques such as SHAP (SHapley Additive exPlanations) enhances the transparency of the model by clearly indicating which factors contribute most to the prediction. This not only builds trust among healthcare professionals but also supports informed clinical decision-making.

II. Literature Survey

Hospital readmission prediction has been widely studied using data analytics and machine learning techniques due to its importance in improving healthcare quality and reducing costs. Several research studies have demonstrated that machine learning algorithms such as decision trees, logistic regression, support vector machines, and neural networks are commonly used for predicting readmission risk. These models have shown promising performance, with many achieving an Area Under Curve (AUC) greater than 0.70, indicating their effectiveness in handling healthcare prediction tasks. Such findings highlight the growing role of data-driven approaches in enhancing patient care and hospital management.

Further studies emphasize the limitations of traditional statistical methods, which often fail to capture complex, non-linear relationships present in patient data. In contrast, machine learning models can automatically learn patterns from large and diverse datasets, leading to improved prediction accuracy. Advanced techniques such as Gradient Boosting and ensemble methods have been reported to achieve higher performance, with AUC values reaching up to 0.83. This demonstrates the advantage of modern algorithms in providing more reliable and accurate predictions compared to conventional approaches.

Recent research has also explored the application of deep learning and ensemble models for hospital readmission prediction. While these methods improve predictive accuracy, they often lack interpretability, making it difficult for healthcare professionals to trust and adopt them in real-world scenarios. To overcome this limitation, the focus has shifted toward explainable artificial intelligence (XAI) techniques. Methods such as SHAP (SHapley Additive exPlanations) are increasingly used to provide clear insights into how individual features influence model predictions, thereby enhancing transparency and trust.

Additionally, contemporary studies highlight the importance of integrating both clinical data and broader contextual factors, such as social determinants of health, to

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

improve prediction performance. By combining multiple data sources, models can provide a more comprehensive understanding of patient risk. Overall, the literature indicates that while machine learning significantly improves readmission prediction, challenges related to data quality, interpretability, and practical implementation remain. This project builds upon existing research by integrating accurate machine learning models with SHAP-based explainability to develop a reliable, transparent, and practical system for predicting hospital readmission risk.

III. System Analysis

The Hospital 30-Day Patient Readmission Risk Prediction system is designed to analyze patient data and predict the likelihood of readmission using machine learning techniques. The system focuses on improving healthcare quality by identifying high-risk patients before discharge. It processes clinical and demographic data such as age, diagnosis, medical history, treatment details, and length of hospital stay. The system identifies patterns and relationships within the data to detect potential readmission risks. It addresses challenges such as data imbalance and missing values in healthcare datasets. Advanced preprocessing techniques are used to clean and prepare the data. Multiple machine learning models are applied to improve prediction accuracy. The system incorporates SHAP explainability to provide transparent insights into predictions. It evaluates model performance using metrics like accuracy, precision, recall, and F1-score. The system supports better decision-making for healthcare providers. It is scalable and adaptable to different hospital datasets. Overall, it enhances patient care and reduces readmission rates.

Existing System

The existing system for predicting patient readmissions primarily relies on traditional statistical methods and manual assessment by healthcare professionals. These approaches use limited patient information and predefined rules. Data analysis is often performed manually or using basic tools, making the process time-consuming. Existing systems lack advanced predictive capabilities and fail to capture complex relationships in patient data. They do not effectively handle large-scale healthcare datasets. There is limited integration of clinical and demographic data. The systems lack real-time prediction capabilities. Visualization and reporting features are minimal. Existing methods do not provide clear explanations for predictions. Healthcare professionals must rely on experience rather than data-driven insights. These systems are less efficient in identifying high-risk patients. Overall, existing systems are less accurate and lack scalability.

Disadvantages of Existing System (Points)

- Relies on manual analysis and traditional methods
- Limited use of patient data features
- Cannot handle large and complex datasets efficiently
- Lack of real-time prediction capabilities
- Low prediction accuracy and reliability
- No explainability for predictions

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

- Time-consuming and resource-intensive

Proposed System

The proposed system introduces a machine learning-based approach for predicting hospital readmission risk. It integrates both clinical and demographic data for comprehensive analysis. The system applies preprocessing techniques to handle missing values and inconsistencies. Multiple algorithms such as Logistic Regression, Random Forest, and Gradient Boosting are used for prediction. It incorporates SHAP explainability to interpret model predictions. The system identifies key factors influencing readmission risk. It provides real-time prediction results for healthcare professionals. Interactive dashboards display patient risk scores and insights. The system improves early detection of high-risk patients. It supports targeted interventions and personalized care plans. The platform is scalable and efficient for large datasets. Overall, it enhances healthcare decision-making and reduces readmission rates.

Advantages of Proposed System (Points)

- High prediction accuracy using advanced machine learning models
- Integration of clinical and demographic data
- SHAP-based explainability for transparent predictions
- Early identification of high-risk patients
- Real-time prediction and analysis
- Reduces hospital readmission rates
- Supports data-driven decision-making
- Scalable for large healthcare datasets

IV. Methodology

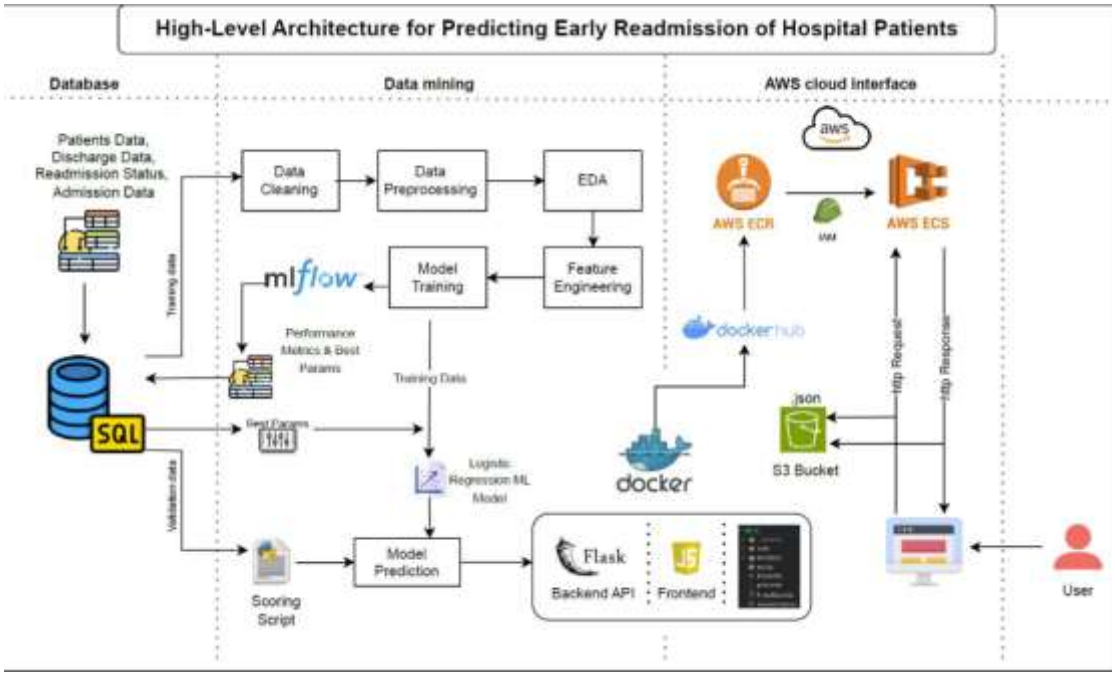
The system follows a structured methodology for predicting readmission risk. Initially, patient data is collected from hospital records. Data preprocessing is performed to handle missing values and inconsistencies. Feature selection is applied to identify relevant clinical and demographic attributes. The dataset is analyzed for class imbalance and appropriate techniques are used. The data is split into training and testing sets. Machine learning models such as Logistic Regression, Random Forest, and Gradient Boosting are trained. Model performance is evaluated using accuracy, precision, recall, and F1-score. SHAP is applied to interpret model predictions and identify important features. The best-performing model is selected for deployment. Predictions are generated for new patient data. Results are visualized through dashboards for easy interpretation.

System Architecture

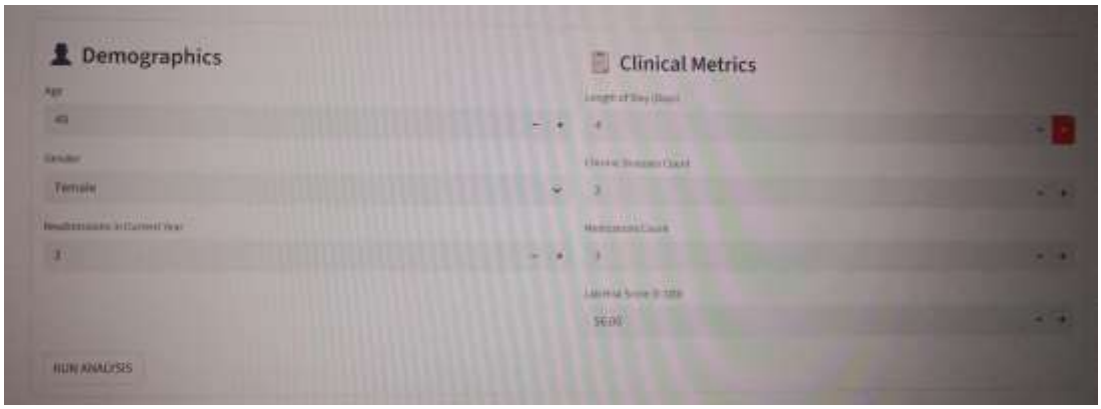
The system architecture consists of several interconnected components. The Data Collection Module gathers clinical and demographic patient data. The Data Preprocessing Module cleans and prepares the data. The Feature Engineering Module extracts relevant features. The Imbalance Handling Module addresses skewed data

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

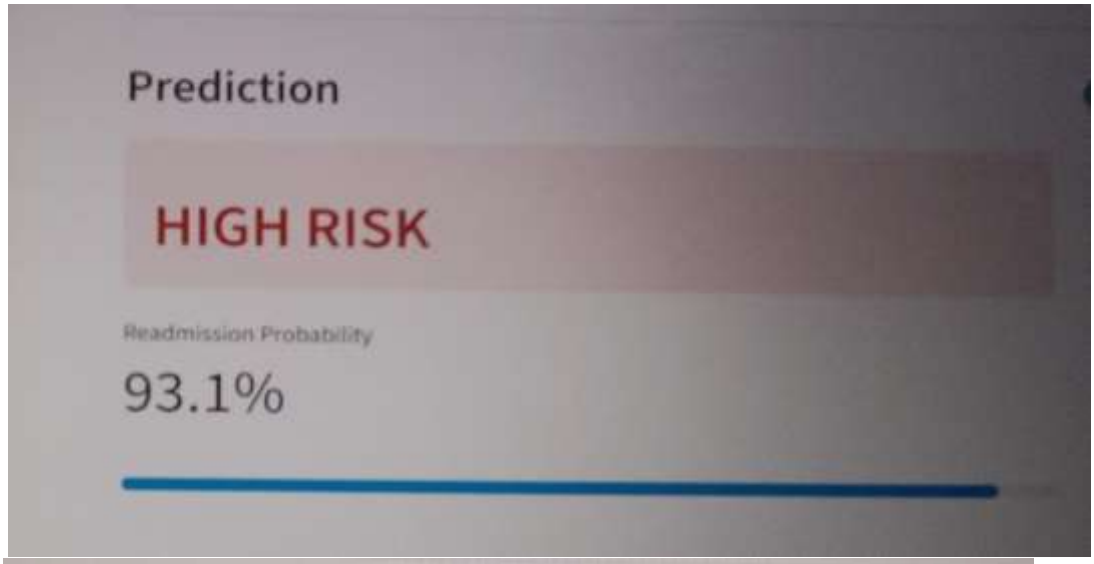
distribution. The Model Training Module applies machine learning algorithms. The Evaluation Module assesses model performance. The Explainability Module uses SHAP to interpret predictions. The Prediction Module generates readmission risk scores. The Visualization Module presents insights through dashboards. The Storage Layer manages datasets and model outputs. The Backend is implemented using Python and machine learning libraries. All components work together to provide accurate and efficient readmission prediction.

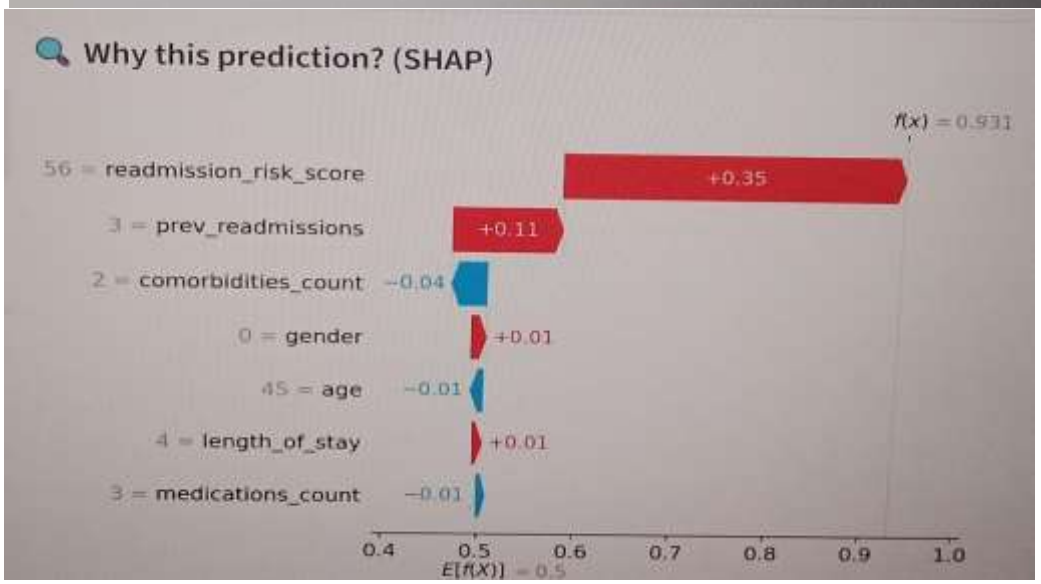


V. Result and Output



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





VI. Conclusion

In conclusion, this project successfully demonstrates the application of data analytics and machine learning techniques to predict hospital readmission within 30 days. By

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

analyzing both clinical and demographic patient data, the system effectively identifies high-risk patients prior to discharge, enabling healthcare professionals to take timely preventive measures. The use of advanced machine learning models such as Random Forest and Gradient Boosting significantly improves prediction accuracy compared to traditional approaches, ensuring more reliable outcomes.

Furthermore, the integration of SHAP (SHapley Additive exPlanations) enhances the transparency and interpretability of the model by clearly highlighting the key factors influencing each prediction. This makes the system more trustworthy and practical for real-world medical decision-making. The inclusion of an interactive web-based dashboard further improves usability by presenting insights in a clear and understandable format for healthcare providers.

Overall, this project contributes to better patient care by supporting early intervention strategies, reducing unnecessary hospital readmissions, and optimizing the use of healthcare resources. It emphasizes the importance of data-driven decision-making in modern healthcare systems and provides a strong foundation for future improvements, such as real-time data integration, advanced predictive models, and large-scale implementation across healthcare institutions.

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.

1. [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.
2. Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
3. 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.
4. 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.
5. 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.

6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
12. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
13. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
14. Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. European Journal of Advances in Engineering and Technology, 12(4), 76–81.
15. Viswanathan, V. (2024). Embedding Ethical Principles into Generative AI Workflows for Project Teams.
16. Mudusu, S. K. (2026, April 15). The secure intelligence framework: Architecting AI systems for a data-driven world. CIO (Foundry Expert Contributor Network).
17. Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.
18. Mudusu, S. K. (2025, June 3). Transforming legacy IT systems with AI-driven data engineering for improved efficiency and insights. Hampton Global Business Review (HGBR).
19. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
20. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
21. Chowdhury, A. K., Muhit, M. M. I., & Islam, M. M. (2023). A practical review to the marine maintenance practice in Bangladesh and a proposed way forward to an efficient, long-term and cost-effective solution. In Proceedings

of the 13th International Conference on Marine Technology (MARTEC 2022). https://doi.org/10.2139/ssrn.4445071

22. Manoharan, D. (2025). Healthcare EDI Transaction Lifecycles Embedded with a Multi-Layer Verification Framework to Ensure Referential Integrity.
23. Ravishankara, M. (2026, February). CircuChain: Disentangling Competence and Compliance in LLM Circuit Analysis. In SoutheastCon 2026 (pp. 1-7). IEEE.
24. Doragacharla, V. R. (2023). Comprehensive Benchmarking Analysis of Auto Scaling Approaches in Cloud Native Streaming Pipelines During Flash Sales and Holiday Traffic Peaks. Available at SSRN 6566479.
25. P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Innovative Engineering and Management Research (IJIEMR).
26. Kumar Adabala, P. (2021). Optimizing ERP Modernization: A Smart Data Migration Framework Approach. International Journal of Enhanced Research in Science, Technology &Amp; Engineering, 10(07), 61–72. https://doi.org/10.55948/ijerste.2021.0708
27. Kavuri, S. (2025). Critical Review of Software Testing Problems in the Current Decade. International Journal on Science and Technology, 16(2). https://doi.org/10.71097/ijsat.v16.i2.9469
28. Srikanth Kavuri. (2024). Probabilistic Generative Modeling for Synthesizing High-Coverage Test Data in Safety-Critical Software Applications. Computer Fraud and Security, 633–642. https://doi.org/10.52710/cfs.838
29. Venkata Pavan Kumar Gummadi. (2024). API Design and Implementation: RAML and OpenAPI Specification. Journal of Electrical Systems, 16(4), 76–85. https://doi.org/10.52783/jes.9329
30. Venkata Pavan Kumar Gummadi. (2025). MuleSoft's Role in Advancing Sustainable Digital Infrastructure: An Enterprise Integration Perspective. Journal of Information Systems Engineering and Management, 10(62s), 1313–1321. https://doi.org/10.52783/jisem.v10i62s.13783