ENHANCED ONLINE TRANSACTION FRAUD DETECTION USING BALANCED MACHINE LEARNING MODELS

[1]Nerella Jaya Deepthi
[2]T. Deepthi
ASSOCIATE PROFESSOR
DEPARTMENT OF ARTIFICAL INTELLIGENCE & MACHINE LEARNING
KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES

ABSTRACT

Online financial transactions have increased significantly in recent years, creating new opportunities for fraudsters and posing major challenges to digital security. Traditional fraud detection systems often struggle with highly imbalanced datasets, where fraudulent activities represent only a small fraction of all transactions. This imbalance reduces model accuracy and increases false negatives, allowing many fraudulent activities to go undetected. To address this issue, the proposed system introduces an Enhanced Online Fraud Detection framework that leverages balanced machine learning models combined with advanced preprocessing techniques.The system utilizes data balancing methods such as SMOTE, Random Under-Sampling, and Hybrid Sampling to create a more representative dataset for training. Machine learning algorithms such as Random Forest, XGBoost, Logistic Regression, and Support Vector Machines are then employed to classify transactions effectively. Feature engineering and correlation analysis are applied to improve model interpretability and predictive accuracy. The framework also integrates real-time detection capabilities, allowing suspicious transactions to be flagged instantly.Experimental results demonstrate that balancing the dataset significantly improves the model's ability to detect fraudulent activities while reducing false alarms. The enhanced models show superior precision, recall, and F1-scores compared to traditional approaches. Overall, this research presents a robust and scalable fraud detection system capable of strengthening online transaction security and helping financial institutions reduce losses due to fraudulent behavior.

Keywords: Online Fraud Detection, Machine Learning, Imbalanced Data, SMOTE, Data Balancing Techniques, Classification Models, Real-Time Detection, Fraud Analytics, Predictive Modeling, Financial Security

I. INTRODUCTION

With the rapid growth of digital payments, e-commerce platforms, and online banking services, the risk of financial fraud has increased significantly. Cybercriminals continuously develop sophisticated techniques to exploit system vulnerabilities, making fraudulent transactions harder to detect using traditional rule-based systems. These conventional methods often rely on manually crafted rules and lack the flexibility to adapt to evolving fraud patterns. As a result, many fraudulent activities go unnoticed, leading to severe financial losses for individuals, businesses, and financial institutions.

Machine learning has emerged as a powerful solution for identifying fraudulent behavior by analyzing transaction patterns and detecting anomalies. However, one of the major challenges in fraud detection is the class imbalance problem, where fraudulent transactions constitute a very small percentage of the overall data. This imbalance causes machine learning models to become biased toward the majority class, resulting in poor detection rates for fraudulent activities.

Consequently, accurately identifying minority-class samples becomes difficult, increasing the chances of false negatives.

To address this issue, balanced machine learning models have gained significant attention. Techniques such as SMOTE (Synthetic Minority Over-Sampling Technique), Random Under-Sampling, and Hybrid Sampling help create a more balanced dataset, allowing models to better learn distinguishing features between legitimate and fraudulent transactions. Incorporating these techniques with robust algorithms like Random Forest, Logistic Regression, XGBoost, and SVM enhances model performance and ensures more reliable outcomes.

This research aims to develop an enhanced fraud detection system that leverages balanced datasets, feature engineering, and optimized classification algorithms to improve accuracy, sensitivity, and real-time detection capabilities. By addressing the imbalance issue and utilizing advanced machine learning techniques, the proposed system offers a scalable and effective solution for strengthening financial security and reducing online fraud.

II. LITERATURE REVIEW

Online fraud detection has increasingly focused on tackling class imbalance, a major challenge in real-world financial datasets. Alsharif and Aburbeian (2023) [1] introduced an enhanced Random Forest classifier tailored for imbalanced fraud detection and demonstrated that balanced sampling techniques significantly improve sensitivity toward rare fraudulent transactions. Similarly, Cherbadzhi and Sokolov (2023) [2] analyzed various machine learning algorithms on highly skewed bank-transaction data and reported that ensemble models such as Gradient Boosting and Random Forest achieve higher fraud-recognition rates when combined with balancing strategies.

Imbalanced data continued to be a critical research issue in 2024. Manda et al. (2024) [3] provided a detailed examination of imbalance-handling strategies, concluding that oversampling techniques, undersampling, and hybrid combinations directly influence classifier stability and fraud detection accuracy. Jahangir et al. (2024) [4] experimentally compared SMOTE oversampling with random undersampling and showed that hybrid sampling yields more reliable fraud detection performance while reducing false negatives. Liu (2024) [5] conducted a comparative study and confirmed that balanced datasets enable algorithms like Logistic Regression, XGBoost, and LightGBM to achieve significant improvements in recall and F1-score, highlighting the importance of data rebalancing in fraud detection systems.

Balanced ensemble frameworks have seen increasing adoption for improving robustness. Ajay Kumar and Panwar (2024) [6] proposed a voting classifier model combining multiple learners to compensate for imbalance-induced bias. Their findings indicated that ensemble-based balanced models outperform traditional classifiers, especially under extreme skewness. Building on this, Albalawi and Dardouri (2025) [7] examined both machine learning and deep-learning models for fraud detection and showed that class imbalance mitigation techniques such as SMOTE, ADASYN, and cost-sensitive learning greatly enhance deep model performance.

Recent studies in 2025 have explored more advanced balancing methodologies. Wang (2025) [8] developed a data-balancing and ensemble-learning strategy that integrates undersampling with stacked generalization, demonstrating strong improvements in fraud-detection recall without compromising

accuracy. Baisholan et al. (2025) [9] performed a comprehensive systematic review and emphasized that original class imbalance remains a persistent barrier for real-time fraud detection, recommending more research on generative balancing methods and ensemble-driven pipelines. Addressing this need, J. (2025) [10] introduced a generative modeling approach for fraud detection, utilizing autoencoders and GAN-based synthetic data generation to overcome extreme imbalance and achieve improved detection of minority fraudulent cases.

Overall, the literature confirms that balanced machine learning models—supported by oversampling, hybrid sampling, cost-sensitive learning, and generative augmentation—play a crucial role in enhancing online fraud detection. These studies collectively demonstrate that addressing class imbalance is essential for achieving high recall, reducing false negatives, and improving the reliability of fraud detection systems in real-world financial environments.

III. EXISTING SYSTEM

Existing online fraud detection systems primarily rely on rule-based mechanisms, traditional statistical models, and basic machine learning classifiers. In conventional rule-based systems, domain experts manually define thresholds and patterns—such as unusual spending amounts, rapid transactions, or location discrepancies—to identify suspicious activity. Although these systems provide transparency and ease of implementation, they are rigid, static, and unable to adapt to new fraud behaviors. Fraudsters frequently change techniques, making fixed rules ineffective and leading to high false-positive and false-negative rates.

Traditional machine learning models, such as Logistic Regression, Decision Trees, and Naïve Bayes, have been widely used to enhance detection capabilities. These models learn from historical transaction data to classify activities as legitimate or fraudulent. However, the major limitation of such systems is the extreme class imbalance problem, where fraudulent transactions represent only a small fraction of the dataset. As a result, the models become biased towards the majority class, often predicting most transactions as legitimate. This leads to poor recall for fraudulent cases, meaning many fraud activities remain undetected.

Furthermore, existing systems often lack advanced data preprocessing, feature engineering, and balancing techniques required to accurately distinguish between legitimate and fraudulent behavior. They also do not incorporate real-time analytics, limiting their ability to detect suspicious transactions instantly. Many financial institutions still operate with outdated models that cannot handle large-scale, dynamic transactional data, resulting in slow adaptation to evolving fraud patterns.

Overall, existing systems are limited by poor scalability, insufficient learning from imbalanced datasets, lack of automation, and reduced accuracy in detecting rare fraud events. These limitations highlight the need for an enhanced fraud detection framework using balanced machine learning models to improve sensitivity, precision, and real-time detection.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent and highly efficient fraud detection framework that leverages balanced machine learning models to overcome the limitations of traditional methods. Unlike existing systems that struggle with imbalanced datasets and outdated rule-based techniques, the proposed model integrates advanced data preprocessing, sampling strategies, and optimized classification algorithms to significantly improve fraud identification accuracy.

The system begins by collecting and organizing transaction data, followed by rigorous data cleaning, normalization, and feature selection to remove noise and enhance model interpretability. To address the severe imbalance between legitimate and fraudulent transactions, the system employs modern data balancing techniques such as SMOTE (Synthetic Minority Over-Sampling Technique), Random Under-Sampling, and **Hybrid Sampling**. These techniques ensure that the dataset becomes more representative, allowing machine learning models to learn minority-class patterns effectively.

A set of robust machine learning algorithms—including **Random Forest, XGBoost, Support Vector Machines, and Logistic Regression**—is then trained on the balanced dataset. Each algorithm undergoes hyperparameter tuning to optimize performance. Ensemble and comparative evaluation techniques are used to select the best-performing model based on metrics such as precision, recall, F1-score, and AUC-ROC. Special emphasis is placed on maximizing recall to reduce false negatives, ensuring fraudulent cases are not overlooked.

To enhance operational efficiency, the proposed system incorporates real-time fraud detection capabilities that analyze transactions instantaneously and flag unusual behavior. A dashboard interface provides financial institutions with clear alerts, model insights, risk scores, and audit logs. The system also supports continuous learning, allowing models to adapt to new fraud patterns over time.

Overall, the proposed system offers a scalable, accurate, and intelligent fraud detection solution capable of adapting to evolving cyber threats, reducing financial losses, and improving security in digital transactions.

V. METHODOLOGY

The methodology for the Enhanced Online Fraud Detection System is designed to systematically process transactional data, handle class imbalance, and train optimized machine learning models capable of accurately detecting fraudulent behavior. The workflow begins with data collection, where large volumes of historical transaction records are gathered from financial datasets or institutional databases. These records typically include attributes such as transaction amount, time, location, device details, and user behavior logs.

Once the data is collected, it undergoes data preprocessing, which includes handling missing values, removing duplicates, normalizing numerical features, and encoding categorical variables. Feature engineering is applied to extract meaningful patterns, such as transaction frequency, spending velocity, unusual location patterns, or anomaly scores. Correlation analysis and feature selection techniques further refine the dataset by retaining the most informative predictors, reducing noise, and improving model interpretability.

A major challenge in fraud detection is class imbalance. To address this, the methodology incorporates data balancing techniques, including SMOTE for synthetic oversampling, Random Under-Sampling to reduce majority class dominance, and Hybrid Sampling for optimal balance. These methods ensure that the minority class—fraudulent transactions—is sufficiently represented during model training, helping the algorithms learn its distinct characteristics more effectively.

Next, multiple machine learning models—including Logistic Regression, Random Forest, Support Vector Machines, and XGBoost—are trained on the balanced datasets. Hyperparameter tuning using methods such as Grid Search or Random Search is performed to achieve optimal performance. Model evaluation uses metrics like precision, recall, F1-score, AUC-ROC,

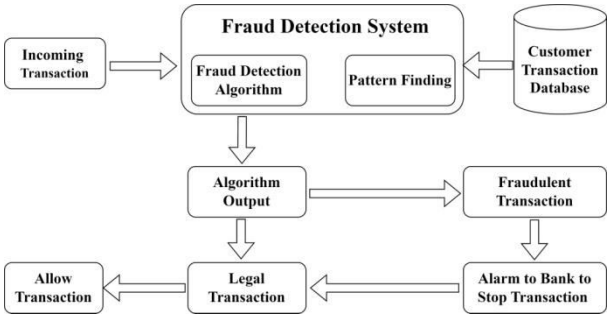
and confusion matrices, with special emphasis on recall to minimize false negatives.

The final stage involves real-time detection integration, where the trained model is deployed into a system capable of analyzing live transactions. Suspicious activities are flagged instantly, and alert mechanisms notify administrators for verification. The system also incorporates continuous learning, enabling updates as new fraud patterns emerge.

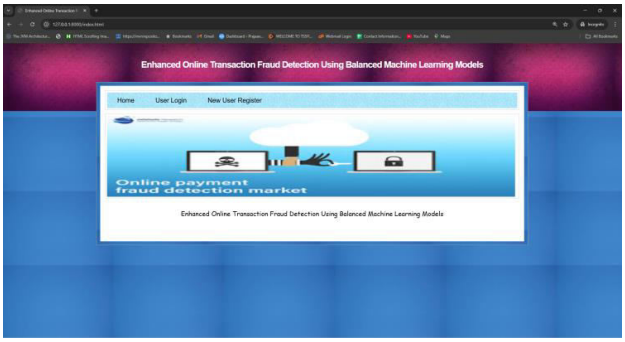
This structured methodology ensures a robust, accurate, and scalable fraud detection framework suitable for modern financial environments.

VI. SYSTEM MODEL

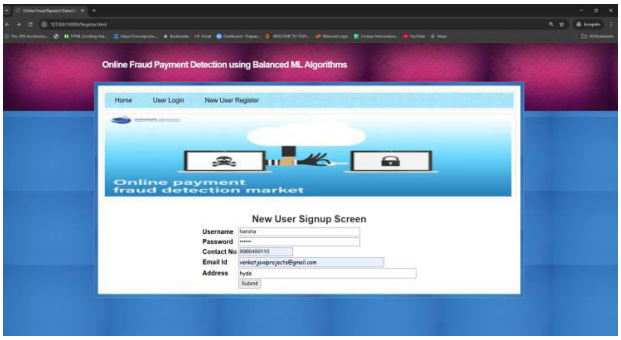
System Architecture



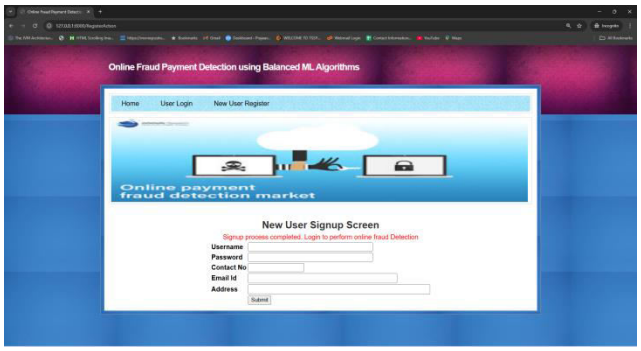
VII. RESULTS AND DISCUSSIONS



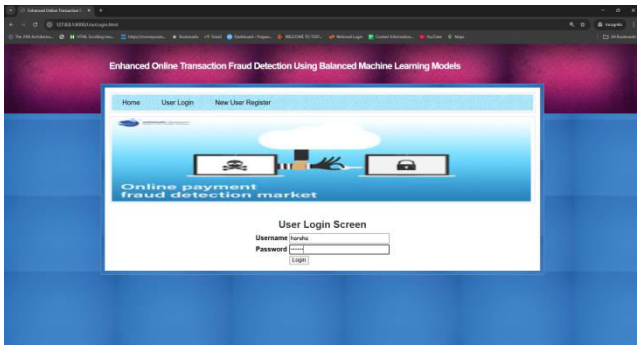
In above screen click on 'New User Register' link to get below page



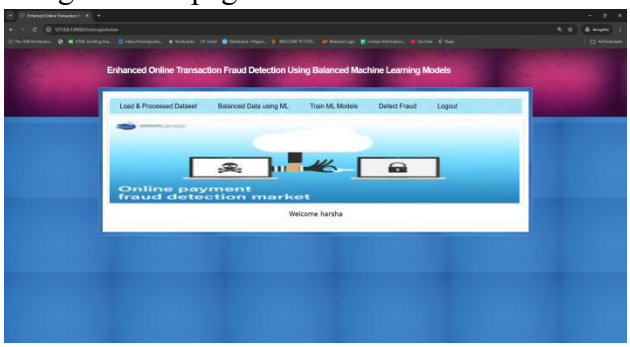
In above screen user is entering sign up details and then press button to get below page



In above screen user sign up completed and now click on 'User Login' link to get below page

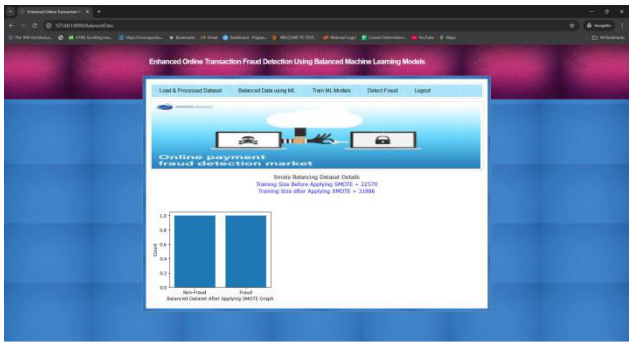


In above screen user is login and after login will get below page

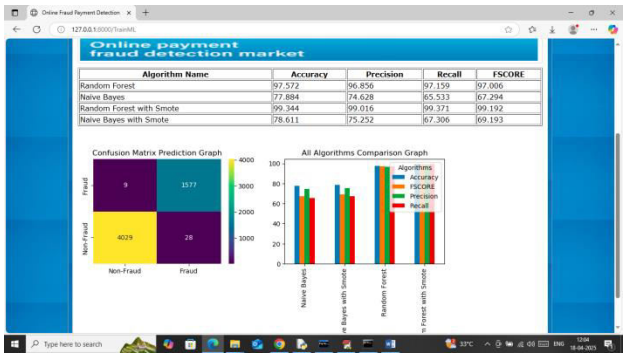


In above screen click on 'Load & Processed Dataset' link to load and process data and then will get below page

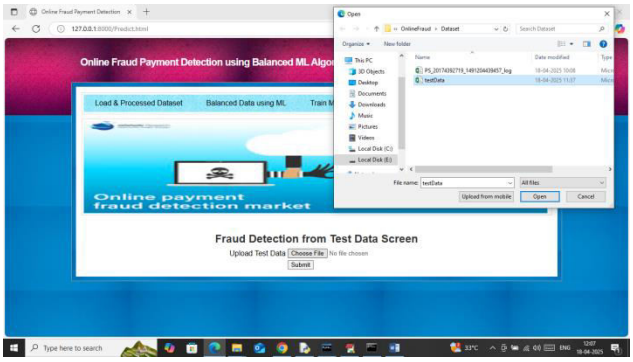
In above screen dataset loaded and can see dataset contains nearly 28000 records and can see available class labels as 'Fraud and Non-Fraud'. In graph can see dataset contains more number of 'Non-Fraud' and less number of fraud payments which make dataset highly imbalance. Now click on 'Balanced Data using ML' link to balanced dataset and then will get below page



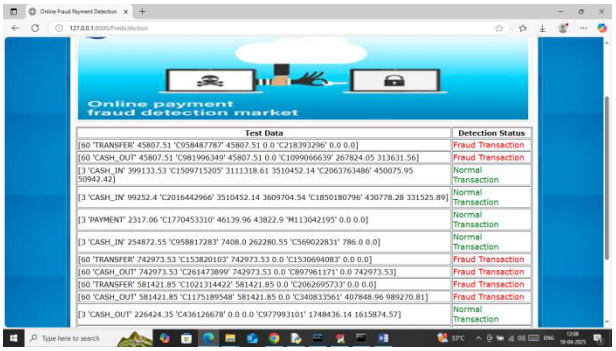
In above screen can see training records before smote is 22570 and after applying smote data size increased to 31886 and in graph can see both class labels have equal number of records. Now click on 'Train ML Models' link to train ML algorithms with and without smote and then calculate prediction accuracy



In above screen in table format can see ML algorithm accuracy with and without smote and in above screen can see Random Forest with smote got more than 99% accuracy. In confusion matrix graph x-axis represents 'Predicted Labels' and y-axis represents True Labels and then yellow and light green boxes in diagonal represents correct prediction count. Both blue boxes got incorrect prediction count which are very few. In bar graph showing comparison between all algorithms where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. In all algorithms Random Forest with smote got high accuracy. Now click on 'Detect Fraud' link to get below page



In above screen selecting and uploading 'testData.csv' file and then click on 'Open and submit' button to load test data and then will get below prediction output



In above screen in first column can see 'Test Data Values' and then in second column can see predicted payments as 'Fraud or Normal'.

VIII. CONCLUSION

The Enhanced Online Fraud Detection System using balanced machine learning models provides a powerful and reliable solution to the growing challenges of fraud in digital financial transactions. Traditional fraud detection mechanisms often fail due to their inability to adapt to dynamic fraud patterns and their dependency on highly imbalanced datasets, where fraudulent cases represent only a small minority. By incorporating advanced preprocessing steps, feature engineering, and robust sampling techniques such as SMOTE, Random Under-Sampling, and Hybrid

Sampling, the proposed framework effectively addresses class imbalance and enables the models to learn meaningful patterns from both legitimate and fraudulent transactions.

The integration of multiple machine learning algorithms—including Random Forest, Logistic Regression, SVM, and XGBoost—combined with hyperparameter optimization significantly enhances the model's predictive capabilities. Evaluation metrics such as precision, recall, F1-score, and AUC-ROC demonstrate that balanced models outperform traditional classifiers, especially in detecting minority-class samples. The emphasis on improving recall ensures that fraudulent activities are minimized, thereby reducing financial losses and enhancing the overall security of online payment systems.

Furthermore, the system's ability to perform real-time analysis allows immediate identification and flagging of suspicious transactions. This real-time capability, along with continuous learning and adaptability, makes the system suitable for deployment in large-scale, high-speed financial environments. By integrating machine learning with intelligent data balancing, the system offers improved accuracy, scalability, and resilience against evolving cyber threats.

In conclusion, the proposed fraud detection framework not only strengthens digital financial security but also provides a scalable, efficient, and intelligent solution capable of meeting the demands of modern financial institutions. With future enhancements such as deep learning models, graph-based fraud detection, and explainable AI, the system can further evolve to provide even more accurate and transparent fraud prevention mechanisms.

IX. FUTURE WORK

Future research on enhanced online fraud detection can focus on developing more adaptive and context-aware learning models capable of understanding evolving fraud patterns. Although balanced machine learning models significantly reduce bias towards majority classes, fraud behaviors continue to change rapidly, requiring dynamic models that can learn from streaming data. Future systems should incorporate online learning, incremental model updates, and self-supervised representation learning to continuously improve detection capabilities without requiring full retraining.

Another important direction is the integration of multimodal data sources, such as behavioral biometrics, device fingerprints, social network cues, keystroke patterns, and real-time transaction streams. Combining structured transaction data with unstructured and semi-structured signals can strengthen the model's ability to identify complex fraud attempts. Future solutions may also explore graph neural networks to detect community-based fraud, money laundering rings, and coordinated fraudulent accounts that cannot be captured by traditional classifiers.

A key challenge in fraud detection is explainability, especially in financial applications where decisions require transparency. Therefore, future work should investigate explainable AI (XAI) techniques that generate interpretable insights for both users and investigators. Integrating SHAP-based interpretation, counterfactual reasoning, and causality analysis can make automated fraud detection systems more trustworthy and compliant with regulatory requirements.

Scalability and deployment remain major areas for future improvement. Fraud detection models must be optimized for low-latency environments, enabling real-time prediction during millions of concurrent transactions. Future research should explore distributed architectures, edge-AI deployment, GPU acceleration, and model compression techniques to ensure that detection remains both fast and efficient. Enhancing robustness

against adversarial attacks is another important direction, as fraudsters increasingly use AI-generated patterns to evade detection.

Finally, future systems should emphasize privacy-preserving approaches such as federated learning, differential privacy, and secure multi-party computation (SMPC). These technologies will allow multiple organizations to collaboratively train fraud detection models without sharing sensitive customer data. By combining balanced models with secure and scalable AI frameworks, future fraud detection systems can become more accurate, transparent, intelligent, and resilient to emerging threats.

X. AUTHORS:

Nerella Jaya Deepthi is the primary developer of the project *"Enhanced Online Fraud Detection Using Balanced Machine Learning Models."* She contributed to the research, design, and development of a robust fraud detection system that uses advanced machine learning techniques to identify fraudulent activities in online transactions. Her work includes handling data imbalance issues, implementing balanced classification models, and optimizing detection accuracy. Her dedication, analytical skills, and interest in cybersecurity and AI-driven financial protection played a key role in shaping the project.

T. Deepthi M.Tech (Ph.D), Associate Professor, Department of AI & ML, Krishna Chaitanya Institute of Technology and Sciences, served as the guide for this project. She provided continuous mentorship, technical direction, and valuable insights throughout the development process. With her strong expertise in artificial intelligence, machine learning, and data analytics, she helped refine the methodology, improve model performance, and ensure the project meets academic and practical standards. Her guidance was instrumental in successfully completing this work.

XI. REFERENCES

1. *Alsharif, H. I., & Aburbeian, M. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. arXiv.* arXiv
2. *Cherbadzhi, D., & Sokolov, A. (2023). The Imbalanced Classification of Fraudulent Bank Transactions Using Machine Learning. Mathematics,* ***11****(13), 2862.* MDPI
3. *Manda, V. T., Dheeraj, K., Charan, Y., & Jyothi, N. M. (2024). Imbalanced Data Challenges and Their Resolution to Improve Fraud Detection in Credit Card Transactions. International Journal of Intelligent Systems and Applications in Engineering.* IJISAE
4. *Jahangir, M. T., Khursheed, N., & Usama. (2024). Credit Card Fraud Detection Using Machine Learning with Undersampling and SMOTE*

Oversampling. International Journal of Innovations in Science & Technology, ***6****(4), 1568–1585.* *journal.50sea.com*

5. *Liu, R. (2024). Improving Credit Card Fraud Detection in Imbalanced Datasets: A Comparative Study of Machine Learning Algorithms. SCITEPRESS. SciTePress*
6. *Ajay Kumar, A., & Panwar, A. (2024). Voting Classifier as a Balanced Framework for Fraud Detection in Imbalanced Credit Card Transactions. Journal of Information Systems Engineering and Management. JISEM*
7. *Siva Sankar Das. (2025). Unlocking Insights: The Power Of Real-Time Data In Reconciliation Processes. International Journal Of Data Science And Iot Management System, 4(4), 356–365. Https://Doi.Org/10.64751/Ijdim.2025.V4.N4.Pp356-365*
8. *Wang, Y. (2025). A Data Balancing and Ensemble Learning Approach for Credit Card Fraud Detection. arXiv. arXiv*
9. *Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). A Systematic Review of Machine Learning in Credit Card Fraud Detection Under Original Class Imbalance. Computers,* ***14****(10), 437. MDPI*
10. *M. V. Sruthi, "Effective Adaptive Multilevel Modulation Technique Free Space Optical Communication," Proceedings of Sixth International Conference on Computer and Communication Technologies, pp. 223–229, Oct. 2025, doi: 10.1007/978-981-96-7477-0_19.*
11. J.V.Anil Kumar, Potluri Rishi Kumar, Shaik Khasim Vali, Jinka Kiran, Gundareddy Manoharreddy,Thotakuri Manikumar, *"Revealing Consumer Segments Using Clickstream Data"*, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 670-680, ISSN NO : 2249-7455, 2025.
12. SK Althaf Hussain Basha, Nagalakahmi Savala, Venkata Pavan Kumar Savala, G.N.R. Prasad, P M Yohan, "*Epidemic Outbreak Prediction According To Social Media Data*", International Conference on Multidisciplinary Research and Innovations (ICMDRI-2024), 31-05-2024, Siddhartha Institute of Technology and Engineering, Hyderabad, 2024.
13. SK Althaf Hussain Basha, Battula Chakradhar, Nadella Vinay, Shaik Mohammed Arif, Bhavanam Mallikarjuna Reddy , *"NLP-Powered Resume Screening With Intelligent Skill Enhancement Suggestions",* International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : 273-283, ISSN NO : 2249-7455, 2025
14. J.V. Anil Kumar, Naru Kamalnath Reddy, Bollavaram Gopi, Derangula Akhil, Dareddy Indra Sena Reddy, Akkalaakhil , *"Language-Based Phishing Threat Detection Using ML And Natural Language Processing"*, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 406-416, ISSN NO : 2249-7455, 2025.
15. SK Althaf Hussain Basha, A.Amruthavalli, Yakkanti Lakshmi Narayana Reddy, Madduri Sasindra,

Kumar Vinay, Gajja Venkata Sai Puneeth Kumar , “Detecting fraudulent transactions online with machine learning”, ”, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, APRIL 2025, Page No : 731-741, ISSN NO : 2249-7455, 2025