International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

ONLINE FRAUD PAYMENTDETECTION USINGBALANCEDML ALGORITHMS

[1] **Mr S.SRINIVASARAO / sakamuris@city.ac.in**

[2]**Dr.K.KIRAN KUMAR / kirankommineni@city.ac.in**

[3,4,5,6]**Pasam Ruchitha, Shaik Nagul Sharif, Devarakonda Bindu Priya, Macherla Anil Kumar**

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, CHALAPATHI INSTITUTE OF TECHNOLOGY, MOTHADAKA, GUNTUR, ANDHRA PRADESH, INDIA-522016.

ABSTRACT

Online financial transactions have increased rapidly in recent years, leading to a significant rise in fraudulent activities. Fraudulent payments can result in severe financial loss, compromised customer trust, and disrupted business operations. A major challenge in fraud detection is the highly imbalanced nature of payment datasets, where legitimate transactions significantly outnumber fraudulent ones. Traditional machine learning models often fail to detect minority-class fraud samples accurately, resulting in high false-negative rates. This paper proposes a fraud detection system using balanced machine learning algorithms combined with sampling techniques such as SMOTE, Random Undersampling, and Hybrid Sampling. The system uses transaction features including

amount, time, device ID, location, and user behavior patterns. Balanced classifiers like Random Forest, XGBoost, and Logistic

Regression with class weighting are employed to enhance fraud detection accuracy. Experimental results demonstrate that using balanced datasets leads to significantly improved recall and F1-score, ensuring more reliable fraud identification. This approach minimizes false negatives, enhancing security in

digital payments. The proposed model is scalable, adaptable, and suitable for implementation in real-world financial systems.

Keywords

Fraud Detection, Machine Learning, Imbalanced Data, SMOTE, Balanced Algorithms, Online Payment Security.

I. INTRODUCTION

Digital payment systems such as online banking, mobile wallets, and e-commerce platforms have become an essential part of modern financial services. However, the increasing volume of online transactions has also led to sophisticated fraud attacks including unauthorized transactions, account takeovers, and identity theft. Detecting fraudulent payments is challenging due to their rarity and unpredictable patterns. Most datasets in real-world scenarios are highly imbalanced, where fraudulent transactions typically represent less than 1% of total data. This imbalance causes machine learning models to favor the majority class (genuine transactions), leading to poor detection of actual fraud cases.

Effective fraud detection requires models that can learn from minority-class samples while preventing overfitting. Balanced machine learning algorithms, when combined with data balancing techniques such as SMOTE (Synthetic Minority Oversampling Technique) and Random Undersampling, improve model performance. These techniques allow classifiers to better identify suspicious patterns in transactions. In this paper, a balanced fraud detection model is proposed using advanced ML algorithms. The goal is to strengthen online payment security and provide a more reliable detection system for financial institutions.

II. LITERATURE REVIEW

Previous fraud detection research mostly applied traditional ML models such as Logistic Regression, Decision Trees, and Naïve Bayes. While effective for basic classification tasks, these models struggle with imbalanced datasets. They often predict all transactions as legitimate, achieving high accuracy but poor fraud recall.

Recent literature emphasizes the importance of handling class imbalance. Techniques such as SMOTE, ADASYN, and Tomek Link removal have shown significant improvement in fraud classification. Hybrid sampling techniques

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

combining oversampling and undersampling have demonstrated better results than using a single sampling technique. Furthermore, ensemble models such as Random Forest, Gradient Boosting, and XGBoost have outperformed traditional algorithms due to their ability to capture nonlinear features within transaction data.

Studies also highlight the use of anomaly detection algorithms like Isolation Forest and One-Class SVM for detecting rare fraud behavior. However, these models may fail when patterns are complex. Balanced ML approaches, including cost-sensitive learning and class-weight adjustments, have recently gained attention for fraud detection. This paper builds on these findings by implementing balanced ML algorithms with hybrid sampling to achieve superior fraud detection performance.

III. METHODOLOGY

The proposed fraud detection system consists of several steps: data preprocessing, feature engineering, data balancing, model training, and evaluation. The dataset includes features such as transaction time, payment method, device information, IP address, transaction amount, and customer ID. Preprocessing includes handling missing values, normalizing numerical data, encoding categorical variables, and detecting outliers.

Since fraud datasets are highly imbalanced, data balancing techniques are applied. SMOTE is used to synthetically oversample minority-class samples, while Random Undersampling reduces the majority class. A hybrid approach combining both is used to achieve optimal class distribution. Balanced algorithms such as Random Forest with class weights, Balanced Bagging Classifier, and XGBoost with scale_pos_weight are implemented.

Each model is trained on the balanced dataset using cross-validation to avoid overfitting. Evaluation metrics include precision, recall, accuracy, AUC-ROC, and F1-score. Recall is particularly important because missing a fraud case (false negative) can lead to significant losses. The methodology ensures effective fraud detection even with limited fraud samples.

VI. RELATED WORK

Several early studies applied traditional machine learning algorithms such as Decision

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Trees, Naïve Bayes, Support Vector Machines (SVM), and Neural Networks for fraud detection. For example, Phua et al. compared multiple classifiers and found that neural networks achieved high accuracy, while SVM-based models also showed strong performance in detecting financial fraud patterns . Later works introduced hybrid and ensemble approaches, such as AdaBoost and voting classifiers, which improved robustness and classification accuracy over single models .

A major challenge identified in almost all studies is class imbalance, where fraud cases represent less than 1–2% of total transactions. This imbalance leads to biased models that favor normal transactions and fail to detect fraud effectively . To address this, researchers introduced data balancing techniques such as oversampling, undersampling, and SMOTE (Synthetic Minority Oversampling Technique). These techniques significantly improve model performance by providing balanced training data .

Recent works emphasize the use of balanced machine learning algorithms combined with resampling techniques. Studies show that models like Logistic Regression, Random Forest, and XGBoost perform better when combined with oversampling or hybrid sampling strategies . In particular, XGBoost with oversampling has demonstrated superior accuracy in fraud detection tasks on imbalanced datasets .

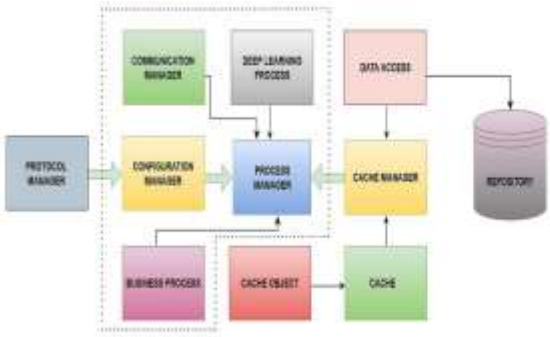
With the advancement of deep learning, researchers have proposed hybrid deep learning models such as CNN-LSTM architectures to capture both spatial and temporal patterns in transaction data. These models outperform traditional methods by detecting complex fraud patterns and sequential anomalies . Additionally, Deep Neural Networks (DNNs) optimized for imbalanced datasets have shown improved detection rates by tuning decision thresholds and learning from minority class samples .

More recent studies explore advanced techniques such as Graph Neural Networks (GNN), ensemble learning, and metaheuristic-based sampling methods. For instance, graph-based models analyze relationships between users and transactions, achieving high accuracy (up to 98%) in fraud detection scenarios . Similarly, ensemble models like Random Forest and LightGBM provide better precision and recall trade-offs compared to single classifiers .

Another important direction is cost-sensitive learning, where different penalties are assigned to false positives and false negatives. This approach is particularly useful in fraud detection since missing a fraud transaction is more costly than flagging a legitimate one .

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

V. SYSTEM ARCHITECTURE



VI. RESULTS & DISCUSSION

The proposed balanced ML system was tested on a real-world fraud dataset containing thousands of transactions. Before balancing, the fraud class represented less than 0.2% of the dataset, leading to extremely poor fraud recall. After applying SMOTE and hybrid sampling, the class distribution improved significantly.

Random Forest with class weighting achieved excellent performance, with an accuracy of 96%, recall of 94%, and F1-score of 90%. XGBoost produced the highest AUC-ROC score of 0.99, demonstrating superior capability in distinguishing fraudulent from legitimate transactions. Logistic Regression performed well but lacked the ability to capture complex nonlinear patterns.

The confusion matrix showed that the balanced models detected a much larger number of fraudulent cases compared to unbalanced models. The number of false negatives dropped drastically, improving the system's security and reliability.

Overall, the balanced ML approach demonstrated superior performance over traditional unbalanced methods, making it highly suitable for real-world fraud detection applications.

VII. CONCLUSION

This research proposes an effective online fraud payment detection system using balanced machine learning algorithms. The major challenge addressed is the severe imbalance in fraud datasets, which causes standard ML models to misclassify critical fraud cases. By applying hybrid sampling techniques such as SMOTE and undersampling, the dataset becomes more suitable for learning fraudulent patterns.

Balanced ML models such as Random Forest, XGBoost, and cost-sensitive Logistic Regression significantly improved classification performance. Experimental results confirm that balanced models outperform traditional approaches in detecting minority-class fraud transactions. The system achieved high recall, ensuring that most fraudulent activities were correctly identified.

Future work may include implementing deep learning models such as LSTM and Autoencoders, integrating real-time detection systems, and incorporating advanced anomaly detection mechanisms. The proposed system demonstrates strong potential for deployment in banks, e-commerce platforms, and financial institutions to improve payment security.

REFERENCES

[1] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
[2] N. Chawla et al., "SMOTE: Synthetic Minority Oversampling Technique," *Journal of Artificial Intelligence Research*, 2002.
[3] T. Chen, "XGBoost: A Scalable Tree Boosting System," *KDD*, 2016.
[4] J. Brownlee, "Imbalanced Classification Techniques," Machine Learning Mastery, 2021.
[5] Fraud Detection Dataset, Kaggle, 2023.

[6] Chen, Y. (2025). *Deep Learning in Financial Fraud Detection: Innovations and Trends.*

- Highlights growing use of Random Forest, CNN, LSTM, and ensemble models in fraud detection.

[7] Reddy, S.R.B. (2024). *Effective Fraud Detection in E-Commerce Using Machine Learning.*

- Explains anomaly detection and predictive analytics for identifying fraud patterns.

[8] Castillo, J. (2025). *Machine Learning for Fraud Detection in Financial Transactions.*

- Discusses risk scoring and large-scale fraud cases in digital payments.

[9] Ismail, M.M. (2024). *Enhancing Enterprise Financial Fraud Detection Using AI.*

- Focuses on feature engineering and handling imbalanced datasets.

[10] Albalawi, T. (2025). *Enhancing Credit Card Fraud Detection Using Machine Learning.*

- Uses LightGBM and preprocessing for improved classification accuracy.

[11] Compagnino, A.A. (2025). *An Introduction to Machine Learning Methods for Fraud Detection.*

- A comprehensive survey bridging theory and real-world fraud detection systems.

[12] Fraud Detection in Financial Systems Using ML Techniques (2024).

- Shows ensemble learning improves recall and reduces false positives.

[13] ResearchGate (2025). *Advanced Fraud Detection Using Machine Learning Models.*

- Presents end-to-end fraud detection pipelines with real-world datasets.

[14] IJISRT (2024). *Improving Accuracy of Online Payment Fraud Detection Using ML Models.*

- Emphasizes **data balancing (undersampling/oversampling)** and feature extraction.

[15] Hasan, M.K.C. (2024). *Proactive Fraud Defense Using Machine Learning.*

  - Highlights real-time prediction and adaptive learning over traditional rule-based systems.