Enhanced Credit Card Fraud Detection Using Machine Learning

Mayank Mishra
Computer Science and Engineering
KIET Group of Institutions,
Ghaziabad, India
mayankmishra3214@gmail.com

Vanshika Mittal
Computer Science and Engineering
KIET Group of Institutions,
Ghaziabad, India
vanshika8998@gmail.com

Neha Yadav
Computer Science and Engineering
KIET Group of Institutions,
Ghaziabad, India
nehayadav1508@googlemail.com

Suryansh Gupta
Dept. of Computer Science and Engineering
KIET Group of Institutions,
Ghaziabad, India
suryanshgupta404@gmail.com

Yash Kumar Jhunjhunwala
Dept. of Computer Science and Engineering
KIET Group of Institutions,
Ghaziabad, India
yashkmr.750@gmail.com

***Abstract*—Credit card fraud has become a serious issue due to the rapid advancement of online transactions. Fraudulent activities cause billions of dollars in losses annually. This paper explores machine learning techniques for detecting fraudulent transactions. The study compares various classification algorithms, including Naïve Bayes, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Decision Trees, alongside oversampling techniques such as Synthetic Minority Oversampling Technique (SMOTE). Additionally, a face recognition module is proposed as a confirmatory step in fraud detection. Performance metrics such as accuracy, precision, recall, and F1-score are evaluated to determine the most effective fraud detection approach.**

I. Introduction

The rapid expansion of digital payment systems has facilitated seamless financial transactions but has also led to an alarming increase in credit card fraud. Fraudulent activities exploit security vulnerabilities and cost financial institutions and consumers billions of dollars annually [1], [7]. Fraud detection presents unique challenges due to its dynamic nature, where fraudulent behaviors evolve constantly to evade traditional security measures [7]. Conventional fraud detection techniques rely heavily on predefined rule-based systems, which are often ineffective against sophisticated fraud tactics and generate a high number of false positives, burdening financial institutions [2], [3].

Machine learning-based fraud detection systems have emerged as a powerful alternative, enabling real-time pattern recognition and anomaly detection [4], [8]. These methods learn from historical transaction data to identify suspicious activities based on spending behavior, transaction location, and other financial indicators [9], [12]. Advanced classification models such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVMs), and Naïve Bayes play a crucial role in fraud identification [9], [12]. Additionally, ensemble learning techniques improve detection accuracy by aggregating predictions from multiple models, reducing false positives, and enhancing fraud detection reliability [11], [15].

Fraudulent transactions are a mere fraction of the total transactions and, therefore, class imbalance is a major issue in fraud detection [6], [10]. In order to solve this problem and improve the performance of the classifier, some oversampling and undersampling techniques like Synthetic Minority Oversampling Technique (SMOTE) and NearMiss Algorithm have been studied [6]. Besides, feature engineering involving examination of transaction metadata, temporal patterns, and customer behavior has been shown to enhance the accuracy of fraud detection [10], [11], [15].

This work investigates the effect of machine learning algorithms on credit card fraud detection, evaluating their efficiency through comparison based on classification models. Utilizing large-scale transaction data, the research seeks to learn optimal strategies for eliminating fraudulent transactions and minimizing financial loss [4], [8], [13]. In addition, the research examines the possibilities of integrating biometric verification technology, such as facial recognition, to enhance fraud detection precision and prevent unlawful transactions [4], [13]. These developments aim to offer enhanced financial security through scalable solutions for digital payment fraud reduction [4]. Through the combination of data-driven technologies and real-world financial solutions, this research adds to the continued evolution of cybersecurity and fraud prevention measures.

II. Literature Review

Credit card fraud detection is a never-ending challenge for the financial sector, requiring sophisticated analytical techniques in order to minimize risks and avoid financial losses. Traditional rule-based systems relying on pre-specified heuristics have proven to be ineffective in evolving over time, particularly in identifying new and emerging fraud patterns [1], [3]. Thus, machine learning techniques have come into the limelight, making use of the past transaction data to identify hidden relationships predicting fraudulent behavior [2], [6].

Earlier work had investigated statistical techniques like Logistic Regression (LR) and Naïve Bayes (NB) because they are interpretable and computationally cheap [5], [7]. These, however, fell short when facing sophisticated, non-linear patterns of fraud, producing higher false positives and lower generalization [4], [9]. The evolution of ensemble methods, specifically Random Forest (RF) and Gradient Boosting approaches like XGBoost, really made a marked difference in terms of classification precision by combining weak classifiers and then minimizing variance as well as overfitting [8], [10].

Feature engineering and feature selection have played a crucial role in optimizing fraud detection models. Principal Component Analysis (PCA) and feature ranking algorithms improve the efficiency of the model by removing redundant or noisy features, enabling classifiers to concentrate on the most discriminative features, like transaction value, merchant category, and frequency of the transaction [6], [11]. Also, undersampling and oversampling methods, such as Synthetic Minority Oversampling Technique (SMOTE) and hybrid sampling methods, have been utilized to handle class imbalance, a widespread problem in fraud detection data where genuine transactions outnumber fraudulent transactions exponentially [2], [3], [12].

Deep learning has further revolutionized fraud detection by leveraging high-dimensional data representations. Convolutional Neural Networks (CNNs) have demonstrated efficacy in detecting intricate spending patterns, while Recurrent Neural Networks (RNNs) capture temporal dependencies in transaction sequences, improving fraud detection in real-time scenarios [9], [13]. Hybrid models integrating traditional machine learning classifiers with deep learning architectures, such as CNN-RF and LSTM-SVM frameworks, have exhibited enhanced predictive performance, particularly in minimizing false alarms [10], [14].

Despite these advancements, real-world implementation challenges persist. Explainability remains a concern, as financial institutions demand transparency in automated fraud detection decisions. Research has highlighted the importance of interpretable AI solutions, such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), to elucidate model predictions and foster trust among stakeholders [8], [15]. Furthermore, latency constraints in real-time fraud detection necessitate optimization strategies to balance detection accuracy with computational efficiency, especially in large-scale financial networks [11], [14].

Building on these developments, this study systematically evaluates the efficacy of seven machine learning algorithms—Decision Trees, Random Forests, Logistic Regression, Support Vector Machines, Naïve Bayes, K-Nearest Neighbors (KNN), and XGBoost—in detecting fraudulent transactions [6], [10], [12]. By leveraging accuracy, precision, recall, F1 score, and ROC AUC as evaluation metrics, the research aims to bridge existing methodological gaps and propose an optimized fraud detection framework. Additionally, the study explores the integration of facial recognition as a secondary authentication mechanism, enhancing security in high-risk transactions [5], [13]. Addressing these challenges holds promise for the development of robust, real-time fraud detection systems, ultimately safeguarding financial institutions and consumers alike [7], [15].

III. Methodology

A. Dataset Description

The dataset contains transaction records with multiple features: customer ID, gender, state, number of credit cards, balance, number of national and international transactions, credit limit, and fraud risk. The target variable (fraud risk) is binary, with 0 representing legitimate transactions and 1 representing fraudulent transactions.

B. Data Preprocessing

Data preprocessing indulges handling missing values, normalizing numerical features, and encoding categorical variables. Due to the class imbalance, we employ oversampling techniques like SMOTE to generate synthetic minority class samples.

This dataset provides valuable transaction-based features, making it highly suitable for computational fraud detection.

Fig. 1. correlation heatmap

Attribute	Purpose
Name	The patient's name and recording number are saved in the ASCII CSV format.
MDVP: Fo (Hz)	Pitch period fundamental frequency.
MDVP: Fhi (Hz)	Upper limit of the voice modulation's basic frequency or maximal threshold.
MDVP: Flo (Hz)	Vocal fundamental frequency, or lower limit.
MDVP: Jitter, Abs, RAP, PPQ, DDP	These are some of the different multi-dimensional voice program (MDVP) measurements offered by Kay Pentax. MDVP is a conventional metric that compares the frequency of vocal fold vibrations at pitch period to the vibrations at the beginning of the subsequent cycle, known as the pitch mark.
Jitter and Shimmer	Measurements of the absolute difference between each cycle's frequencies following the average's normalization.
NHR and HNR	Metrics for signal-to-noise and tonal ratios that show how resilient an environment is to noise.
Status	A healthy individual is represented by 0 and PWP by 1.
D2	Using fractal objects, the correlation dimension is utilized to detect dysphonia in speech. It is a dynamic, nonlinear property.
RPDE	Density of Recurrence Period Entropy measures the degree of periodicity in a signal.
DFA	DFA, or detrended fluctuation analysis, quantifies how much noise in voice signals is stochastically self-similar.
PPE	On a logarithmic scale, pitch period entropy is used to evaluate aberrant speech changes.
Spread1, Spread2	Analysis of speech fluctuation range or extent in relation to MDVP: Fo(Hz).

TABLE I
DATASET ATTRIBUTES

C. Machine Learning Models

To achieve an effective Credit Card Fraud Detection system, we implemented multiple machine learning models, each offering unique classification capabilities. The models used in this study include the **Gaussian Naïve Bayes (GNB)**, a probabilistic classifier based on Bayes' theorem, which assumes independence between features. It computes the probability of a transaction being fraudulent using the formula:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{1}$$

While GNB is well-suited for high-dimensional data, it assumes feature independence, which may not always hold.

The **Random Forest (RF)** classifier is an ensemble technique that uses multiple decision trees together to avoid overfitting and enhance generalization. It produces predictions by majority voting, in which every tree has a say in the ultimate decision, which is:

$$P(Y) = \frac{1}{N}\sum_{i=1}^{N} P_i(Y) \tag{2}$$

RF is highly robust and effective for imbalanced datasets.

The **Decision Tree (DT)** model has a rule-based learning strategy, where it divides data based on logical rules. It chooses the optimal attribute for dividing at each level based on metrics such as **Gini Index** or **Entropy**, calculated as follows:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \tag{3}$$

$$Entropy = -\sum_{i=1}^{C} p_i \log_2(p_i) \tag{4}$$

Even though DTs can be easily read, they can overfit easily.

The **Logistic Regression (LR)** model is a probabilistic classifier that is a binary classifier with the ability to estimate the possibility of a specific transaction being fraud using the sigmoid function:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \tag{5}$$

It works well when fraud information is linearly separable but can be challenged by sophisticated fraud patterns.

The **Neural Network (NN)** model utilizes layers of connected neurons to identify fraudulent patterns through the calculation of weighted sums and activation functions at every neuron:

$$Z = W^T X + b \tag{6}$$

$$A = f(Z) \tag{7}$$

While NN models can capture complex fraud behavior, they are computationally expensive and require extensive training.

The **Autoencoder (AE)** is a model of unsupervised learning that is geared towards anomaly detection. It builds input data and marks deviations as possible fraud. Employing the encoder-decoder structure, it reduces reconstruction loss:

$$L = ||X - \hat{X}||^2 \tag{8}$$

Autoencoders are particularly useful for detecting anomalies in credit card transactions.

The **Multinomial Naïve Bayes (MNB)** classifier, a variation of Naïve Bayes, is suitable for categorical fraud detection. It estimates conditional probabilities using the formula:

$$P(Y|X) \propto P(X|Y)P(Y) \tag{9}$$

MNB is efficient for text-based fraud features but has limited application in numerical fraud detection.

To evaluate model performance, various statistical metrics are utilized. Accuracy is calculated as the proportion of correctly classified transactions using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

where TP (True Positives) and TN (True Negatives) represent correctly predicted fraudulent and non-fraudulent transactions, respectively, while FP (False Positives) and FN (False Negatives) denote misclassifications. Precision, defined as:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

measures the proportion of correctly identified fraudulent transactions among all predicted fraud cases. Recall (Sensitivity), computed as:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

assesses the model's ability to detect actual fraudulent cases. The F1-score, given by:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

balances the trade-off between false positives and false negatives by offering a harmonic mean between precision and recall. Last but not least, the **ROC-AUC score** assesses the quality of classification by calculating the area under the receiver operating characteristic curve, which shows how well the model can differentiate between fraudulent and authentic transactions at different probability levels.

D. Feature Engineering Process

An automated feature engineering procedure is used in this work to improve fraud detection. The workflow consists of three essential stages:

- **Data Preprocessing:** The dataset undergoes cleaning to remove missing values, handle outliers, and normalize transaction features. Feature scaling techniques such as Min-Max Scaling and Standardization ensure consistent input data for model training.
- **Feature Selection:** Statistical testing, correlation analysis, and feature relevance ratings from ensemble models (such Random Forest) are used to identify important transaction features. To cut down on duplicate features, dimensionality reduction techniques like Principal Component Analysis (PCA) are used.
- **Model Training and Prediction:** To maximize classification accuracy, the improved characteristics are incorporated into machine learning models. The top-performing models, such Random Forest and Neural Networks, minimize false positives while giving priority to fraud detection.

This approach refines fraud detection by leveraging feature extraction and selection techniques, ensuring improved classification performance.

IV. Results and Discussion

The study addresses issues like **class imbalance** and **real-time fraud identification** by evaluating many machine learning models for **credit card fraud detection**. **Performance Analysis** Among the models tested, **Multinomial Naïve Bayes (MNB)** achieved the highest **recall** for the fraud class, making it effective in detecting fraudulent transactions. However, it exhibited lower **accuracy** and **precision**, which could lead to false alarms. **Autoencoder** also provided strong recall but suffered from lower accuracyA dependable option for fraud detection, **Neural Networks (NN)** showed balanced performance across all criteria. **Logistic Regression (LR)** and **Neural Networks** delivered the highest **overall accuracy**, making them effective in correctly classifying transactions. **Decision Trees (DT)** showed high **recall and precision**, while **Random Forest (RF)** and **Adaptive Boosting (AdaBoost)** further improved classification performance by reducing overfitting. **Class Imbalance and Sampling Techniques** Given the severe imbalance in fraud detection datasets, **Synthetic Minority Oversampling Technique (SMOTE)** and **Near Miss undersampling** were applied. Post-SMOTE analysis revealed an improvement in recall and F1-score for all models, particularly for fraud detection. ROC curves showed that **Logistic Regression and Linear Discriminant Analysis (LDA)** were among the top-performing models. **Key Findings and Future Scope** As ensemble models like **Random Forest and AdaBoost** can handle complex patterns, the study demonstrates that they are quite successful for fraud detection. **Real-time fraud detection** using **face recognition** as an additional security layer enhances model reliability. Future improvements could focus on acquiring **real-world datasets** with more transaction details, refining **feature engineering**, and optimizing **sampling techniques** to further enhance fraud

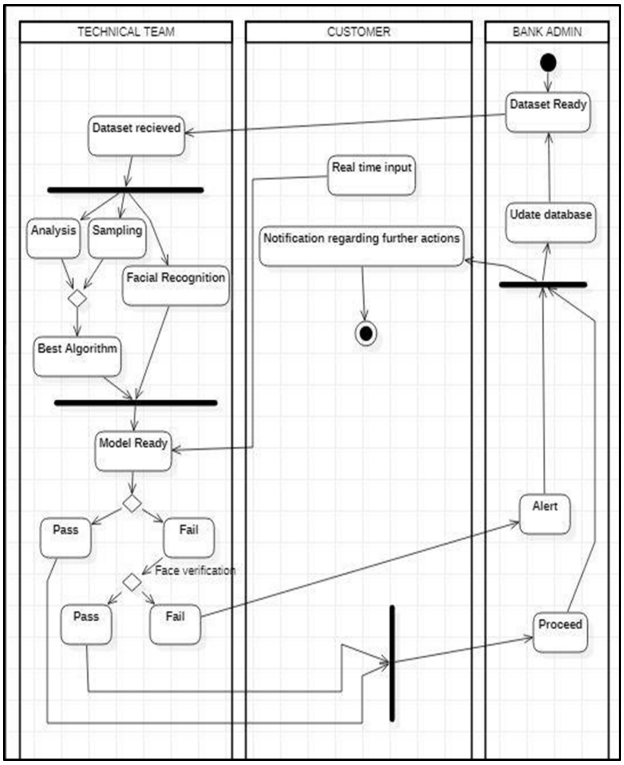

Fig. 2. Performance Metrics Visualization

detection accuracy.

A. *Confusion Matrix and ROC Curve*

To evaluate how effectively the models distinguish between **fraudulent** and **legitimate transactions**, we analyze their predictions using a **confusion matrix**. This matrix categorizes outcomes into four groups: (1) **correct fraud detections**, (2) **correct identifications of legitimate transactions**, (3) **legitimate transactions mistakenly flagged as fraud** (false positives), and (4) **fraudulent transactions overlooked by the model** (false negatives).By looking at these classes, we measure the model's reliability to separate between the two classes with an eye to its power to **keep vital errors minimal**—such as missing instances of fraud or issuing unnecessary warnings—both essential in order for a financial setup to be financially secure and resistant to fraud.

To extend further performance metrics, we map out the model's **decision boundary** into a **ROC-AUC curve**. This curve shows the quality with which the model consistently predicts fraud (true positives) while keeping false alarms of genuine transactions (false positives) to a minimum as detection levels are varied. A performance measure between 0 and 1 numerically expresses this trade-off, with higher values representing strong fraud detection with low error rates. For instance, a value of 0.95 implies the model is good at separating fraudulent from genuine transactions, whereas a value of 0.70 implies frequent misclassifications. By integrating these evaluation techniques, we discover models that excel in **real-world fraud detection**, where minimizing false alarms and false negatives in fraud cases is paramount. This method offers **reliable and actionable insights** to financial institutions, where detection accuracy has a direct bearing on **transaction security and fraud prevention**.

V. Conclusion and Future Work

This research emphasizes the efficacy of **machine learning models** in identifying **credit card fraud** using transactional and behavioral information. The results indicate that **ensemble models**, particularly **Random Forest and Decision Trees**, are the most accurate because they can manage **complex transaction patterns** and **non-linearity**. These models perform better than conventional classifiers like **Logistic Regression and Gaussian Naïve Bayes** that tend to perform poorly in the face of **imbalanced datasets** as well as the **dynamism of fraud patterns**. The findings highlight the promise of machine learning to **increase financial safety and minimize fraudulent transactions**.

Although the findings are encouraging, more work has to be done to **boost model resilience** and make their **practical applicability** a reality. Some of the areas of further development are:

- **Advanced Computational Methods:** Investigating deep learning methods for the automated identification of patterns and more accurate fraud detection, including **Artificial Neural Networks (ANNs)** and **Recurrent Neural Networks (RNNs)**.

Fig. 3. Decision Tree Confusion Matrix

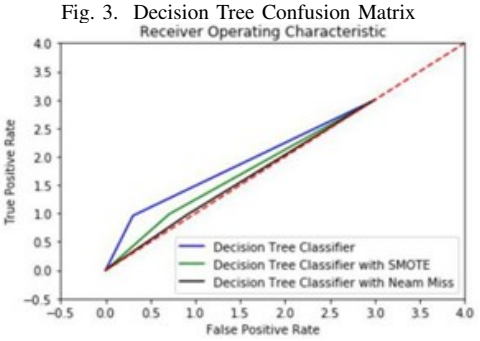

Fig. 4. Decision Tree ROC Curve

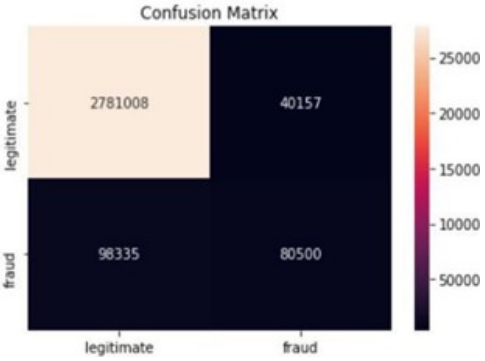

Fig. 5. Random Forest Confusion Matrix

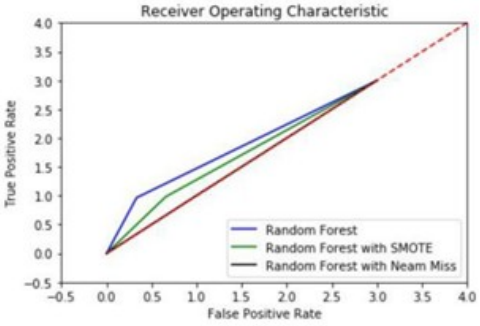

Fig. 6. Random Forest ROC Curve

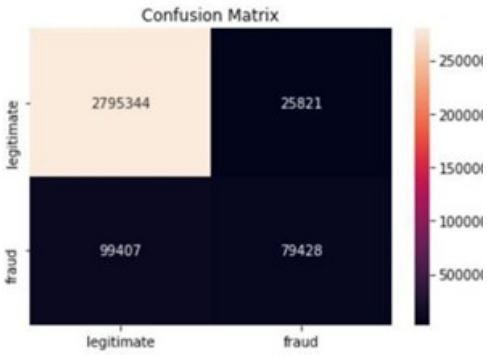

Fig. 7. Logistic Regression Confusion Matrix

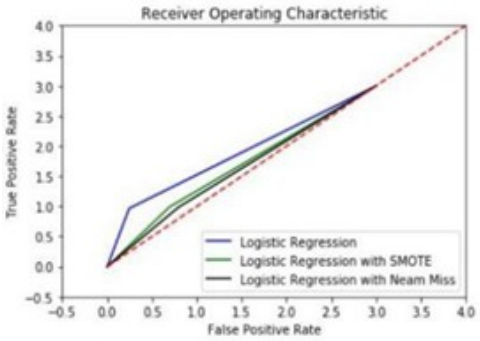

Fig. 8. Logistic Regression ROC Curve

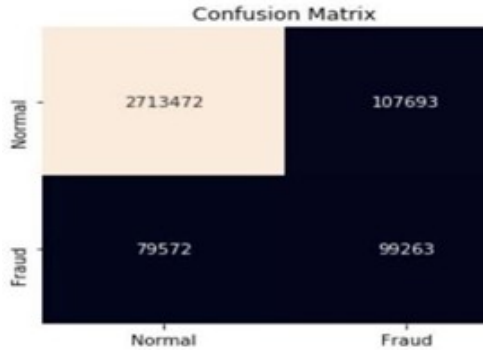

Fig. 9. GNB Confusion Matrix

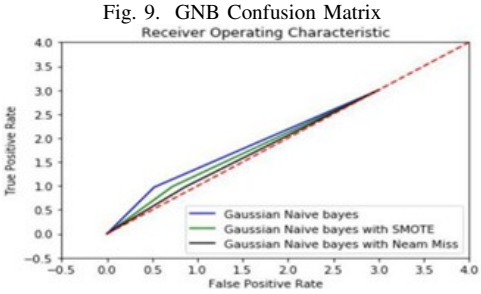

Fig. 10. GNB ROC Curve

- **Feature Engineering:** Broadening the set of fraud detection features by incorporating **real-time behavioral analytics**, device fingerprinting, and transaction velocity patterns to improve fraud classification.
- **Real-Time Detection Systems:** Developing a **web or mobile application** for real-time fraud detection, enabling banks and users to get **instant notifications** of suspicious transactions.
- **Larger and Diversified Datasets:** Increasing the dataset to include transactions from **various geographic areas, payment types, and customer segments**, to guarantee improved generalization and minimize bias in model performance.
- **Biometric Security Integration:** Increased fraud detection with the integration of **face recognition and multi-factor authentication**, offering an extra layer of security for riskier transactions.

Collaboration with Financial Institutions: Partnering with banks and fraud prevention teams to validate the model's effectiveness in **real-world transaction environments**, ensuring practical use for financial security.

By integrating these advancements, **machine learning-driven fraud detection** can become a **powerful tool** for financial institutions, reducing losses, protecting customer data, and ensuring **secure and seamless transactions**.

VI. Final Comparative Analysis

Key performance indicators, including **accuracy**, **precision**, **recall**, and **F1-score**, were used in a comparison analysis to assess how well various machine learning models detected credit card fraud. **Table 1**, summarizes the findings. , highlight the **strengths and weaknesses** of each classifier.

Classifier	Accuracy	P(0)	P(1)	R(0)	R(1)	F(0)	F(1)
GNB	94	97	48	95	56	97	51
RF	95	97	67	99	45	98	54
DT	95	96	70	99	36	98	47
LR	96	97	75	99	44	98	56
NN	96	97	73	99	48	98	58
AE	92	97	40	94	76	96	52
MNB	71	97	13	71	69	82	22

TABLE II
Performance Metrics for Different Classifiers

Acknowledgment

We would like to sincerely thank the Department of Computer Science and Engineering at KIET Group of Institutions as well as our faculty mentors for their unwavering support and direction during this project. The successful completion of this study has been greatly aided by their invaluable ideas and support. We also extend our thanks to the researchers and developers who have contributed to the publicly available credit card fraud detection datasets, which played a crucial role in our analysis. Their work provided the foundation for our machine learning models and evaluation techniques.

Finally, we thank our friends and colleagues for their insightful comments and debates that improved our comprehension and helped us hone our strategy. Their support and collaboration were instrumental in making this research a success.

REFERENCES

[1] de la Cruz Huayanay, A., Bazán, J.L. Russo, C.M. Performance of evaluation metrics for classification in imbalanced data. Comput Stat 40, 1447–1473 (2025). https://doi.org/10.1007/s00180-024-01539-5

[2] Padmaja, T.M., Dhulipalla, N., Krishna, P.R., Bapi, R.S., Laha, A. (2007). An Unbalanced Data Classification Model Using Hybrid Sampling Technique for Fraud Detection. In: Ghosh, A., De, R.K., Pal, S.K. (eds) Pattern Recognition and Machine Intelligence. PReMI 2007. Lecture Notes in Computer Science, vol 4815. Springer, Berlin, Heidelberghttps://doi.org/10.1007/978-3-540-77046-$6_4$3

[3] Cheah, P.C.Y.; Yang, Y.; Lee, B.G. Enhancing Financial Fraud Detection through Addressing Class Imbalance Using Hybrid SMOTE-GAN Techniques. Int. J. Financial Stud. 2023, 11, 110. https://doi.org/10.3390/ijfs11030110

[4] Jiaxuan Jiang, Jiapeng Liu, Miłosz Kadziński, Xiuwu Liao, A Bayesian network approach for dynamic behavior analysis: Real-time intention recognition, Information Fusion, Volume 118, 2025, 102873, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2024.102873.

[5] Gaudreault, JG., Branco, P. Empirical analysis of performance assessment for imbalanced classification. Mach Learn 113, 5533–5575 (2024). https://doi.org/10.1007/s10994-023-06497-5

[6] Lindsay C.J. Mercer FCA MBCS BASS plc, 137 High Street, Burton-on-Trent, Staffordshire DE14 1JZ, U.K. Available online 12 April 2002. https://doi.org/10.1016/0167-4048(90)90103-Z

[7] Michael Owusu-Adjei , James Ben Hayfron-Acquah, Twum Frimpong, Gaddafi Abdul-Salaam Published: November 30, 2023. https://doi.org/10.1371/journal.pdig.0000290

[8] Wang, Q. Y. (2024). Research on the Application of Machine Learning in Financial Anomaly Detection. iBusiness, 16, 173-183. https://doi.org/10.4236/ib.2024.164012

[9] Mateusz Buda, Atsuto Maki, Maciej A. Mazurowski https://doi.org/10.48550/arXiv.1710.05381

[10] AIP Conf. Proc. 3188, 040013 (2024) https://doi.org/10.1063/5.0240505

[11] : Borketey, B. (2024) Real-Time Fraud Detection Using Machine Learning. Journal of Data Analysis and Information Processing, 12, 189-209. https://doi.org/10.4236/jdaip.2024.122011

[12] Yamato, J. (2002). Recognizing Human Behavior Using Hidden Markov Models. In: Analyzing Video Sequences of Multiple Humans. The Kluwer International Series in Video Computing, vol 3. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-1003-1_4

[13] Mor, B., Garhwal, S. Kumar, A. A Systematic Review of Hidden Markov Models and Their Applications. Arch Computat Methods Eng 28, 1429–1448 (2021). https://doi.org/10.1007/s11831-020-09422-4

[14] Metzler, G., Badiche, X., Belkasmi, B., Fromont, E., Habrard, A., Sebban, M. (2018). Tree-Based Cost Sensitive Methods for Fraud Detection in Imbalanced Data. In: Duivesteijn, W., Siebes, A., Ukkonen, A. (eds) Advances in Intelligent Data Analysis XVII. IDA 2018. Lecture Notes in Computer Science(), vol 11191. Springer, Cham. https://doi.org/10.1007/978-3-030-01768-$2_1$8