JMRR
Innovate •Share •Explore

JOURNAL OF MULTIDIMENSIONAL RESEARCH & REVIEW

http://www.jmrr.org Volume: 6, Issue: 1, April 2025 | ISSN: 2708-9452

# Credit Card Fraud Detection Using Machine Learning

**Dr Gayathiri A**
Assistant Professor,Department of Computer Sciences and Applications Vivekanandha College of Arts and Sciences for Women [Autonomous],Tiruchengode, Namakkal, Tamilnadu, India.
**Nehasri M**
MCA Student,Department of Computer Science and Applications
Vivekanandha College of Arts and Sciences for Women [Autonomous],Tiruchengode
Namakkal, Tamilnadu, India.

## Abstract

Credit card fraud detection is a crucial application of machine learning that helps financial institutions identify fraudulent transactions and prevent financial losses. Traditional rule-based fraud detection systems struggle with evolving fraud patterns, making machine learning models more effective due to their ability to learn complex relationships in transaction data. This study employs a Support Vector Machine (SVM) classifier to detect fraudulent transactions in a highly imbalanced dataset. The dataset is analyzed for fraud distribution, and key features are selected using correlation analysis. To address class imbalance, a subset of non-fraudulent transactions is sampled to balance the dataset. The SVM model is trained on this balanced dataset and evaluated using a confusion matrix to measure accuracy and fraud detection rate. Additionally, a weighted SVM classifier is implemented to enhance fraud detection by assigning higher importance to fraudulent transactions. The results demonstrate the effectiveness of machine learning techniques in credit card fraud detection, highlighting the impact of feature selection, class balancing, and model weighting in improving fraud classification accuracy.

**Keywords:** SVM, Classification technique, Transactions, Credit card fraud detection

# 1 Introduction

## 1.1 Credit Card

Credit card fraud has become a significant concern in the financial sector, with fraudulent transactions causing substantial economic losses and security risks. Traditional fraud detection systems rely on rule-based methods, which struggle to adapt to evolving fraud techniques. Machine learning offers a more effective approach by analyzing large datasets and identifying fraudulent patterns based on historical transaction data. However, a major challenge in fraud detection is the highly imbalanced nature of the data, where fraudulent transactions constitute a small fraction of the total transactions. To address this, techniques such as feature selection, sampling, and weighted classification are employed. In this study, a Support Vector Machine (SVM) classifier is utilized for fraud detection, with a focus on improving classification performance through feature selection and class balancing. By evaluating the model's accuracy and fraud detection rate, this research highlights the potential of machine learning in enhancing financial security and preventing fraudulent activities.

## 1.2 Machine Learning

Machine learning is a branch of artificial intelligence that enables systems to learn patterns from data and make predictions or decisions without being explicitly programmed. It is widely used across various domains, including healthcare, finance, cybersecurity, and autonomous systems, due to its ability to analyze large datasets and uncover complex relationships. Machine learning techniques are broadly categorized into supervised, unsupervised, and reinforcement learning. Supervised learning, which includes algorithms like Support Vector Machines (SVM), Decision Trees, and Neural Networks, is commonly used for classification and regression tasks. Unsupervised learning, such as clustering and anomaly detection, helps identify hidden patterns in unlabeled data. Reinforcement learning, inspired by behavioral psychology, enables systems to learn through rewards and penalties. Despite its advantages, machine learning faces challenges such as data quality, model interpretability, and ethical concerns. Nevertheless, its continuous advancements are driving innovation in automation, decision-making, and predictive analytics across industries.

## 1.3 SVM Algorithm

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data points belonging to different classes in a high-dimensional space. SVM uses support vectors, which are the data points closest to the hyperplane, to maximize the margin between different classes, ensuring better generalization. It can handle both linear and non-linear classification problems by utilizing kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid to map data into higher-dimensional spaces where it becomes linearly separable. SVM is particularly effective in scenarios with high-dimensional data and small sample sizes, making it widely used in applications like image recognition, text classification, and fraud detection. Despite its advantages, SVM can be computationally intensive for large datasets and requires careful selection of hyper parameters such as the kernel type and regularization parameter to achieve optimal performance.

### 1.4 Fraud Detection

Fraud detection is a critical application of data analysis and machine learning, aimed at identifying and preventing fraudulent activities and transactions in various domains. As technology advances and financial transactions increasingly shift to digital platforms, the risk of fraudulent behavior has grown, making fraud detection more crucial than ever. Fig 0.1 shows that the whether it is credit card fraud, insurance fraud, identity theft, or online scams, businesses and financial institutions employ sophisticated fraud detection systems to safeguard their assets, protect their customers, and maintain trust in their services. Fraud detection systems analyze vast amounts of transactional and behavioral data to identify irregularities and suspicious patterns. The impact of fraud can be significant, resulting in financial losses, damaged reputations, and compromised customer trust. Fraud detection systems play a crucial role in preventing such consequences by proactively identifying and flagging suspicious transactions. For financial institutions, effective fraud detection not only protects their assets but also ensures compliance with regulatory requirements and enhances customer confidence.

### 1.5 Objectives

- Implement a Support Vector Machine (SVM) classifier to accurately detect fraudulent credit card transactions.
- Apply techniques such as feature selection, data sampling, and weighted classification to improve fraud detection in highly imbalanced datasets.
- Evaluate the impact of feature selection and class balancing on the accuracy, precision, recall, and fraud detection rate of the SVM model.
- Analyze the effectiveness of a standard SVM classifier versus a weighted SVM in improving fraud detection without compromising overall accuracy.
- Provide insights into how machine learning can enhance fraud detection systems, helping financial institutions reduce losses and improve transaction security

## 2 Literature Review

Credit card fraud detection (CCFD) has become a critical concern due to increasing online transactions and technological advancements. This section summarizes recent research exploring machine learning (ML) and deep learning approaches for fraud detection. ML and DRL Based on Resampling Approaches for CCF Detection Tran Khanh Dang et al. addressed the imbalanced dataset problem in CCF using resampling methods (SMOTE and ADASYN) and ML algorithms. Two resampling approaches were tested—resampling before and after train-test split. Models achieved over 99% accuracy with full-dataset resampling but showed significant performance drops when resampling only training data, especially for logistic regression. Deep reinforcement learning (DRL) performed poorly (34.8 % accuracy), showing limitations for imbalanced data handling[**1**].

Systematic Review of Fraud Detection in the Era of Disruptive Technologies Asma Cherif et al. reviewed 40 studies from 2015–2021, categorizing research by ML techniques, class imbalance handling, and emerging technologies like IoT and AI. The review found limited deep learning research and emphasized the need for big data analytics, cloud computing, and robust ML frameworks to tackle increasingly sophisticated fraud techniques in the post-COVID era[**2**].
Interpretable Autoencoders for Fraud Detection Jacobo Chaquet-Ulldemolins et al. proposed an interpretable ML approach using autoencoders for nonlinear analysis while maintaining compliance with regulations. A novel single transaction-level explanation (STE) technique was introduced to improve traceability. Their method outperformed previous models in accuracy (by 5.5% and 1.5%) and enabled individualized decision interpretations, addressing the ”black-box” issue in AI[**3**].
Hybrid ML Architecture for CCF Detection Esraa Faisal Malik et al. developed and evaluated seven hybrid ML models for fraud detection using a real-world dataset. The AdaBoost + LGBM hybrid model performed best. The study highlighted the importance of exploring multiple hybrid models over single algorithms to improve fraud detection performance in evolving financial ecosystems[**4**].

# 3 Proposed System

The proposed system is designed to classify transaction data using machine learning techniques, focusing on detecting anomalies that indicate fraudulent activities. A Support Vector Machine (SVM) classifier is utilized due to its ability to handle high-dimensional data and effectively separate different classes through an optimal decision boundary. The system begins with data preprocessing, where missing values are handled, and relevant features are selected to enhance model performance. Since the dataset exhibits an imbalanced distribution, techniques such as under sampling and weighted classification are incorporated to ensure that the minority class is adequately represented during training. The classifier is trained on a refined dataset, and its performance is assessed using evaluation metrics such as accuracy, precision, recall, and detection rate. Additionally, a comparative analysis is conducted between standard and weighted SVM approaches to determine their effectiveness in identifying patterns within the data. By implementing feature selection and balancing techniques, the system aims to improve classification efficiency, reducing misclassification errors and enhancing overall model robustness.

## 3.1 Advantages

- The proposed system enhances classification accuracy by utilizing a Support Vector Machine (SVM) classifier, which efficiently handles high-dimensional data and constructs an optimal decision boundary for effective separation of different classes.

- The integration of feature selection techniques improves performance by eliminating less relevant attributes, reducing computational complexity, and increasing interpretability.

- Class balancing strategies address data distribution challenges, ensuring that all categories contribute meaningfully to the learning process, leading to a more stable and fair classification model.

- The incorporation of weighted classification enhances the detection capability for less frequent categories, improving the overall classification reliability.
- The evaluation of different classification approaches allows for a thorough comparative analysis, ensuring the selection of the most effective method to optimize performance and reduce misclassification errors.

# 4 Methodology

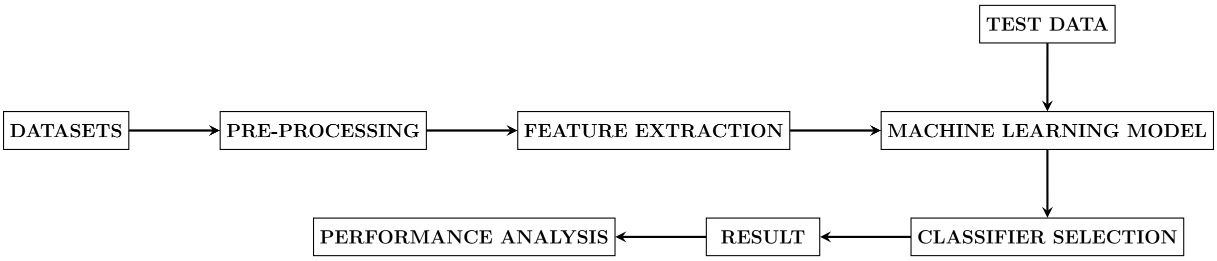

Figure 1: Flowchart

## 4.1 Load Data

This module is responsible for importing structured transaction data from a predefined source into a suitable format for analysis. The dataset is loaded into a structured framework where each data point represents an instance with multiple attributes. Proper data loading ensures that the information is accessible without inconsistencies or missing entries. Handling different data formats, ensuring compatibility with the classification algorithm, and verifying the completeness of the dataset are crucial steps in this module. The Support Vector Machine (SVM) algorithm requires well-structured input data for effective classification, making this step essential for the subsequent stages.

## 4.2 Data Preprocessing

The pre-processing module cleans and transforms the dataset to improve its quality and usability for classification. Steps in this module include handling missing values through imputation techniques, normalizing numerical attributes to standardize the data, encoding categorical variables into numerical representations, and removing duplicate records. Additionally, data transformation techniques such as logarithmic scaling or outlier detection are applied to enhance classification accuracy. Proper preprocessing ensures that the SVM classifier receives well-refined data, reducing the risk of biases and inconsistencies that could negatively impact classification performance.

## 4.3 Feature Selection

This module focuses on identifying the most relevant attributes that contribute significantly to classification accuracy while eliminating redundant and less informative features. Feature selection improves computational efficiency and prevents overfitting by reducing the dimensionality of the dataset. Techniques such as correlation analysis, mutual information, or ranking-based selection methods are applied to retain only the most significant variables. Since SVM relies on finding an optimal hyperplane to separate different classes, selecting the most distinguishing features enhances its ability to accurately classify instances while minimizing unnecessary computations.

## 4.4 Training and Testing

This module is responsible for dividing the dataset into separate subsets for model training and evaluation. The dataset is split into training and testing sets using a predefined ratio, ensuring that the model learns from a sufficient amount of data while being evaluated on unseen instances. The SVM classifier is trained on the training subset, where it learns patterns and relationships between different attributes to create a decision boundary that effectively separates categories. Kernel functions such as linear, polynomial, or radial basis function (RBF) can be utilized depending on the complexity of the dataset. Once trained, the model is tested on the evaluation subset to assess its generalization capability and ensure it performs well on unseen instances.

## 4.5 Evaluation and Performance

This module measures the effectiveness of the trained SVM classification model using various performance metrics. Accuracy, precision, recall, F1-score, and detection rate are computed to evaluate the model's efficiency in distinguishing between different classes. A confusion matrix is generated to analyze classification errors and assess the model's strengths and weaknesses. Additionally, comparative analysis is conducted between a standard SVM and a weighted SVM classifier to determine which approach provides better classification performance. By evaluating these metrics, necessary adjustments can be made to optimize the model and enhance its predictive capabilities.

# 5 Result and Discussion

| **Algorithm** | **Training Accuracy** | **Validation Accuracy** |
|---|---|---|
| **SVM** | 95.744% | 98.449% |

Table 1: Training and Validation Accuracy for SVM Algorithm

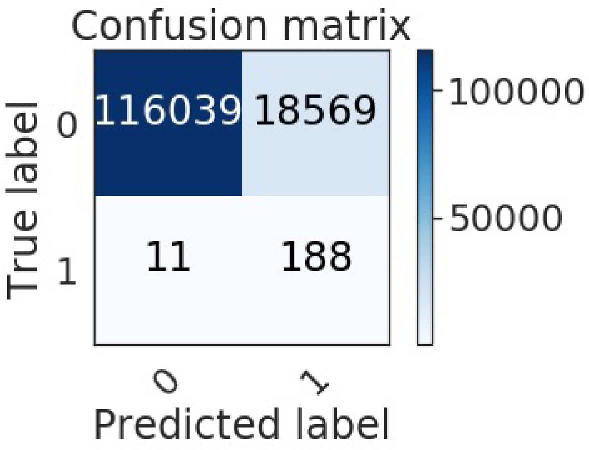

Figure 2

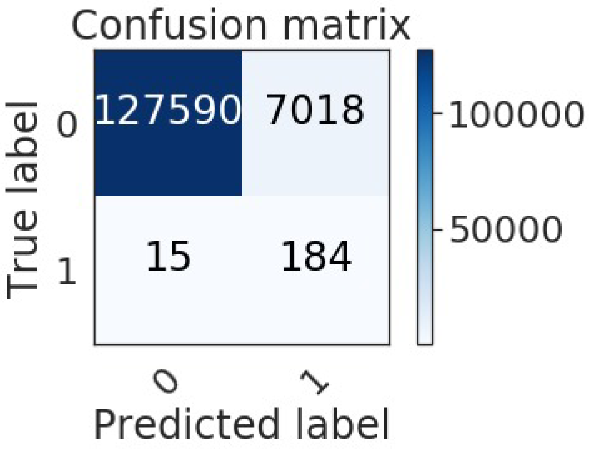

Figure 3

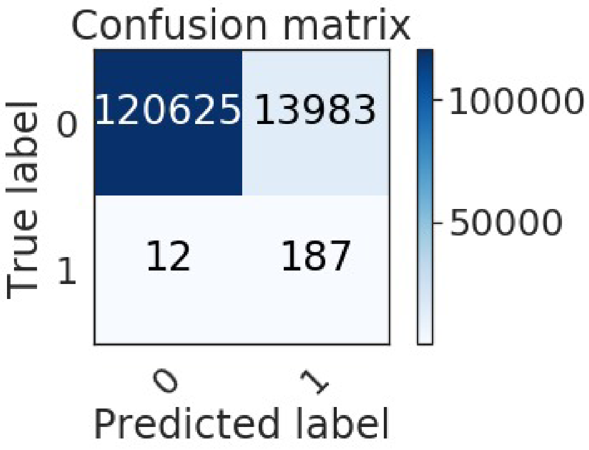

Figure 4

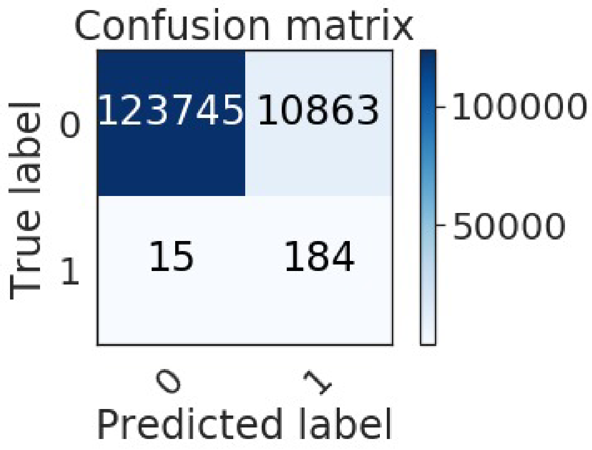

Figure 5

# 6 Conclusion

The classification system demonstrates the effectiveness of machine learning techniques in identifying patterns within structured data. By utilizing the Support Vector Machine (SVM) algorithm, the system efficiently processes input data, applies feature selection, and balances class distribution to improve classification accuracy. The evaluation results highlight the model's ability to distinguish between different categories with high precision, ensuring reliable predictions. Various performance metrics confirm the impact of preprocessing and optimization techniques in enhancing classification outcomes. The structured approach to data processing, model training, and evaluation ensures that the system maintains consistency and efficiency. Overall, the implementation of this classification system contributes to the development of more accurate and computationally efficient models, reinforcing the importance of machine learning in pattern recognition and decision-making applications.

# References

[1] M. Habibpour, H. Gharoun, M. Mehdipour, A. Tajally, H. Asgharnezhad, A. Shamsi, A. Khosravi, M. Shafie-Khah, S. Nahavandi, and J. P. S. Catalao, "Uncertainty-aware credit card fraud detection using deep learning," arXiv preprint arXiv:2107.13508, 2021.

[2] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, no. 1, pp. 145–174, Jan. 2023.

[3] T. K. Dang, T. C. Tran, L. M. Tuan, and M. V. Tiep, "Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems," *Appl. Sci.*, vol. 11, no. 21, p. 10004, Oct. 2021.

[4] J. Chaquet-Ulldemolins, F.-J. Gimeno-Blanes, S. Moral-Rubio, S. Muñoz Romero, and J.-L. Rojo-Álvarez, "On the black-box challenge for fraud detection using ma-

chine learning (I): Linear models and informative feature selection," *Appl. Sci.*, vol. 12, no. 7, p. 3328, Mar. 2022.

[5] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, "Credit card fraud detection using a new hybrid machine learning architecture," *Mathematics*, vol. 10, no. 9, p. 1480, Apr. 2022.

[6] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *J. Big Data*, vol. 8, no. 1, p. 151, Dec. 2021.

[7] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," *IEEE Access*, vol. 10, pp. 16400–16407, 2022.

[8] E. Btoush, X. Zhou, R. Gururaian, K. Chan, and X. Tao, "A survey on credit card fraud detection techniques in banking industry for cyber security," in *Proc. 8th Int. Conf. Behav. Social Comput. (BESC)*, Oct. 2021, pp. 1–7.

[9] I. D. Mienye and Y. Sun, "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data," *Inform. Med. Unlocked*, vol. 25, Jan. 2021, Art. no. 100690.

[10] S. A. Ebiaredoh-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye, "A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease," *Bioengineering*, vol. 9, no. 8, p. 350, Jul. 2022.

[11] V. K. S. Varun Kumar, V. G. Vijaya Kumar, A. Vijayshankar, and K. Pratibha, "Credit card fraud detection using machine learning algorithms," *Int. J. Eng. Res. Technol. (IJERT)*, vol. 9, no. 7, Jul. 2020.

[12] M. A. Sharma, B. R. Ganesh Raj, B. Ramamurthy, and R. H. Bhaskar, "Credit card fraud detection using deep learning based on auto-encoder," *ITM Web Conf.*, vol. 50, 2022.

**Cite this article:**

Dr Gayathiri A & Nehasri M, "Credit Card Fraud Detection Using Machine Learning", Journal of Multidimensional Research and Review (JMRR), Vol.6, Iss.2, pp.158-166, 2025