# CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

**D.Navya Sai Lakshmi,M.C.A Student , Amritha sai institute of science and technology, Kanchikacharla (Mandal), A.P- 521180**

**M.Anuradha,Assistant professor , Amritha sai institute of science and technology, Kanchikacharla (Mandal), A.P- 521180**

## ABSTRACT

The exponential growth of online transactions has led to a significant rise in credit card fraud, causing substantial financial losses to financial institutions and customers. Detecting fraudulent transactions is a challenging task due to the high volume of transaction data and the evolving nature of fraud patterns. This paper presents a machine learning-based approach for credit card fraud detection using the Random Forest algorithm. The proposed system analyzes historical transaction data and classifies transactions as genuine or fraudulent. Various preprocessing techniques such as data cleaning, normalization, and feature extraction are applied to enhance model performance. The system is evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that the Random Forest algorithm provides superior performance compared to traditional methods, achieving high accuracy and robustness in fraud detection.

## 1. INTRODUCTION

The advancement of digital payment systems has transformed financial transactions, making them faster and more convenient. However, this growth has also increased the risk of fraudulent activities. Credit card fraud is one of the most common financial crimes, where unauthorized users gain access to card details and perform illegal transactions.

Fraud detection systems must process large volumes of data in real time while maintaining high accuracy. Traditional rule-based systems are no longer sufficient due to their inability to adapt to new fraud patterns. Machine Learning (ML) techniques provide an intelligent solution by learning patterns from historical data and predicting fraudulent activities.

This work focuses on supervised learning using the Random Forest algorithm, which is known for its efficiency, accuracy, and resistance to overfitting. The model classifies transactions into two categories:

- Genuine Transaction (0)
- Fraudulent Transaction (1)

## 2. LITERATURE SURVEY

Numerous research works have been carried out in the domain of credit card fraud detection:

Sudhamathy G. proposed a decision tree-based classification model using R programming for credit risk analysis. The model involved preprocessing and achieved effective prediction accuracy.

Wei Sun et al. introduced Support Vector Machines (SVM) for credit risk assessment. SVM demonstrated strong classification capabilities and adaptability in financial datasets.

Amlan Kundu et al. proposed a hybrid model combining multiple machine learning algorithms such as Random Forest, Decision Trees, SVM, and Naïve Bayes. The study achieved accuracy levels above 90%, proving the effectiveness of ensemble techniques.

Y. Sahin and E. Duman compared Decision Trees and SVM for fraud detection and concluded that machine learning models significantly outperform traditional statistical approaches.

These studies highlight that ensemble methods and supervised learning algorithms provide better fraud detection performance.

## 3. EXISTING SYSTEM

The existing fraud detection system is based on unsupervised learning techniques such as clustering and Artificial Neural Networks (ANN).

**Working:**

- Data is normalized
- Cluster analysis is applied
- Neural networks classify transactions

**Limitations:**

- Low accuracy (~50%)
- High false positives
- Inefficient for large-scale data

- Lack of real-time detection
- Complex training process

These limitations necessitate the development of a more efficient and accurate system.

## 4. PROPOSED SYSTEM

The proposed system uses the Random Forest algorithm for classification. Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy.

**Architecture of Proposed System:**

1. Data Collection
2. Data Preprocessing
3. Feature Extraction
4. Model Training
5. Fraud Prediction
6. Performance Evaluation

**Advantages:**

- High accuracy and efficiency
- Handles large datasets
- Reduces overfitting
- Works well with imbalanced data

The system is designed to automatically learn from transaction patterns and detect anomalies effectively.

## 5. ALGORITHMS USED

### 5.1 Random Forest Algorithm

Random Forest is a supervised learning algorithm that constructs multiple decision trees during training and outputs the majority class.

**Working Principle:**

- Random sampling of data (Bootstrap Sampling)

- Construction of multiple decision trees
- Random selection of features at each split
- Aggregation using majority voting

**Mathematical Representation:**

Let:

- T1,T2,T3,...,TnT_1, T_2, T_3, ..., T_nT1,T2,T3,...,Tn be decision trees
- Final output = Majority Vote of all trees

**Advantages:**

- High accuracy
- Robust to noise
- Prevents overfitting
- Works for both classification and regression

## 6. METHODOLOGY

### 6.1 Data Collection

The dataset is obtained from Kaggle and contains anonymized credit card transaction records. It includes features such as transaction time, amount, and PCA-transformed variables.

### 6.2 Data Preprocessing

- Removal of missing values
- Data normalization
- Handling imbalanced data
- Feature scaling

### 6.3 Feature Extraction

Feature extraction transforms raw data into meaningful features. PCA (Principal Component Analysis) is used to reduce dimensionality and protect sensitive information.

### 6.4 Model Training

- Dataset split into training and testing sets
- Random Forest model is trained
- Hyperparameters are optimized

### 6.5 Model Evaluation

Two evaluation methods:

- Hold-Out Method
- Cross-Validation

## 7. RESULTS

The system is evaluated using the following metrics:

**Performance Metrics:**

- **Accuracy**
  Percentage of correctly classified transactions
- **Precision**
  Ratio of correctly predicted frauds to total predicted frauds
- **Recall** **(Sensitivity)**
  Ability to detect actual fraud cases
- **F1-Score**
  Harmonic mean of precision and recall

**Observations:**

- Accuracy achieved > 90%
- Significant improvement over existing system
- Reduced false positives
- Efficient fraud detection

Graphical representations such as confusion matrix and performance charts are used to visualize results.

## 8. CONCLUSION

This paper presented a machine learning-based approach for credit card fraud detection using the Random Forest algorithm. The proposed system effectively identifies fraudulent transactions with high accuracy and reliability. Compared to traditional systems, it provides better performance, scalability, and adaptability.

Future enhancements may include:

- Integration of deep learning models
- Real-time fraud detection systems
- Deployment using cloud technologies

## 9. REFERENCES

1. Sudhamathy G., "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R," Vol. 8, No. 5.
2. Wei Sun et al., "Credit Risk Assessment Using SVM," Vol. 6, 2006.
3. Amlan Kundu et al., "Credit Card Fraud Detection," IEEE, 2009.
4. Y. Sahin and E. Duman, "Fraud Detection using Decision Trees and SVM."
5. Kaggle, "Credit Card Fraud Detection Dataset."
6. T. Hastie, R. Tibshirani, "Statistical Learning Theory."