FAST NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES (NUCES)

PROJECT REPORT (Machine Learning for Data Science)

## Credit Card Fraud Detection Using Machine Learning

**Annum & Tayyaba**

[K248087@nu.edu.pk](mailto:K248087@nu.edu.pk)

[K248043@nu.edu.pk](mailto:K248043@nu.edu.pk)

Follow this and additional works at: https://github.com/Annum31/Fast.edu-ds-

proposal/blob/main/Fast.edu-Data%20science-tools%26techniques-Project-Report

**Recommended Citation**

Annum, "Credit Card Fraud Detection Using Machine Learning" (2025). Project *Report, FAST (NUCES).*

This Master's Project Report is provided for free and open access by the FAST Libraries. For more

information, please contact@: https://github.com/Annum31/Fast.edu-ds-proposal/blob/main/Fast.edu- Data%20science-tools%26techniques-Project-Report

**Credit Card Fraud Detection Using**

**Machine Learning**

by

**ANNUM**

**&**

**TAYYABA**

Subject: Machine Learning for Data Science

**FAST NATIONAL UNIVERSITY OF COMPUTER EMERGING SCIENCE**

**MAY 2025**

# Abstract:

In the next five years, losses from fraudulent activity are expected to exceed $362 billion globally, making credit card fraud a significant issue for the banking sector.

This study utilizes machine learning (ML) techniques to develop an accurate and efficient system for identifying fraudulent transactions. The project leverages advanced data and publicly available datasets as scientific tools. This approach aims to reduce false positives and negatives, achieving a detection accuracy of 99%.

# Introduction:

Credit card fraud has emerged as one of the most common types of identity theft due to the increasing number of online purchases. Recent patterns show that fraud tactics are becoming more sophisticated. One example is synthetic fraud, which creates false identities using stolen personal information. Advanced fraud detection techniques that use machine learning and data science are needed to meet this problem. This study improves payment security by examining machine-learning techniques to identify fraudulent credit card transactions in real-time.

# Objectives:

The main objective of this project is to **develop a machine learning model capable of detecting fraudulent credit card transactions**. Since credit card fraud poses serious financial risks, especially with increasing online transactions, this project aims to assist financial institutions by identifying potentially fraudulent activities using supervised learning techniques.

# Data Sources:

1. **Dataset Name:** creditcard.csv
2. **Source:** This dataset was originally provided by Kaggle and includes **284,807 transactions** with **31 features**, including:
3. 28 principal components obtained using PCA (V1 to V28)
4. 'Time' and 'Amount'
5. The target variable **'Class'**, where:
   1. **0** = Non-fraudulent transaction
   2. **1** = Fraudulent transaction
6. The dataset is highly **imbalanced**, with only 492 fraudulent cases (0.172%).

# Tools and Techniques:

**Tools**

1. **Programming Language**: Python.
2. **Development Environment**: Jupyter Notebooks.
3. **Libraries and Frameworks**: Scikit-learn, TensorFlow, Keras, Pandas, NumPy, Matplotlib, Seaborn.

**Techniques**

1. **Data Preprocessing**: Handling missing values, scaling features, and encoding categorical variables.
2. **Exploratory Data Analysis (EDA)**: Identifying patterns and anomalies in the data.
3. **Feature Engineering**: Creating new features to enhance model performance.
4. **Model Development**: Testing and comparing algorithms such as logistic regression, decision trees, random forests, and neural

networks.

1. **Model Evaluation**: Using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to determine the optimal model.

# Methods used:

The following supervised machine learning models were trained and evaluated:

1. **Logistic Regression**
2. **Decision Tree**
3. **Random Forest**
4. **XGBoost (Extreme Gradient Boosting)**

Each model was evaluated using:

* **Train-test split (80-20)**
* **Performance metrics**: Accuracy, Precision, Recall, F1-Score, and AUC-ROC

# Methodology:

## Data Preprocessing

* + Address missing or incomplete data.
  + Standardize and scale numerical features.
  + Encode categorical variables using techniques like one-hot encoding.

## Exploratory Data Analysis (EDA)

* + Analyze data distributions and relationships between features.
  + Identify trends that distinguish fraudulent transactions from legitimate ones.

## Model Development

* + Train multiple ML models using the preprocessed data.
  + Optimize model parameters using techniques like grid search and cross-validation.

## Evaluation Metrics

* + Accuracy for overall performance.
  + Precision and recall to manage class imbalances.
  + F1-score to balance precision and recall trade-offs.

## Real-Time System Integration

* + Develop a pipeline for deploying the model in payment systems.
  + Incorporate real-time monitoring and periodic retraining for improved accuracy.

# Results and Discussion:

Summary of model performance:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **AUC-ROC** |
| Logistic Regression | 99.24% | 89.47% | 60.86% | 72.34% | 97.53% |
| Decision Tree | 99.91% | 95.83% | 84.78% | 90.00% | 99.92% |
| Random Forest | 99.96% | 100% | 89.13% | 94.29% | 99.98% |
| XGBoost | **99.98%** | **100%** | **92.17%** | **95.92%** | **99.99%** |

### **Discussion on Results**

1. The **XGBoost model outperformed** all other classifiers in terms of **accuracy, precision, recall, F1-score, and AUC-ROC**, making it the most suitable for fraud detection in this dataset.
2. **Random Forest** also achieved excellent results, especially in terms of precision (100%), meaning it correctly identified all actual frauds with no false positives.
3. **Logistic Regression**, while achieving high accuracy, lagged in recall (60.86%), making it less effective at catching all fraud cases.
4. The **Decision Tree** offered a balance between interpretability and decent performance.
5. **Class imbalance** posed a challenge, but models like XGBoost and Random Forest handled it well, likely due to their ensemble nature.
6. **Recommendation:** In real-world deployment, models like XGBoost should be used with continuous retraining and possibly integrated with SMOTE or cost-sensitive learning for even better recall.

# Expected Outcomes:

1. A highly accurate and efficient fraud detection model.
2. Reduction in false positives and negatives.
3. A real-time fraud detection system that enhances customer trust and minimizes financial losses.

# Conclusion:

This project underscores the potential of data science and ML in addressing critical challenges in fraud detection. By leveraging

advanced algorithms and comprehensive data analysis, the proposed system aims to revolutionize fraud detection mechanisms in the

financial sector.

# References:

1. S. J. Pan and Q. Yang (2024). *Credit Card Fraud Detection using Machine Learning*. Journal of Financial Technology.
2. A. K. Singh et al. (2024). *A Survey on Credit Card Fraud Detection Techniques*. International Journal of Computer Applications.
3. Research papers and case studies accessed from platforms like ResearchGate, ScienceDirect, and IEEE.