**Project Report: Credit Card Fraud Detection**

**Muhammad Zafar Ul Haq**

**Title:** Credit Card Fraud Detection

**Introduction**

**Objective:**

The primary goal of this project is to develop a machine learning model capable of identifying fraudulent credit card transactions. This is critical due to the financial losses and security breaches associated with fraudulent activities.

**Dataset:**

The dataset utilized consists of credit card transactions made by European cardholders over two days in September 2013. It is highly imbalanced, containing 284,807 transactions, of which only 492 are fraudulent.

**Features:**

**Numerical Features:** Principal components obtained through PCA, excluding 'Time' and 'Amount'.

**Time:** Elapsed seconds since the first transaction.

**Amount:** Transaction value.

**Target Variable:** 'Class' (1 indicates fraud, 0 otherwise).

**Importance:**

Detecting fraudulent transactions is crucial for preventing financial losses and ensuring the security of online transactions, which are increasingly common.

**Related Work**

Previous research on fraud detection has explored various machine learning models, such as:

**Logistic Regression**

**Decision Trees**

**Support Vector Machines (SVM)**

**Challenges:**

Imbalanced Dataset: Few fraudulent transactions make model training challenging.

Feature Engineering: Essential for improving model performance.

**Working**

**Data Preprocessing:**

The data underwent cleaning to handle missing values and normalization. It was then split into training and testing sets for model development.

**Models Used:**

1. **Logistic Regression:** Simple and effective for classification tasks.

2**. K-Nearest Neighbors (KNN):** Evaluated and found effective (99% accuracy).

3. **Decision Tree:** Provides interpretability but prone to overfitting.

4. **Support Vector Machine (SVM):** Suitable for high-dimensional data.

**Evaluation Metrics:**

Given the class imbalance, accuracy is not a reliable metric. The primary evaluation metric used is the Area Under the Precision-Recall Curve (AUPRC). Other metrics considered include Precision, Recall, and F1 Score for comprehensive evaluation.

**Application**

**Real-world Implementation:**

The trained models can be integrated into financial systems for real-time fraud detection. They monitor transactions and flag suspicious activities, offering several advantages:

**Reduced False Positives:** Minimizes financial loss.

**Improved Accuracy:** Reduces unnecessary transaction blocks, enhancing customer experience.

**Increased Efficiency:** Prioritizes suspicious cases for fraud investigation teams.

**Conclusion**

**Summary:**

The project successfully implemented multiple machine learning models for credit card fraud detection, with KNN achieving notable performance (99% accuracy).

**Future Work:**

**Enhancing Model Performance:**

Collecting more transaction data for training.

Exploring advanced techniques like ensemble learning to combine multiple models for improved accuracy.