CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

1. Introduction
Credit card fraud is a major issue in the financial sector, causing significant losses to banks and customers every year. With the rapid growth of online and digital transactions, detecting fraudulent activities has become increasingly challenging. Machine Learning provides an effective solution by learning patterns from historical transaction data.

2. Objective
The objective of this project is to build a machine learning model that can accurately classify credit card transactions as fraudulent or legitimate using historical transaction data.

3. Dataset Description
The project uses two datasets: fraudTrain.csv and fraudTest.csv. The target column is **is_fraud**, where 0 represents legitimate transactions and 1 represents fraudulent transactions. The dataset is highly imbalanced.

4. Data Preprocessing
Irrelevant columns such as IDs, names, and timestamps were removed. Categorical variables were encoded using Label Encoding. Features and target variables were separated for training and testing.

5. Model Selection
Logistic Regression was chosen due to its simplicity, efficiency, and effectiveness for binary classification problems.

6. Model Training
The model was trained on the training dataset and learned patterns to differentiate fraudulent transactions from legitimate ones.

7. Model Evaluation
The model was evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Recall for fraud detection was prioritized due to the high cost of missed fraud cases.

8. Results
The model achieved high accuracy and successfully detected fraudulent transactions despite class imbalance.

9. Conclusion
Machine learning techniques can effectively detect fraudulent credit card transactions. Logistic Regression provides a strong baseline model for fraud detection tasks.

10. Future Improvements
Advanced models, imbalance handling techniques, and real-time deployment can further enhance performance.

11. Tools & Technologies
Python, Jupyter Notebook, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn.

12. Final Outcome
The project successfully meets the requirements of Task 2 of the CODSOFT Machine Learning Internship.