**Phase-1 Submission**

**Student Name:**Subash Chandra Bose M

**Register Number:** 712523205060

**Institution:** PPG Institute Of Technology

**Department:** B.tech IT

**Date of Submission:** 02-05-2025

**1. Problem Statement**

*In today’s digital economy, the rise in online transactions has been paralleled by a significant increase in credit card fraud.* *This project aims to leverage Artificial Intelligence (AI) and Machine Learning (ML) techniques to detect and prevent credit card fraud in real-time, ensuring the safety and integrity of financial transactions.*

**2. Objectives of the Project**

* *To build an AI-driven model that accurately identifies fraudulent credit card transactions.*
* *To analyze transaction patterns and identify features most indicative of fraud.*
* *To reduce false positives and maximize detection accuracy.*
* *To simulate a real-time fraud detection environment.*
* *To visualize fraud trends for business and security insights.*

**3. Scope of the Project**

* *Anomaly detection using transaction data.*
* *Model training and evaluation on real-world datasets.*
* *Dashboard for fraud visualization and alert generation.*

**4. Data Sources**

* ***Dataset:*** *Credit Card Fraud Detection Dataset*
* ***Source:*** *Kaggle - Credit Card Fraud Detection*
* ***Type:*** *Public*
* ***Nature:*** *Static dataset containing European credit card transactions from 2013, anonymized with PCA transformations.*
* ***Data source link:*** **GitHub:** <https://github.com/curiousily/Credit-Card-Fraud-Detection>

**5. High-Level Methodology**

* ***Data Collection****:  
  Download the dataset from Kaggle.*
* ***Data Cleaning****:  
  Handle missing values, remove duplicates, and normalize features if necessary.*
* ***Exploratory Data Analysis (EDA****):  
  Use visualizations like histograms, box plots, and correlation heat maps to analyze distributions and spot outliers.*
* ***Feature Engineering****:  
  Apply techniques like oversampling (SMOTE) to handle data imbalance. Create derived features such as transaction time intervals.*
* ***Model Building****:  
  Test various models including Logistic Regression, Decision Trees, Random Forest, XG Boost, and Neural Networks.*
* *Model Evaluation:  
  Use metrics like Precision, Recall, F1 Score, ROC-AUC to evaluate performance.*
* *Visualization & Interpretation:  
  Present results using confusion matrices, ROC curves, and interactive dashboards (e.g., Streamlet).*
* *Deployment:  
  Optional deployment as a Streamlet web app simulating a fraud detection interface.*

**6. Tools and Technologies**

* *Programming Language: Python*
* *Notebook/IDE: Jupyter Notebook, Google Colab*
* *Libraries: pandas, humpy, matplotlib, sea born, scikit-learn, imbalanced-learn, xgboost, tensorflow/keras*
* *Optional Deployment Tools: Streamlet, Flask*

**7. Team Members and Roles**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | *NAME* | *ROLE* |  | | *SUBASH HANDRA BOSE M* | *Team leader* | *Responsible for overall project planning data processing and model handling* | | *THANGAPANDI P* | *Data Analyst* | *Handles exploratory data analysis (EDA),feature engineering ,and insights*  *Generation* | | *SURESH KUMAR U* | Visualization Specialist | *Creates visualization and interprets model result using tools like SHAP and Plotly* | | *SRI SANKARI U* | Development Engineer | *Develops and manages development using tools like streamlet or flask and prepares*  *Final presentation* | | *SOWMIYA* | Documentation & reporting Lead | *Prepares project documentation,creates presentation, and compiles reports for final submission* | | | | | |  |
|  | | | | |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  | | | | |  |

|  |
| --- |
|  |

|  |
| --- |
|  |