**PHASE-1**

**Student Name** : Muhammed Hammad.S  
**Register Number** : 511323244017  
**Institution** : Kingston Engineering College  
**Department** : Computer Science and Business Systems  
**Date of Submission** : 25-04-2025

**Title**: Guarding Transactions with AI-powered credit card fraud detection and prevention

**Problem Statement:**

Credit card fraud is a silent thief — it strikes without warning, draining hard-earned money and shaking trust in digital systems. As online transactions rise, so does the creativity of fraudsters. Traditional systems struggle to keep up, relying on static rules and delayed response. Enter Artificial Intelligence — a guardian at the gate, trained to recognize patterns, detect anomalies, and stop fraud before it happens. This project leverages AI and machine learning to secure financial transactions by predicting and preventing credit card fraud with precision and speed.

**Objectives:**

* Detect fraudulent credit card transactions using AI-based models.
* Prevent fraud in near-real-time using predictive analytics.
* Reduce false positives and improve fraud detection accuracy.
* Analyze behavioral differences between fraudulent and genuine transactions.
* Design an interactive dashboard to visualize risk and transaction insights.

**Scope:**

**Features to Analyze:**

* Transaction time, amount, and frequency
* Geo-location and device information
* Merchant category
* Customer behavior trends and transaction history

**Constraints:**

* Focused only on credit card fraud detection.
* Uses publicly available datasets only.
* Real-time implementation is simulated.
* No direct integration with banking APIs or live data.

**Data Sources:**

* **Primary Dataset:** [**Credit Card Fraud Detection Dataset (Kaggle)**](https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud)
* **Contains anonymized transaction records labeled as fraudulent or genuine.**
* **Highly imbalanced dataset (~0.17% fraud cases), ideal for anomaly detection models.**
* **Additional Data:**
* **Synthetic Transaction Data: Generated using the Faker library to simulate real-world transaction flows and test model scalability.**
* **Geo-location & IP Data: Integrated from open-source APIs (like IPStack or GeoLite2) to identify geographic inconsistencies and location spoofing attempts.**

**High-Level Methodology:**

**1. Data Collection:**

* Downloaded dataset from Kaggle.
* Synthetic data generation for fraud/non-fraud balance if required.
* Real-time data simulation tools (like Streamlit forms or mock APIs) used to mimic live transaction inputs for testing the model pipeline.
* Public data from open financial APIs (e.g., Open Payments API or Plaid Sandbox) considered to simulate additional transaction types and metadata.

**2. Data Preprocessing:**

* Remove nulls, duplicates.
* Normalize and scale numerical data.
* Encode categorical features.
* Apply SMOTE or PCA for imbalance correction.

**3. Exploratory Data Analysis (EDA):**

* Identify fraud frequency across time and amount.
* Use plots to show correlation between variables.
* Analyze outliers and transaction clustering.

**4. Feature Engineering:**

* Add features like transaction velocity, night/day transactions, location mismatch.
* Behavior-based scoring of users.

**5. Model Building:**

* Algorithms: Logistic Regression, Random Forest, XGBoost, Neural Networks, Isolation Forest.
* Classification and anomaly detection for fraud prediction.

**6. Model Evaluation:**

* Metrics: Precision, Recall, F1-score, Accuracy, ROC-AUC.
* Prioritize Recall to catch fraud.
* Use k-fold cross-validation for reliability.

**7. Visualization and Interpretation:**

* Create dashboards showing:
* Fraud frequency graphs
* Location-based risk maps
* High-risk merchant or transaction patterns

**8. Deployment:**

* Use Streamlit for a clean, interactive web app dashboard.
* Backend (optional) built with Flask for API-based predictions.

**Tools and Technologies:**

* **Language**: Python
* **IDE**: Google Colab / Jupyter Notebook
* **Libraries**:
  + Data: pandas, numpy
  + ML Models: scikit-learn, xgboost, keras
  + Visualization: seaborn, matplotlib, plotly
  + Deployment: Streamlit, Flask

**Team Members and Roles:**

| **Name** | **Role** | **Responsibilities** |
| --- | --- | --- |
| **Muhammed Hammad.S** | Project Lead & Data Scientist | Leads the entire project pipeline — from data cleaning and model design to final deployment. |
| **Jayapriyan** | Data Analyst & Feature Engineer | Performs deep EDA, extracts meaningful insights, builds powerful features, and supports model optimization. |