**To develop a fraud detection system for credit card transactions using anomaly detection techniques**

**Problem Statement: -**

The goal of this project is to develop a fraud detection system for credit card transactions using anomaly detection techniques. Given the highly imbalanced nature of fraud in real-world datasets, where fraudulent transactions represent a very small percentage of the total data, this project aims to leverage machine learning methods to accurately identify anomalies that may represent fraudulent behaviour.

**Algorithm used: -**

**1. Isolation Forest:**

Best for: Unsupervised anomaly detection, especially on large datasets.

Why: It is fast, scalable, and efficient at detecting outliers.

**2. Autoencoders (Deep Learning):**

Best for: Complex data relationships and non-linear patterns.

Why: Autoencoders learn patterns in the data, and fraud (anomalies) tends to have a higher reconstruction error.

**3. XGBoost (Supervised Learning):**

Best for: When labelled data is available and you need high accuracy.

Why: It handles class imbalance well and is powerful for supervised learning tasks.

**4. Local Outlier Factor (LOF):**

Best for: Smaller datasets with clear local anomalies.

Why: It works well for local density deviations in fraud data.

**Datasets: -**

UCI Machine Learning Repository

Credit Card Data (default of credit card clients): This dataset contains 30,000 samples of credit card users with default payments as a form of fraud detection.

UCI Credit Card Dataset

**Expected Output: -**

**1. Anomaly Scores or Probability:**

* Each transaction is assigned an anomaly score or probability indicating the likelihood of it being fraudulent.
* Higher scores correspond to transactions more likely to be frauds.

**2. Classification (Fraud/Not Fraud):**

* The system classifies each transaction as either fraudulent or non-fraudulent based on a predefined threshold of the anomaly score.

**3. Confusion Matrix (Evaluation):**

* True Positives (TP): Fraudulent transactions correctly identified.
* False Positives (FP): Non-fraudulent transactions wrongly flagged as fraud.
* True Negatives (TN): Non-fraudulent transactions correctly identified.
* False Negatives (FN): Fraudulent transactions missed by the system.

**4. Precision, Recall, F1-Score:**

* Precision: Measures how many identified frauds are actually frauds.
* Recall: Measures how many actual frauds were correctly identified.
* F1-Score: Harmonic mean of precision and recall, indicating the balance between them.

**5. ROC Curve and AUC Score:**

* A Receiver Operating Characteristic (ROC) curve to visualize the trade-off between the true positive rate and false positive rate at different thresholds.
* AUC (Area Under Curve) score quantifying the performance of the model (the closer to 1, the better).

**6. Transaction Risk Alerts:**

* The system may trigger alerts when a transaction exceeds a certain risk threshold.

**Team members: -**

2320030234: - KRITHIK REDDY**(LEAD)**

2320030482: - BHARADWAJ

2320030XXX: - GALIB HUSSAIN