****

**Assessment Report**

on

**“Credit card Fraud detection”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Team No : 13

Section: C

Name : Harsh vardhan shukla , 202401100300116

Name : Divyam Singh , 202401100300104

Name : Deepak , 202401100300092

Name : Avinash Kumar , 202401100300079

**Under the supervision of**

“Mr Shivansh Sir”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

Credit card fraud is a serious financial crime that causes billions in losses worldwide. Traditional rule-based systems struggle to keep up with evolving fraud patterns. This project uses an unsupervised machine learning algorithm, **Isolation Forest**, to detect anomalies in credit card transactions and classify them as fraud or legitimate. The goal is to demonstrate how AI can help financial institutions detect fraudulent transactions proactively and efficiently.

**2. Problem Statement**

To identify fraudulent transactions in a highly imbalanced dataset using unsupervised anomaly detection. The challenge lies in accurately detecting rare fraud cases from the majority of legitimate transactions.

**3. Objectives**

* Load and explore a real-world credit card transaction dataset.
* Preprocess the data including scaling of features.
* Train an **Isolation Forest** model on normal transactions.
* Evaluate the model using classification metrics and visualizations.
* Plot ROC curve and confusion matrix for interpretability.

**4. Methodology**

**Data Collection:**  
The dataset was sourced from Kaggle, containing 284,807 transactions made by European cardholders in 2013, with only 492 fraud cases.

**Preprocessing Steps:**

* Amount and Time were scaled using **StandardScaler**.
* Original unscaled columns were removed.
* Class distribution was visualized to understand data imbalance.

**Model Building:**

* **Isolation Forest** was trained only on non-fraudulent transactions.
* The model learned the "normal" behavior and flagged outliers as fraud.

**Evaluation:**

* Predictions were compared to actual labels.
* Metrics like precision, recall, and AUC were calculated.
* Confusion matrix and ROC curve were plotted for better understanding.

**5. Data Preprocessing**

* StandardScaler was used to normalize Amount and Time features.
* Class was used to separate training data (normal only) and for evaluating on the entire set.
* PCA was applied to visualize fraud and non-fraud cases in two dimensions.
* Imbalanced nature of the data (only 0.17% fraud) was handled using the model’s contamination parameter.

**6. Model Implementation**

**Isolation Forest** was chosen because it is effective for anomaly detection in unsupervised settings. It isolates outliers using decision trees and is well-suited for highly imbalanced problems like fraud detection.

**7. Evaluation Metrics**

* **Precision**: How many detected frauds were actually fraud.
* **Recall**: How many actual frauds were correctly detected.
* **F1 Score**: Harmonic mean of precision and recall.
* **ROC AUC**: Measures ability of model to distinguish fraud vs non-fraud.
* **Confusion Matrix**: Visually represents prediction outcomes.

**8. Results and Analysis**

* The model achieved good **recall**, showing its ability to flag most frauds.
* ROC AUC score indicated decent separability between classes.
* Confusion matrix showed few false positives, meaning the model is reliable.
* PCA scatterplot helped visualize fraud clusters separated from normal transactions.

**9. Conclusion**

This project demonstrated that unsupervised learning with Isolation Forest can be effective in detecting credit card fraud, especially when labeled data is limited. While performance was satisfactory, further enhancements like using supervised methods (Random Forest, XGBoost) and applying oversampling (SMOTE) can improve detection. Visualization played a key role in understanding model behavior and result interpretation.

**10. References**

* Kaggle: Credit Card Fraud Detection Dataset
* scikit-learn documentation
* seaborn & matplotlib visualization libraries
* Research papers on fraud detection and anomaly detection algorithms







