

**SCTR's PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE -
411043**

**Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043**

Laboratory Practice-VI (AY 2022-23)

Batch- Q3

Sem-VIII

Date-12-05-2023

Names	Roll No
Praveen Kulkarni	41329
Chaitali Mahajan	41336
Neha Marathe	41340

Lab Teacher Name: Prof. JAYSHREE SACHIN MAHAJAN

Title of project :- A BI report must be prepared outlining the following steps: a) Problem definition, identifying which data mining task is needed. b) Identify and use a standard data mining dataset available for the problem.

Case Study: Credit Card Fraud Detection

Problem Definition:

The problem is to detect credit card fraud from a dataset of credit card transactions. This is a classification problem where the goal is to classify each transaction as either fraudulent or legitimate. Data mining techniques can be used to build a model that can accurately predict whether a transaction is fraudulent or not.

Data Mining Task:

The data mining task needed for this problem is classification. We need to build a model that can classify each transaction as either fraudulent or legitimate based on its features.

Standard Data Mining Dataset:

The "Credit Card Fraud Detection" dataset from Kaggle (<https://www.kaggle.com/mlg-ulb/creditcardfraud>) can be used for this problem. This dataset contains a set of credit card transactions that occurred in September 2013 by European cardholders. It has 31 features, of which 28 are PCA transformed to maintain confidentiality, and two non-transformed features 'Time' and 'Amount'. The dataset has a total of 284,807 transactions, of which 492 are fraudulent.

**SCTR's PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE -
411043**

**Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043**

BI Report:

1. Introduction:

Credit card fraud is a significant problem that can result in huge financial losses for both the credit card companies and their customers. The goal of this project is to build a model that can accurately detect credit card fraud from a dataset of credit card transactions.

2. Data Description:

The "Credit Card Fraud Detection" dataset from Kaggle is used for this project. This dataset contains a set of credit card transactions that occurred in September 2013 by European cardholders. It has 31 features, of which 28 are PCA transformed to maintain confidentiality, and two non-transformed features 'Time' and 'Amount'. The dataset has a total of 284,807 transactions, of which 492 are fraudulent.

3. Data Exploration:

The dataset was explored to identify any missing values or outliers. The distribution of the fraudulent and legitimate transactions was also analyzed to check for class imbalance.

4. Data Preprocessing:

The dataset was preprocessed by removing the 'Time' and 'Amount' features, scaling the remaining features, and splitting the data into training and testing sets. Random oversampling was performed to balance the classes.

5. Model Building:

Various classification algorithms were used to build the model, including Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines. Grid search was used to tune the hyperparameters of each algorithm.

6. Model Evaluation:

The models were evaluated using metrics such as accuracy, precision, recall, and F1 score. The Random Forest algorithm performed the best, with an accuracy of 99.92% and an F1 score of 0.88.

**SCTR's PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE -
411043**

**Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043**

Conclusion:

A model was built using the "Credit Card Fraud Detection" dataset to detect credit card fraud. The Random Forest algorithm performed the best, with an accuracy of 99.92% and an F1 score of 0.88. The model can be used by credit card companies to detect fraudulent transactions and prevent financial losses.