

***­­***

***Machine Learning Project***

***Credit Card Fraud Detection***

# Team 17

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Code | Sec | BN |
| Abdelrahman Hamdy Ahmed | 9202833 | 1 | 38 |
| Ziad Sherif | 9202586 | 1 | 26 |
| Zeyad Tarek | 9202588 | 1 | 27 |
| Abdelhameed Emad | 9202758 | 1 | 34 |

Date: 4/08/2024 Presented to: Eng. Mohamed Shawky

|  |  |  |  |
| --- | --- | --- | --- |
| **Workload Division** | | | |
| **Abdelrahman** | **Zeyad Tarek** | **Ziad Sherif** | **Abdelhameed** |
| * **Data Preprocessing** * Data Splitting * Scaling * Removing outliers * Data Distribution & Correlation * **Data Visualization** * **Feature Engineering** * Oversampling * Undersampling * Dimensionality Reduction * **Grid Search** | | * **Model Selection & Training** * Logistic Regression * KNN * SVM * Decision Tree * Random Forest * **Cross-validation** * **Performance Analysis** * ROC Curve * Confusion Matrix * Accuracy Measures | |

**Project Outline**

**1. Problem Description:**

Problem: Detecting credit card fraud is crucial for banks to protect customers from unauthorized charges. Using machine learning, banks analyze transaction data to flag suspicious activities based on patterns and anomalies, ensuring security against evolving fraud tactics. Regular updates and improvements in detection methods are essential to stay ahead of fraudsters.

Problem Definition and Motivation: The goal is to develop a **machine learning model** capable of accurately **detecting fraudulent credit card transactions** based on transaction data. This is motivated by the need to prevent financial losses for both customers and credit card companies, as well as to maintain trust in the financial system. Credit card fraud detection is crucial for protecting customers and minimizing financial losses. Machine learning models, including supervised algorithms like logistic regression and random forests, as well as unsupervised techniques like anomaly detection, are employed to identify fraudulent transactions. Evaluation metrics such as precision, recall, and AUC-ROC ensure the effectiveness of these models. Once trained, the selected model can be deployed to monitor real-time transactions, aiding financial institutions in promptly detecting and mitigating fraudulent activity.

**2. Evaluation Metrics:**

* **ROC (Receiver Operating Characteristic)**: ROC curves visualize the performance of a classification model across various threshold settings. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR). A higher area under the ROC curve indicates better model performance, with an AUC (Area under Curve) of 1 representing a perfect classifier.
* **Accuracy**: Accuracy is the proportion of correctly classified instances out of the total instances. It's a simple and intuitive metric, but it can be misleading in the presence of class imbalance. Accuracy doesn't account for the types of errors (false positives and false negatives) made by the classifier.
* **Recall (Sensitivity)**: Recall measures the ability of a classifier to find all relevant instances in the dataset. It is calculated as the ratio of True Positives to the sum of True Positives and False Negatives. High recall indicates that the classifier is good at identifying positive instances, but it may also classify some negative instances as positive.
* **Precision**: Precision measures the accuracy of positive predictions made by the classifier. It is calculated as the ratio of True Positives to the sum of True Positives and False Positives. High precision indicates that the classifier is conservative in labeling instances as positive, but it may miss some positive instances.

**3. Dataset and References:**

* **Dataset:** The dataset contains transactions made by credit cards in September 2013 by European cardholders. It includes 492 frauds out of 284,807 transactions, with highly unbalanced classes.
* **References:**
  + The dataset is sourced from Kaggle: Credit Card Fraud Detection
  + Additionally, a simulator for transaction data has been released as part of the practical handbook on Machine Learning for Credit Card Fraud Detection (Fraud Detection Handbook).

**Results & Conclusion**