Project Overview: Credit Fraud Detection with Imbalanced Datasets

Objective: The project focuses on detecting fraudulent credit card transactions within a heavily imbalanced dataset. It employs resampling techniques and machine learning models to improve the detection accuracy of fraudulent cases, addressing the imbalance challenge effectively.

Data:

- Description: A credit card transactions dataset with anonymized features, including transaction time and amount, where the majority of transactions are legitimate, and a small fraction are fraudulent.

# Key Project Components

1. Data Analysis and Preprocessing:

- Explored data distribution to identify skewed patterns.

- Scaling and PCA Transformation: Prepared the features using scaling and PCA to handle dimensionality while preserving important patterns.

2. Resampling Techniques for Imbalance:

- Random UnderSampling and SMOTE: Applied NearMiss (for undersampling) and SMOTE (for oversampling) techniques to create a more balanced dataset.

- Implemented anomaly detection methods to aid in identifying fraudulent transactions.

3. Modeling:

- Trained various classifiers, including Logistic Regression, Decision Trees, Random Forest, and Neural Networks, evaluating each for their effectiveness in handling imbalanced data.

- Tuned models with hyperparameter optimization, using class weighting and cross-validation.

4. Evaluation Metrics:

- Focused on suitable metrics like F1-Score, Precision, Recall, ROC-AUC, and Confusion Matrix to accurately measure performance, avoiding misleading accuracy metrics.

# Accessing the Project on GitHub

1. Clone the Repository:

```bash

git clone https://github.com/yourusername/credit-fraud-detection.git

```

2. Navigate to the Project Directory:

```bash

cd credit-fraud-detection

3. Run the Notebook:

Open the notebook file (`credit-fraud-dealing-with-imbalanced-datasets.ipynb`) in Jupyter Notebook or Google Colab to review the entire workflow, from data preprocessing through model evaluation.

---

This overview provides a clear breakdown of each step, highlighting the project’s methodology and approach to handling data imbalance. Let me know if you need any adjustments or further details!