Credit Card Fraud Detection Project Documentation

Objective

The objective of this project is to build an effective credit card fraud detection system using machine learning techniques. We'll analyze a dataset containing historical credit card transactions labeled as either fraudulent or legitimate.

Steps

1. Data Preparation

- **Load the Dataset:**
  - Load the credit card dataset (`creditcard.csv`).
  - Explore the dataset by checking the first few rows using `df.head()`.
  - Examine the class distribution using `df.Class.value_counts()`.

2. Data Splitting

- **Prepare Training and Testing Data:**
  - Separate features (independent variables) and the target variable.
  - Split the data into training and testing sets using `train_test_split`.

3. Model Building

- **Logistic Regression:**
  - Perform hyperparameter tuning using `GridSearchCV`:
    - Explore different regularization strengths (`C`) and penalties (`l1` or `l2`).
  - Evaluate the model using metrics like accuracy, precision, recall, and F1-score.
- **Random Forest Classifier:**
  - Train a Random Forest classifier:
    - Consider class weights to handle class imbalance.
  - Evaluate the model performance.

4. Handling Class Imbalance

- **Under-Sampling (NearMiss):**
  - Reduce the majority class instances to balance the dataset.
  - Evaluate the model performance after under-sampling.
- **Over-Sampling (RandomOverSampler):**
  - Increase the minority class instances to balance the dataset.
  - Evaluate the model performance after over-sampling.

- **Combined Over-Sampling and Under-Sampling (SMOTETomek):**
    - Use a combination of synthetic minority oversampling and under-sampling.
    - Evaluate the model performance.

Conclusion

In this project, we explored various techniques for credit card fraud detection. We trained and evaluated models using logistic regression and random forest classifiers. Additionally, we addressed class imbalance using under-sampling and over-sampling methods.