**Credit Card Fraud Detection Using Machine Learning**

**Objective:**

To detect fraudulent credit card transactions using a machine learning-based classification approach. The goal was to develop and evaluate models that can identify fraud from normal transactions with high accuracy, precision, and recall.

**Project Overview:**

This project focused on building machine learning models for detecting fraudulent credit card transactions. The dataset used contained a high class imbalance between fraud (minority) and normal transactions (majority). You employed various techniques to preprocess the data, handle imbalance, and build multiple models to compare their performance.

**Key Contributions:**

* **Data Preprocessing:**
  + **Handled duplicates:** Detected and removed 1,081 duplicate records from the dataset, reducing it to 283,726 rows.
  + **Standardization:** Applied StandardScaler to normalize features such as Time and Amount for uniformity, aiding in model convergence.
  + **Feature Scaling:** Ensured all features were on the same scale by standardizing numerical values.
* **Class Imbalance Handling:**
  + **Undersampling:** Addressed the imbalance by undersampling the majority class (normal transactions) to match the number of fraud cases.
  + **Oversampling using SMOTE:** Employed Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples for the minority class (fraud transactions), ensuring the model was trained with an equal number of fraud and normal transactions.
* **Exploratory Data Analysis (EDA):**
  + Performed a correlation analysis and visualized the relationships between features using a heatmap.
  + Visualized the distribution of fraud and normal transactions using a count plot to highlight the class imbalance.
* **Model Development:**
  + Trained and compared three machine learning models: Logistic Regression, Decision Tree Classifier, and Random Forest Classifier.
  + Evaluated the models using metrics like **Accuracy Score**, **ROC-AUC Score**, **Confusion Matrix**, and **Classification Report**.
* **Results on Imbalanced Data:**
  + **Logistic Regression:** Achieved an accuracy of 99.91% and an ROC-AUC score of 0.776.
  + **Decision Tree Classifier:** Achieved an accuracy of 99.91% and an ROC-AUC score of 0.879.
  + **Random Forest Classifier:** Achieved an accuracy of 99.95% and an ROC-AUC score of 0.866.
* **Results after Addressing Class Imbalance (SMOTE):**
  + **Logistic Regression:** Achieved an accuracy of 94.51%, and a well-balanced **ROC-AUC score** of 0.945.
  + **Decision Tree Classifier:** Achieved an accuracy of 99.81% with an ROC-AUC score of 0.998.
  + **Random Forest Classifier:** Achieved an outstanding accuracy of 99.99% and an ROC-AUC score of 0.999.
* **Model Deployment:**
  + **Model Saving and Loading:** Successfully saved the best performing model using joblib for future predictions, ensuring reusability.
  + Implemented a **Random Forest Classifier** model for real-time fraud detection based on trained features, with a prediction function to classify new transactions.

**Exceptional Achievements:**

* Demonstrated the ability to handle highly imbalanced data effectively using both under sampling and oversampling techniques.
* Successfully optimized and evaluated multiple models, achieving near-perfect precision and recall in detecting fraudulent transactions after oversampling.
* Incorporated a thorough approach to model evaluation using detailed metrics, ensuring that the model performance was not just high in accuracy but also in recall and precision for the minority class (fraud).
* Automated the model deployment process, enabling real-time fraud detection using the trained model.