🚀 **Project Overview** The main objective of this project is to detect fraudulent transactions in a highly imbalanced credit card dataset, where only 0.17% of the transactions are fraudulent.

📁 Dataset

- The dataset was obtained from Kaggle and contains anonymized transaction details.
- **Features:** 30 numerical features resulting from a PCA transformation for privacy.
- **Target Variable:** Class (0 for legitimate, 1 for fraudulent).

🛠️ Technologies Used

- Python
- Pandas for data manipulation
- Matplotlib & Seaborn for data visualization
- Scikit-learn for machine learning algorithms and model evaluation

🔧 Methodology

Data Preprocessing

- Checked for missing values and data inconsistencies.
- Performed exploratory data analysis (EDA) to understand the dataset.

Model Building

- Applied a Random Forest Classifier to build the initial model.

Model Evaluation

- Evaluated using metrics like Accuracy, Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC).
- Visualized the confusion matrix to understand misclassifications.

📊 Evaluation Metrics

Metric	Score
Accuracy	99.95%
Precision	98.66%
Recall	75.51%
F1-Score	85.55%
Matthews Correlation Coefficient	86.29%

🔍 Key Findings

- The model achieved high accuracy and precision, indicating that most non-fraudulent transactions were correctly identified.
- The recall score was lower, showing that some fraudulent transactions were missed—a common challenge with imbalanced datasets.
- The confusion matrix visualized misclassifications, confirming the impact of data imbalance.

🚀 Next Steps

- Apply Data Balancing Techniques such as SMOTE or undersampling to improve recall.
- Experiment with advanced models like XGBoost or LightGBM.
- Fine-tune hyperparameters for better optimization.
- Implement real-time detection techniques for faster fraud recognition.

📈 Results

- The model correctly identified most legitimate transactions.
- Some fraudulent transactions were missed, highlighting the challenge of imbalanced datasets.
- Visualization of the confusion matrix provided deeper insight into model performance.

✅ **Conclusion** This project provided valuable learning on handling imbalanced datasets and evaluating machine learning models. While the model performed well initially, future improvements such as data balancing and model tuning could enhance recall and overall robustness.