**📝 Task 2: Credit Card Fraud Detection – Final Report**

**📌 Objective**

Build a machine learning model to detect fraudulent credit card transactions using real-world anonymized data. The goal is to classify each transaction as either **legitimate (0)** or **fraudulent (1)**.

**📁 Dataset**

* **Source**: [Kaggle - Credit Card Fraud Detection](https://www.kaggle.com/datasets/kartik2112/fraud-detection)
* **Rows**: 284,807 transactions
* **Features**: 30 anonymized variables (V1 to V28, Time, Amount)
* **Target**: Class (0 = Normal, 1 = Fraud)
* **Note**: Highly imbalanced dataset (only ~0.17% are frauds)

**🛠️ Preprocessing Steps**

1. Checked for and removed **missing values**.
2. Separated features (X) and labels (y).
3. **Standardized** the feature values using StandardScaler.
4. Split dataset into **training (80%)** and **testing (20%)** sets using stratification.

**🤖 Model Used**

* **Logistic Regression**
  + Fast and interpretable baseline model
  + Well-suited for binary classification tasks

**📊 Model Evaluation**

**✅ Confusion Matrix**

| **Actual \ Predicted** | **0 (Normal)** | **1 (Fraud)** |
| --- | --- | --- |
| 0 (Normal) | 56851 | 13 |
| 1 (Fraud) | 36 | 62 |

**✅ Classification Report**

* **Accuracy**: 99.91%
* **Precision (Fraud)**: 83%
* **Recall (Fraud)**: 63%
* **F1-Score (Fraud)**: 72%

**📈 Key Insights**

* The model is **highly accurate** in identifying normal transactions.
* It detected **63% of fraud cases**, which is decent for an imbalanced dataset.
* Performance can be further improved using:
  + **Oversampling (SMOTE)**
  + **Ensemble models** (Random Forest, XGBoost)

**🔗 Repository**

[GitHub Repo Link – <https://github.com/thiruvarul11/CODSOFT.git> ]

**✅ Conclusion**

Successfully implemented a machine learning pipeline to detect fraud in credit card transactions using Logistic Regression. The model provides a solid baseline and can be improved with further optimization techniques.