✅ RD INFRO TECHNOLOGY Internship Report

Intern Name: Giridharan D
Project Title: Anomaly Detection Using Machine Learning
Task: Task 5 – Build a Machine Learning Model for Anomaly Detection

Tools Used:

- Python
- Scikit-learn
- Pandas
- Google Colab
- GitHub (for project hosting)

🔍 Objective

The main goal of this project is to build a machine learning model to detect anomalies or fraudulent transactions using a real-world dataset (creditcard.csv). This is a classic use case of **unsupervised learning**, ideal for identifying rare and unusual patterns.

📂 Dataset Description

- **Name**: creditcard.csv
- **Source**: Publicly available dataset for fraud detection
- **Features**: 30 columns (numerical and scaled), with a Class column indicating:
  - 0 = Normal Transaction
  - 1 = Fraudulent Transaction

🦴 Steps Performed

1. **Import Libraries**
   Loaded essential libraries like pandas, numpy, matplotlib, sklearn.
2. **Data Cleaning**
   - Checked for and removed any rows with missing values (NaN) in the Class column.
3. **Feature Selection**
   - Removed the Class column temporarily to separate features (X) and labels (y).
4. **Model Used**
   - **Isolation Forest**, an unsupervised algorithm suitable for anomaly detection.
   - Tuned with: n_estimators=100, contamination=0.001.
5. **Training and Prediction**
   - Trained the Isolation Forest on the dataset.
   - Predictions were mapped as:
     - 1 → Normal (0)
     - -1 → Anomaly (1)
6. **Evaluation Metrics**
   - **Confusion Matrix**
   - **Classification Report** with:
     - Precision
     - Recall
     - F1-Score

📊 Sample Output

lua

CopyEdit

Confusion Matrix:

[[282569 31]

[192 351]]

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	282600
1	0.92	0.65	0.76	543
accuracy			1.00	283143
macro avg	0.96	0.83	0.88	283143
weighted avg	1.00	1.00	1.00	283143

📌 Conclusion

- Successfully built a working anomaly detection system using Isolation Forest.
- Achieved good accuracy and recall, especially important in fraud detection scenarios.
- Demonstrated hands-on understanding of handling real-world, imbalanced data.

🔗 GitHub Repository Link:

https://github.com/GIRIDHARAN-D46/RD-INFRO-TECHNOLOGY.git