Fraud Detection System using Machine Learning

Data Analytics Internship

Anisha More

Problem Statement

Detect fraudulent transactions

Handle imbalanced dataset

Apply classification algorithms

Evaluate detection performance

Dataset Overview

284,807

Total Records

transactions

492

Fraud Cases

30

Features

numerical attributes

Highly imbalanced dataset

Data Challenges

- Class imbalance problem
- Fraud cases < 0.2%
- Need for anomaly detection
- Risk of overfitting

Data Preprocessing

Scaled numerical features

Handled imbalance (undersampling)

Train-Test Split (80:20)

Feature selection applied

Models Used

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier

Model Performance

Logistic Regression Accuracy

Decision Tree Accuracy

Random Forest Accuracy

Precision & Recall evaluated

Evaluation Metrics

Confusion Matrix

Precision Score

Recall Score

F1-Score

Key Insights

- Random Forest most reliable
- High recall reduces fraud loss
- Class imbalance critical factor
- Feature scaling improved results

Conclusion

- Built fraud detection system
- Handled imbalanced data
- Compared multiple models
- Achieved high detection accuracy