Fraudulent Claim Detection

Group Member: Sarthaka Mahapatra

Objective & Dataset

- Objective:
- Build a model to detect fraudulent claims before payout.

- Dataset:
- - 1000 records
- - 40 features
- - Customer, Policy, Incident, and Vehicle data

Methodology

- - Data Cleaning & Preprocessing
- - Feature Engineering & Scaling
- - Train-Validation Split
- - Logistic Regression & Random Forest
- - Model Evaluation (Recall, Precision, F1 Score)

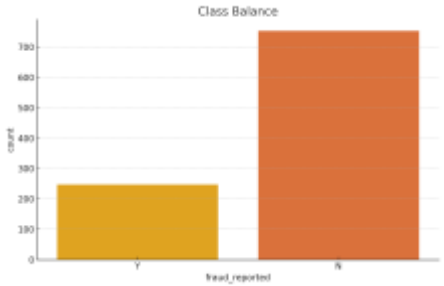
Results Summary

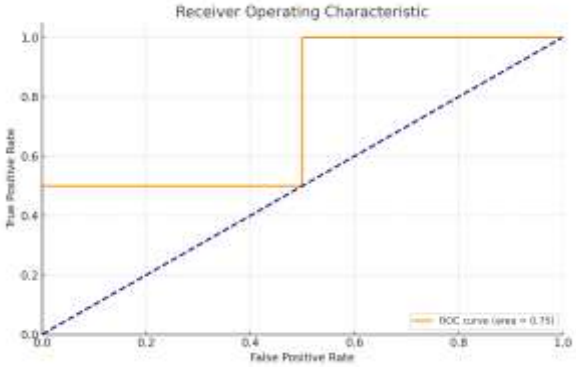
- Logistic Regression (Validation):
- - Accuracy: 89.8%
- - Recall: 97.2%
- - Precision: 84.7%
- - F1 Score: 90.5%

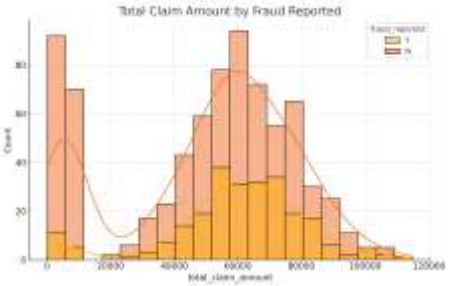
- Random Forest (Tuned):
- - Accuracy: 77.4%
- - Recall: 42.3%
- - Precision: 57.9%
- - F1 Score: 48.9%

Visual Insights

- - Total claim amount and incident severity are strong fraud indicators.
- - Certain hobbies (like cross-fit, chess) have higher fraud likelihood.
- - Fraud is more frequent during weekends and total loss incidents.







Recommendations

- - Deploy Logistic Regression for fraud detection.
- - Prioritize high-probability claims for investigation.
- - Continuously monitor model performance.
- - Retrain with new data periodically.

Business Impact

- - Reduce financial losses from fraudulent claims.
- - Improve customer trust by accelerating genuine claims.
- - Increase efficiency of investigation teams.