Fraud Detection Project Documentation

1. Model Performance Summary

Supervised Models Performance

Logistic Regression

Precision: 1.0

Recall: 0.95

F1-Score: 0.97

XGBoost Classifier

Precision : 1.00

Recall : 0.97

F1-Score : 0.99

Isolation Forest

Anomaly Detection Rate: 1%

Detected Anomalies: 12967

Approach:

- Identifies transactions significantly different from normal patterns
- Uses ensemble of decision trees to isolate anomalies

2. Visualizations

Confusion Matrix

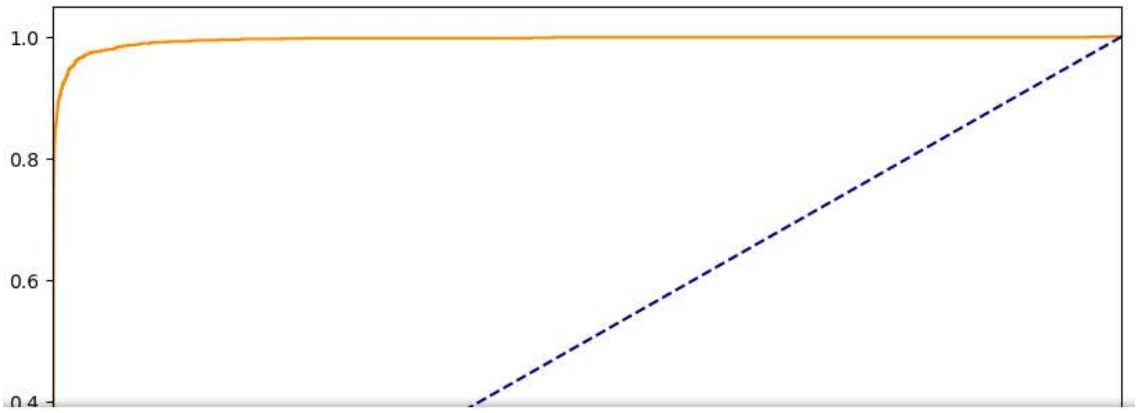
Logistic Regression Results:

	precision	recall	f1-score	support
0	1.00	0.95	0.97	257834
1	0.08	0.76	0.15	1501
accuracy			0.95	259335
macro avg	0.54	0.86	0.56	259335
weighted avg	0.99	0.95	0.97	259335

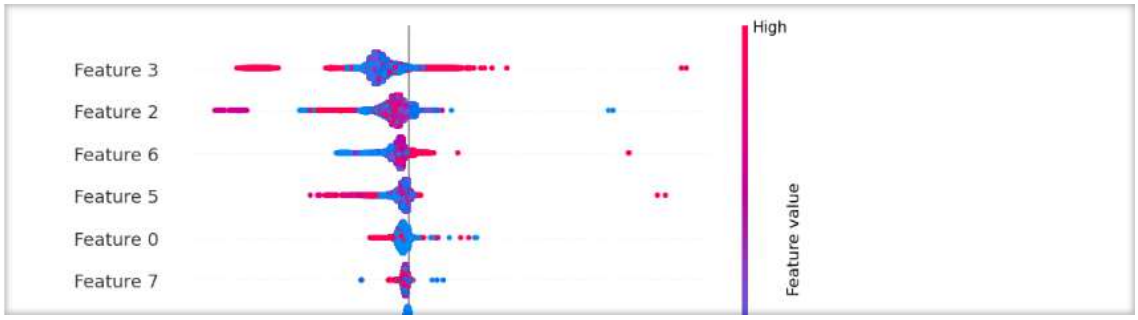
XGBoost Results:

	precision	recall	f1-score	support
0	1.00	0.97	0.99	257834
1	0.16	0.97	0.28	1501
accuracy			0.97	259335
macro avg	0.58	0.97	0.63	259335
weighted avg	0.99	0.97	0.98	259335

ROC Curve



SHAP Feature Importance



3. Anomaly Detection Insights

Anomaly Characteristics

- Unusual transaction amounts
- Geographical inconsistencies
- Unexpected merchant categories
- Temporal pattern deviations

Example Anomalous Transactions

1. High-value transaction in unusual location
2. Rapid successive transactions
3. Transactions outside typical customer behavior

4. Feature Engineering Highlights

Engineered Features

- Transaction time of day
- Geographical distance
- Customer age
- Transaction frequency indicators

Feature Impact

- Spatial features critical in detecting unusual transactions
- Temporal features help identify suspicious activity patterns

5. Recommendations

Model Improvements

- Incorporate more external data sources
- Implement ensemble methods
- Develop real-time fraud scoring system

Deployment Considerations

- Regular model retraining
- Continuous monitoring of performance
- Adaptive threshold management

6. Limitations

- Relies on historical data patterns
- May miss novel fraud techniques

- Performance depends on data quality and representativeness