## Fastag fraud detections.

Evaluation metrics and analysis report.

Outline

[Fastag fraud detections. 1](#_Toc169256187)

**2...……………………………………………………………………………………………………………………….** Outline.

3……………………………………………………………………………………………………………Dataset Overview.

3……………………………………………………………………………………………………………Data Preprocessing.

3……………………………………………………………………………………………………………EDA.

3……………………………………………………………………………………………………………Model Development.

4……………………………………………………………………………………………………………Handling imbalance.

4……………………………………………………………………………………………………………Model Evaluation.

4……………………………………………………………………………………………………………Recommendations.

5……………………………………………………………………………………………………………Recommendations.

5……………………………………………………………………………………………………………Conclusion.

Data overview:

**Dataset Size**: 5000 transactions.

**Features**:

Transaction\_ID, Timestamp, Vehicle\_Type, FastagID, TollBoothID, Lane\_Type, Vehicle\_Dimensions, Transaction\_Amount, Amount\_paid, Geographical\_Location, Vehicle\_Speed, Vehicle\_Plate\_Number.

**Target Variable**: Fraud\_indicator (Binary: Fraud or Not Fraud).

Data Preprocessing:

**Handling Missing Values:**

**FastagID**: 549 missing values (imputed with 'Unknown').

**Feature Engineering:**

 Converted Timestamp to datetime format.

 Label encoded categorical features.

 Scaled numerical features using MinMaxScaler.

Exploratory Data Analysis (EDA):

**Visualization**:

 Distribution of Fraud Indicator: Majority class (Not Fraud) dominates the dataset.

 Transaction Amount Distribution: Skewed distribution with most transactions having lower amounts.

 Vehicle Speed Distribution: Normally distributed with a mean around 67 km/h.

 Transaction Amount by Fraud Indicator: Higher transaction amounts slightly associated with fraud.

Model Development:

**Model Used**: Logistic Regression.

**Model Performance**:

**1.Accuracy: 99%**

**2.Precision:**

 Fraud: 98%

 Not Fraud: 100%

3.Recall:

 Fraud: 100%

 Not Fraud: 93%

4.F1-Score:

 Fraud: 99%

 Not Fraud: 97%

Handling Imbalance:

**SMOTE Oversampling**:

* Applied to balance the classes in training data.
* Improved model performance on minority class (fraud detection).

­­­ Model Evaluation:

**Evaluation Metrics**:

* Confusion Matrix Analysis:
  + True Positive (TP): Correctly predicted fraud transactions.
  + True Negative (TN): Correctly predicted non-fraud transactions.
  + False Positive (FP): Incorrectly predicted as fraud (Type I error).
  + False Negative (FN): Incorrectly predicted as non-fraud (Type II error).

**Analysis**:

* **Precision-Recall Trade-off**: Balanced precision and recall, crucial for fraud detection to minimize false positives and false negatives.
* **Model Robustness**: Handled class imbalance effectively using SMOTE, enhancing detection capabilities.
* **Model Limitations**: Convergence issues noted with the logistic regression solver, suggesting potential improvements in model tuning or alternative models.

Recommendations:

 **Further Investigation**:

* Explore ensemble methods (e.g., Random Forest, Gradient Boosting) for potentially higher accuracy and robustness.
* Investigate feature importance for better insights into fraud detection patterns.

 **Deployment Considerations**:

* Implement real-time data ingestion and prediction for Fastag fraud detection system.
* Monitor model performance metrics post-deployment for continuous improvement.

Conclusion:

 **Summary**: Developed and evaluated a logistic regression model for Fastag fraud detection, achieving high accuracy and balanced precision-recall trade-off.

 **Future Directions**: Continual refinement of models and integration with advanced analytics for enhanced fraud detection capabilities.