FAST TAG FRAUD DETECTION

Presented by: Catharin Nivitha

Introduction

Fast Tag is a revolutionary system designed to facilitate faster toll payments on highways using radio-frequency identification (RFID) technology. As convenient it seems, there lies the risk of fraudulent activities associated with it. This project focuses on developing a robust fraud detection system for Fast Tag transactions to ensure the integrity and security of the system.

Objectives

- Implementing real-time monitoring of Fast Tag transactions.
- Developing algorithms to detect suspicious patterns and anomalies.
- Enhancing the security and trustworthiness of the Fast Tag system.

Methodologies

- Data Collection
- Feature Engineering
- Machine Learning Model
- Training data
- Testing data

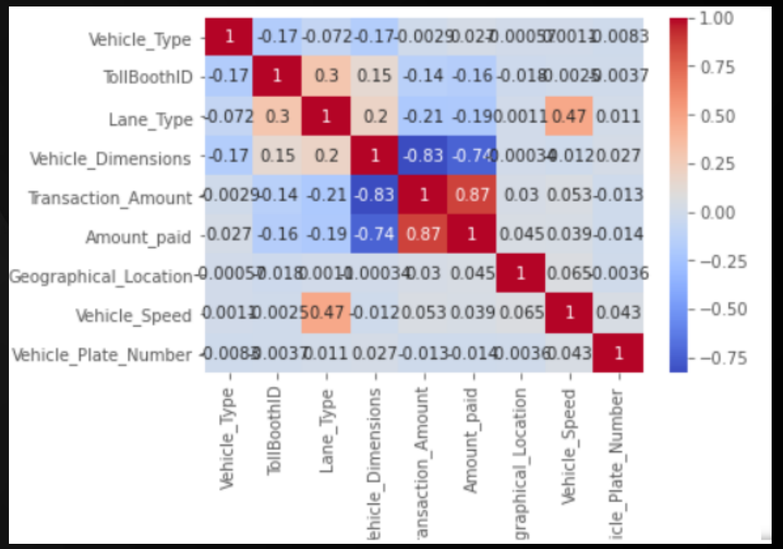
Data Collection

The dataset that was attached with the project description is used to model the fraud detection system. It contains features such as transaction details, vehicle information, geographical location, and transaction amounts.

Feature Engineering

Bivariate analysis and non parametric methods are employed on the features to select the appropriate features for the model. Through the analysis the features selected are

Transaction ID, Vehicle Type, TollBoothID, Lane Type, Vehicle Dimensions, Amount paid, Geographical Location, Vehicle Speed



Machine Learning Model

- The Random Forest model has ensemble learning approach, which combines multiple decision trees to handle complex and imbalanced data effectively.
- It can handle imbalanced datasets without extensive preprocessing, provide feature importance insights, reduce overfitting, and scale well to large datasets makes it ideal for real-time fraud detection systems.
- Its interpretability allows for the explanation of flagged transactions to stakeholders and investigators.
- The Random Forest model offers a balanced combination of accuracy, robustness, interpretability, and scalability, making it well-suited for identifying fraudulent activities in Fast Tag transactions.

Training Data

- 80% of the data is used to train the data
- The ‘train_test_split’ from sklearn library is used to split the data into testing and training data
- Data preprocessing such as handling missing values, encoding categorical variables, scaling numerical features, etc are done before training the data

Training Data

- 20% of the data is used to test the model.
- The prediction is done for the 20% data and the predictions made are tested with the actual values.

ACCURACY OF THE MODEL

99.1%

THANK YOU