ML Problem Statement

Mentorness Internship Program

Problem Statement:

This internship project focuses on leveraging machine learning classification techniques to develop an effective fraud detection system for Fastag transactions. The dataset comprises key features such as transaction details, vehicle information, geographical location, and transaction amounts. The goal is to create a robust model that can accurately identify instances of fraudulent activity, ensuring the integrity and security of Fastag transactions.

Dataset Description:

1. Transaction_ID: Unique identifier for each transaction.
2. Timestamp: Date and time of the transaction.
3. Vehicle_Type: Type of vehicle involved in the transaction.
4. FastagID: Unique identifier for Fastag.
5. TollBoothID: Identifier for the toll booth.
6. Lane_Type: Type of lane used for the transaction.
7. Vehicle_Dimensions: Dimensions of the vehicle.
8. Transaction_Amount: Amount associated with the transaction.
9. Amount_paid: Amount paid for the transaction.
10. Geographical_Location: Location details of the transaction.
11. Vehicle_Speed: Speed of the vehicle during the transaction.
12. Vehicle_Plate_Number: License plate number of the vehicle.
13. Fraud_indicator: Binary indicator of fraudulent activity (target variable).

Project Objectives:

1. Data Exploration:
   - Explore the dataset to understand the distribution of features and the prevalence of fraud indicators.

2. Feature Engineering:
   - Identify and engineer relevant features that contribute to fraud detection accuracy.

3. Model Development:
   - Build a machine learning classification model to predict and detect Fastag transaction fraud.
   - Evaluate and fine-tune model performance using appropriate metrics.

4. Real-time Fraud Detection:

- Explore the feasibility of implementing the model for real-time Fastag fraud detection.

5. Explanatory Analysis:

- Provide insights into the factors contributing to fraudulent transactions.

Challenges:

- Imbalanced dataset issues due to the likely low occurrence of fraud.

- Feature engineering to capture nuanced patterns indicative of fraud.

Evaluation Criteria:

- Model performance assessed using metrics such as precision, recall, F1 score, and accuracy.

Deliverables:

- Trained machine learning model for Fastag fraud detection.

- Evaluation metrics and analysis report.

- Documentation on relevant features and their impact on fraud detection.

Expected Outcome:

- An effective and scalable Fastag fraud detection system capable of minimizing financial losses and ensuring the security of digital toll transactions.