

# Interim Documentation

## Project: Fraud Detection for E-commerce and Banking Transactions

**Team:** Adey Innovations Inc. Data Science Team

**Prepared by:** [Your Name]

**Date:** [Insert Date]

### 1. Project Overview

This project aims to develop machine learning models to detect fraudulent transactions in e-commerce and banking systems. The primary objectives are:

- To analyze and preprocess transaction data.
- To create engineered features for fraud detection.
- To train and optimize fraud detection models.
- To deploy real-time fraud detection systems with monitoring.

### 2. Progress Summary

#### 2.1 Data Collection & Cleaning

##### ☑ Completed Tasks:

- Collected **Fraud\_Data.csv** and **IpAddress\_to\_Country.csv** datasets.
- Handled missing values through imputation and removal.
- Removed duplicate records and corrected data types.

##### ☐ Challenges:

- Some missing values in IP addresses may require external sources for enrichment.

#### 2.2 Exploratory Data Analysis (EDA)

##### ☑ Completed Tasks:

- Conducted **univariate and bivariate analysis** to understand fraud patterns.
- Identified **fraudulent transaction trends** based on transaction amount and time.
- Merged datasets for **geolocation-based fraud analysis**.

##### ☐ Challenges:

- Some transactions lack geolocation data, requiring imputation strategies.

## 2.3 Feature Engineering

### ✅ Completed Tasks:

- Created features such as **transaction velocity, frequency, and time-based indicators**.
- Engineered new features:
  - **Hour of the day & Day of the week** to detect unusual transactions.
  - **IP-based country mapping** to track suspicious activities.

### ❏ Challenges:

- Need further testing to confirm feature effectiveness in improving model accuracy.

## 2.4 Model Development

### ✅ Completed Tasks:

- Baseline models trained using **Logistic Regression and Decision Trees**.
- Initial model evaluation with **Precision, Recall, and F1-score metrics**.

### ❏ Challenges:

- Current models have high false positive rates, requiring further tuning.
- Need to experiment with **advanced models (Random Forest, XGBoost, Neural Networks)**.

## 3. Next Steps

Task	Expected Completion
Feature selection and optimization	Week 9
Train and evaluate advanced ML models	Week 9
Implement real-time fraud detection system	Week 10
Model deployment and monitoring setup	Week 10

## 4. Risks & Mitigation Strategies

Risk	Impact	Mitigation Strategy
High false positive rate	Reduces trust in fraud detection	Fine-tune models using better feature selection and threshold tuning
Missing geolocation data	Reduces accuracy in detecting location-based fraud	Use external APIs to enrich missing IP data
Imbalanced dataset	Model may favor non-fraudulent transactions	Use <b>SMOTE (Synthetic Minority Over-sampling Technique)</b> to balance data

## 5. Conclusion

The project is progressing well, with data preprocessing and feature engineering nearly complete. The next phase will focus on model improvement and deployment. The team is addressing challenges such as high false positives and missing data to enhance fraud detection accuracy.

---