Air University Aerospace and Aviation Campus Kamra

Data Science

ASSIGNMENT 04

Submitted On: Dec 20, 2025

Name: Ahmad Faraz

Registration No :215154

Department: BSCS-VII

Mr. Ghulam Ali

Table of Contents

1. Introduction

This assignment focuses on transitioning the Ethereum Fraud Detection system from model development to real-world deployment. The trained machine learning model from Assignment 03 is deployed using a lightweight web framework, feedback is collected from users, and an improvement roadmap is proposed based on real usage insights.

2. Deployment Tool Comparison

1. Streamlit (Selected Tool)

- Rapid development and minimal boilerplate
- Ideal for data science demos and ML models
- Built-in UI components

2. Flask

- Flexible backend framework
- Requires manual UI and routing
- More setup overhead

3. FastAPI

- High performance and async support
- Better for production APIs
- Overkill for academic ML deployment

Justification: Streamlit was selected due to its simplicity, fast prototyping, and suitability for local and academic deployments.

3. Model Utilization from Assignment 03

A. Models Developed in Assignment 03

Assignment 03 trained and serialized two main models for fraud detection:

1. Logistic Regression Model (logistic_regression_model.pkl)

- Baseline interpretable model
- Uses scaled features
- Fast predictions with probability outputs
- Suitable for simple, linear fraud patterns

2. Random Forest Model (random_forest_model.pkl)

- Ensemble-based model
- Captures non-linear relationships
- Superior performance in Assignment 03
- Better at handling complex fraud patterns

3. Feature Scaler (scaler.pkl)

- Fitted StandardScaler from Assignment 03 training data
- Required for Logistic Regression input preprocessing
- Ensures feature consistency across deployment

B. Assignment 03 Training Process

- Training Data: 75% of cleaned Ethereum fraud dataset
- Testing Data: 25% with stratified sampling (preserved class imbalance)
- Features Used: Numeric transactional and ERC20-based metrics
- Target Variable: FLAG (1 = Fraudulent, 0 = Legitimate)
- Performance: Random Forest achieved superior ROC-AUC and recall

C. Model Selection for Deployment

Random Forest selected for Production Deployment because:

- Higher accuracy and ROC-AUC score
- Better recall for detecting fraudulent addresses (critical for security)
- Non-linear pattern recognition capabilities
- Probability outputs for risk scoring

4. Deployment Process

The model is deployed locally using Streamlit. Users can input transaction features and receive a fraud risk prediction.

Steps to Run Locally:

1. Install Streamlit: pip install streamlit
2. Save the app code as app.py

3. Run: streamlit run app.py

5. Feedback Collection & Analysis

The deployed application was shared with at least 15 users. Feedback was collected using a Google Form and stored in a CSV file named Suggestions_Dataset.csv.

Feedback fields included:

- Usability
- Prediction relevance
- Suggestions for improvement

6. Complete Model Utilization Pipeline

Includes data flow diagram and model comparison table.

Conclusion: Random Forest selected for primary deployment due to superior fraud detection capability (higher recall) and better overall performance.

7. Improvement Plan and Versioning

Version 2.0 Planned Improvements:

- Add more input features for better accuracy
- Improve UI clarity and tooltips
- Add probability-based risk scoring

Prioritization: Accuracy and usability improvements were prioritized based on recurring user feedback.

8. Assumptions & Limitations

Assumptions:

- Cleaned data represents real-world behavior
- Users input realistic transaction values

Limitations:

- Class imbalance may bias predictions
- Model generalization limited to Ethereum network
- Local deployment scalability constraints

9. Conclusion

This assignment demonstrated the complete lifecycle of a data science project, from model deployment to feedback-driven improvement. Deployment revealed practical challenges and user expectations, guiding the roadmap for future versions.