PREDICTIVE ANALYTICS

PROJECT REPORT

TITLE:-

"UPI Risk Intelligence System – Fraud Detection"

Submitted by :-

Aditya Raj

Registration No:- 12305995

Programme: B.Tech(CSE)

Section: K23KS

Course Code:- INT-234

Under the Guidance of

(Dr. Madhu Bala (UID-31770))

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Aditya Raj bearing Registration no. 12305995 has completed INT-234 project titled, **“UPI Risk Intelligence System – Fraud Detection.”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

School of Computer Science & Engineering
Lovely Professional University
Phagwara, Punjab.

DECLARATION

I, Aditya Raj, Student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 14/12/2025 Signature

Registration No. 12305995 Aditya Raj

Acknowledgement

I would like to express my sincere gratitude to my faculty guide for providing continuous support, guidance, and motivation throughout the completion of this project. I am also thankful to the Lovely Professional University for providing the necessary resources and learning environment. Finally, I thank my family and friends for their encouragement and support.

TABLE OF CONTENTS

1. Introduction

In recent years, digital payment systems have transformed the financial ecosystem of India. Among these systems, **Unified Payments Interface (UPI)** has emerged as one of the most widely used platforms due to its speed, convenience, and accessibility. Millions of transactions are processed daily through UPI for personal and commercial purposes.

However, the rapid increase in UPI usage has also led to a rise in **fraudulent transactions**, including unauthorized payments, phishing attacks, and fake merchant scams. Traditional rule-based fraud detection systems often fail to detect complex and evolving fraud patterns.

This project focuses on developing a **machine learning-based UPI fraud detection system** that can intelligently analyze transaction details and predict the likelihood of fraud. The system combines data preprocessing, feature engineering, and a Random Forest classification model with a **Streamlit-based interactive dashboard** for real-time analysis.

2. OBJECTIVE OF THE PROJECT:-

The primary objectives of this project are as follows:

- To study and analyze UPI transaction data
- To identify important features that contribute to fraudulent behavior
- To build a machine learning model capable of detecting fraud
- To classify transactions into low, moderate, and high-risk categories
- To deploy the trained model using a user-friendly Streamlit web application
- To enhance digital payment security using predictive analytics

3. Source of Dataset

Dataset Link : https://www.kaggle.com/datasets/skullagos5246/upi-transactions-2024-dataset

The dataset used in this project consists of **simulated UPI transaction records** representing real-world transaction behavior. The data is stored in CSV format and includes both legitimate and fraudulent transactions.

Key attributes of the dataset include:

- Transaction amount
- Hour of transaction
- Day of week
- Weekend indicator
- Device type (Android / iOS)
- Network type (4G / 5G / WiFi)
- Sender bank name
- Receiver bank name
- Fraud flag (target variable)

This dataset is suitable for supervised machine learning and fraud detection analysis.

4. DATASET PREPROCESSING

Dataset preprocessing is a critical step to ensure accuracy and reliability of the machine learning model.

4.1 Data Cleaning

1) Column names were standardized for consistency
2) Missing and inconsistent values were identified
3) Bank risk values were filled using mean imputation where required

4.2 Encoding of Categorical Variables

Machine learning models require numerical inputs. Therefore:

- Categorical variables such as day of week, device type, network type, sender bank, and receiver bank were encoded using **One-Hot Encoding**

4.3 Feature Scaling

- Numerical features like transaction amount and transaction hour were scaled using **StandardScaler**
- This ensured equal contribution of features during model training

5. FEATURE ENGINEERING

- To improve fraud detection accuracy, additional features were created:
- Late Night Transaction: Identifies transactions occurring during late night hours
- High Amount Transaction: Flags transactions above the 90th percentile amount
- Sender Bank Risk Score: Historical fraud probability of sender bank
- Receiver Bank Risk Score: Historical fraud probability of receiver bank
- These engineered features help capture hidden fraud patterns and behavioral anomalies.
- Outlier Detection and Removal

6. MODEL DEVELOPMENT

The project uses a **Random Forest Classifier**, an ensemble machine learning algorithm known for high accuracy and robustness.

Model Architecture:

- ColumnTransformer for preprocessing
- StandardScaler for numerical features
- OneHotEncoder for categorical features
- RandomForestClassifier with class balancing

The dataset was split into **80% training data** and **20% testing data** using stratified sampling to handle class imbalance.

7. MODEL EVALUATION

The model performance was evaluated using standard classification metrics:

- Accuracy
- Precision
- Recall
- F1-Score

Random Forest performed effectively in detecting fraudulent transactions due to its ability to handle complex feature interactions and imbalanced data

8. SYSTEM ARCHITECTURE & DEPLOYMENT

The system follows a modular architecture:

1. Transaction input from user
2. Data preprocessing pipeline
3. Machine learning fraud prediction
4. Risk scoring and classification
5. Streamlit-based visualization

The trained model is saved using **Joblib** and dynamically loaded into the Streamlit application.

9. STREAMLIT APPLICATION DESCRIPTION

The Streamlit dashboard provides:

- Secure login authentication
- Transaction input form
- Fraud probability score
- Risk level classification (Low / Moderate / High)
- Recommended action (Allow / Review / Block)
- AI assistant for fraud-related guidance

This interface makes the system interactive and easy to use.

10. RESULTS & DISCUSSION

The system successfully identifies high-risk transactions and provides real-time fraud probability. Transactions with high risk are flagged for review or blocking, while low-risk transactions are allowed smoothly.

Feature engineering and bank risk analysis significantly improved the model's prediction capability.

---

11. CONCLUSION

This project successfully demonstrates the application of **machine learning and predictive analytics** in detecting fraudulent UPI transactions. The combination of Random Forest classification, feature engineering, and Streamlit deployment provides an effective and scalable fraud detection solution.

The project highlights the importance of intelligent systems in enhancing digital payment security.

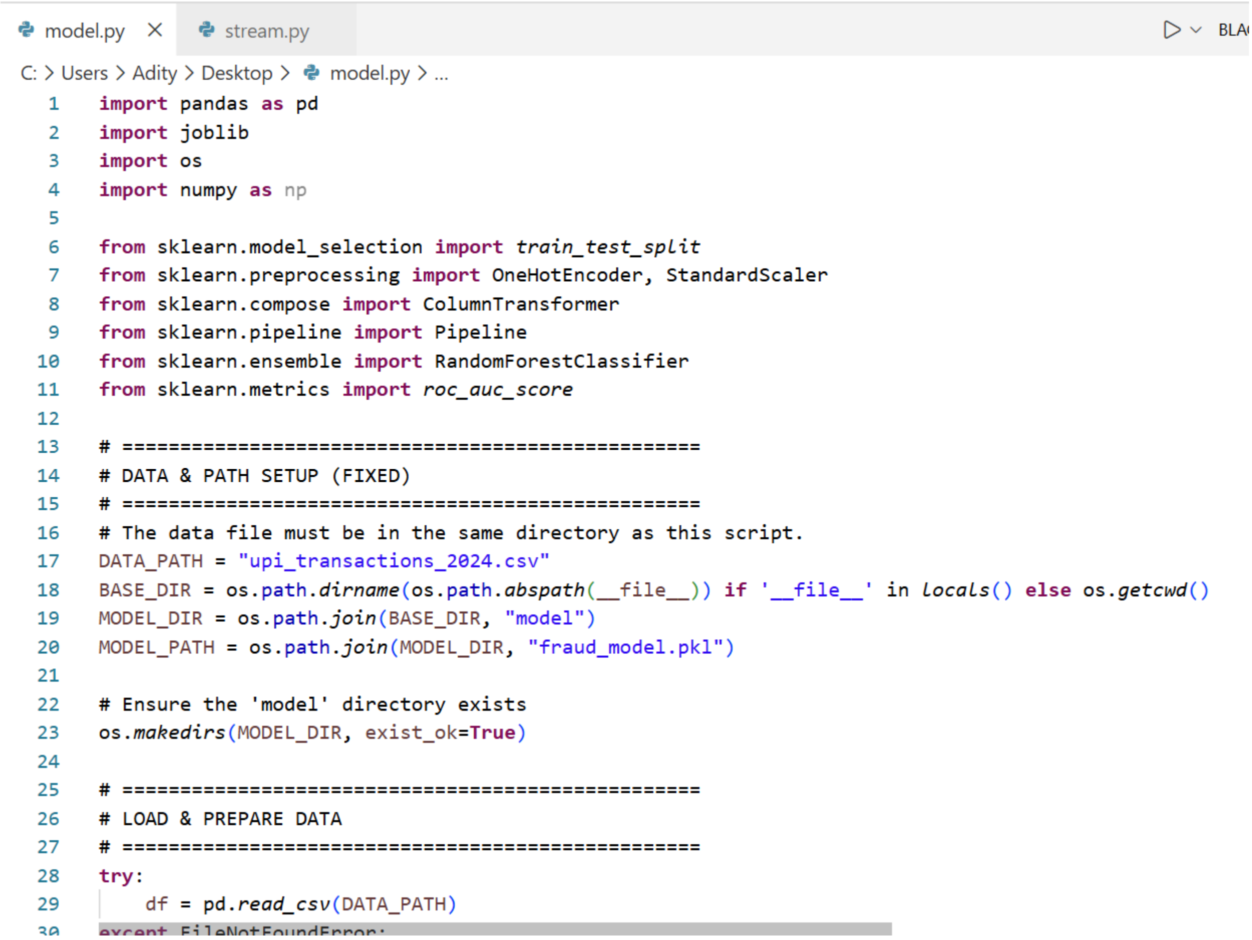
12. FUTURE SCOPE

The system can be enhanced in the future by:

- Integrating real-time UPI transaction APIs
- Using deep learning models for better accuracy
- Deploying on cloud platforms
- Adding SMS/email fraud alerts
- Developing a mobile application version

13. REFERENCES

- Scikit-learn Documentation
- Streamlit Official Documentation
- Digital Payment Security Reports



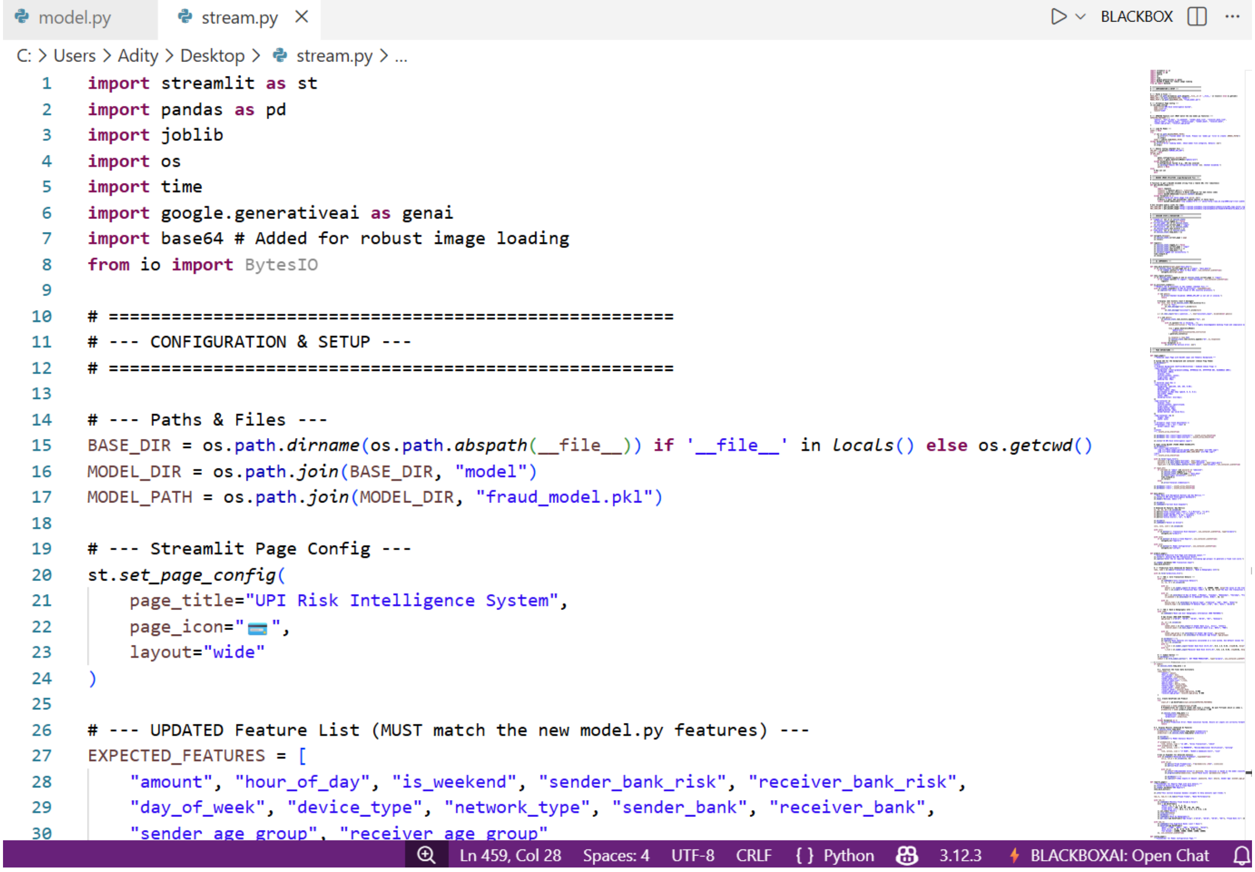
```python
import pandas as pd
import joblib
import os
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score

# ==================================================
# DATA & PATH SETUP (FIXED)
# ==================================================
# The data file must be in the same directory as this script.
DATA_PATH = "upi_transactions_2024.csv"
BASE_DIR = os.path.dirname(os.path.abspath(__file__)) if '__file__' in locals() else os.getcwd()
MODEL_DIR = os.path.join(BASE_DIR, "model")
MODEL_PATH = os.path.join(MODEL_DIR, "fraud_model.pkl")

# Ensure the 'model' directory exists
os.makedirs(MODEL_DIR, exist_ok=True)

# ==================================================
# LOAD & PREPARE DATA
# ==================================================
try:
    df = pd.read_csv(DATA_PATH)
except FileNotFoundError:
```



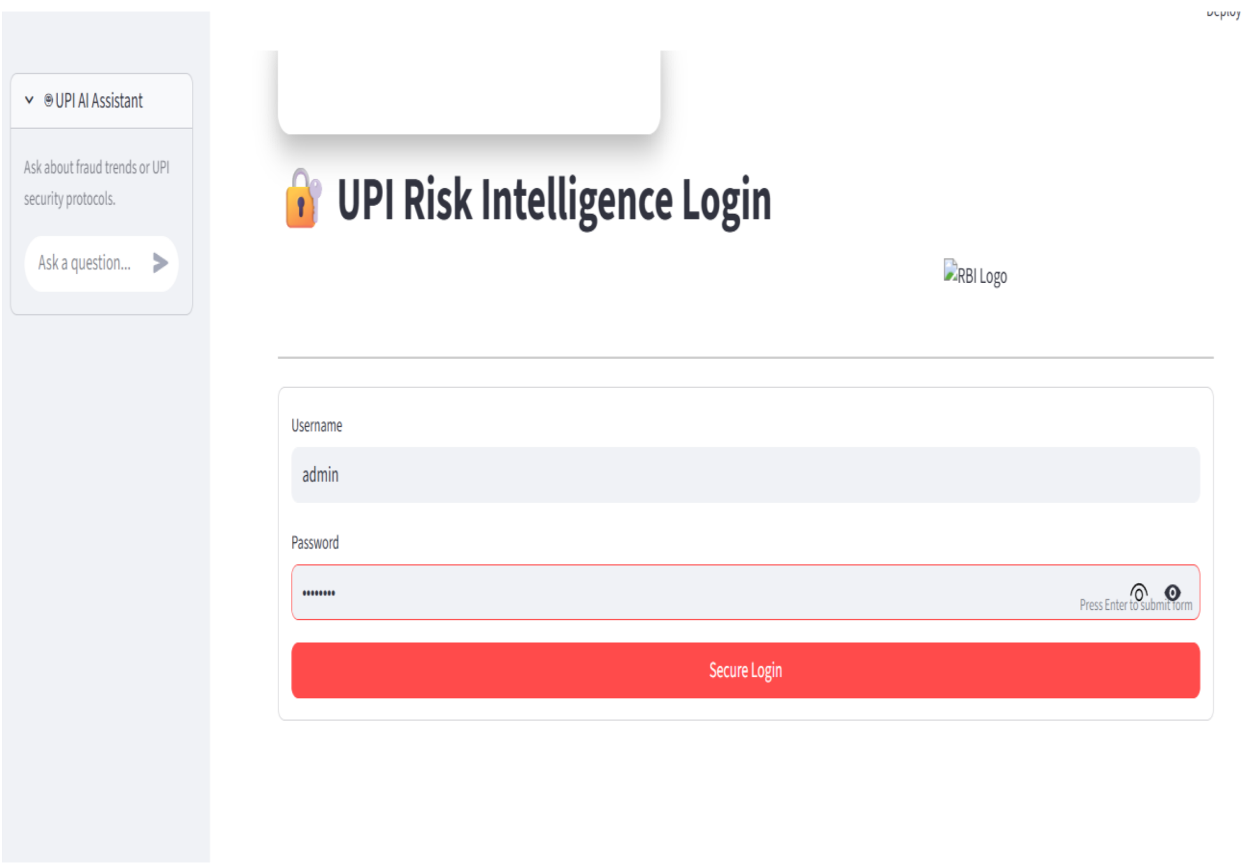
```python
import streamlit as st
import pandas as pd
import joblib
import os
import time
import google.generativeai as genai
import base64 # Added for robust image loading
from io import BytesIO

# ======================================================
# --- CONFIGURATION & SETUP ---
# ======================================================

# --- Paths & Files ---
BASE_DIR = os.path.dirname(os.path.abspath(__file__)) if '__file__' in locals() else os.getcwd()
MODEL_DIR = os.path.join(BASE_DIR, "model")
MODEL_PATH = os.path.join(MODEL_DIR, "fraud_model.pkl")

# --- Streamlit Page Config ---
st.set_page_config(
    page_title="UPI Risk Intelligence System",
    page_icon="💳",
    layout="wide"
)

# --- UPDATED Feature List (MUST match the new model.py features) ---
EXPECTED_FEATURES = [
    "amount", "hour_of_day", "is_weekend", "sender_bank_risk", "receiver_bank_risk",
    "day_of_week", "device_type", "network_type", "sender_bank", "receiver_bank",
    "sender_age_group", "receiver_age_group"
```



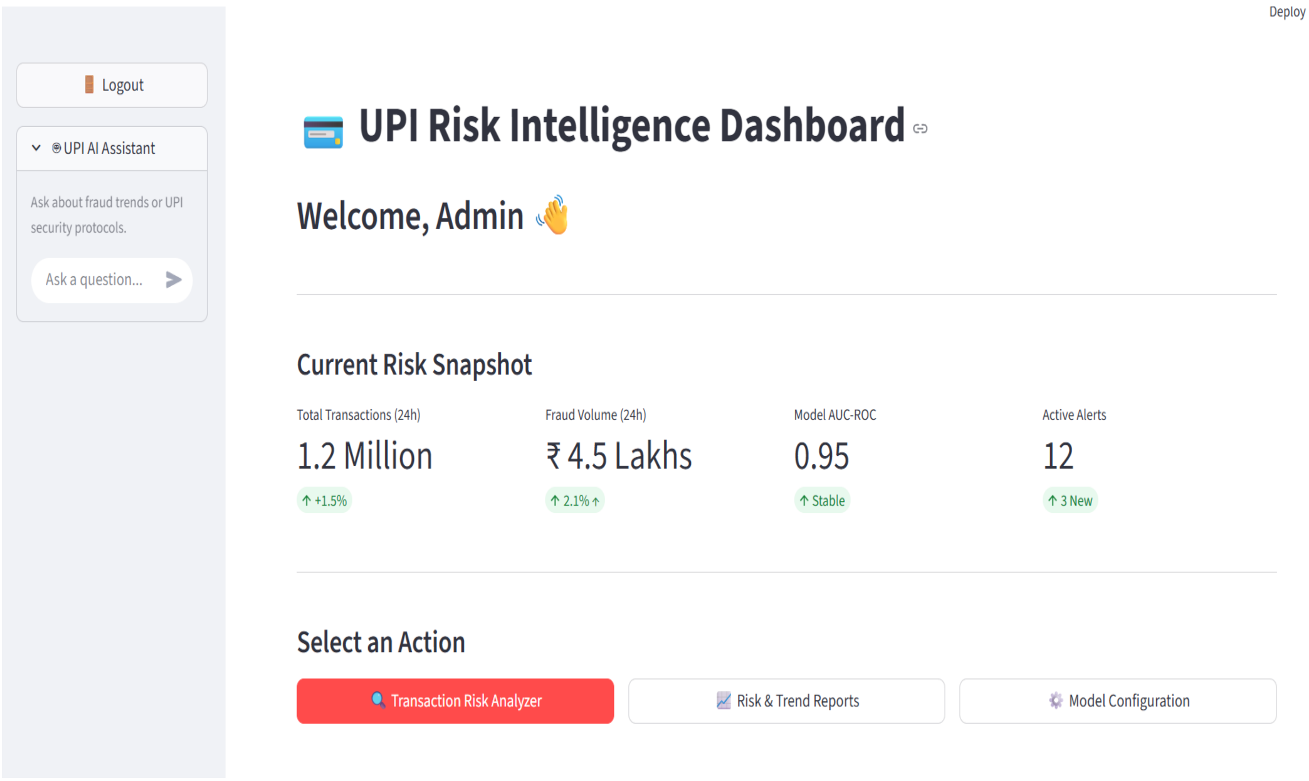
Ask about fraud trends or UPI security protocols.

🔒 UPI Risk Intelligence Login

Username: admin

Password: ••••••••

Press Enter to submit form



Ask about fraud trends or UPI security protocols.

💳 UPI Risk Intelligence Dashboard

Welcome, Admin 👋

Current Risk Snapshot

Total Transactions (24h)	Fraud Volume (24h)	Model AUC-ROC	Active Alerts
1.2 Million	₹ 4.5 Lakhs	0.95	12
↑ +1.5%	↑ 2.1% ↑	↑ Stable	↑ 3 New

Select an Action