# Capstone Project: Fraud Detection System for JPMorgan Chase - Case Study

### Problem Statement

JPMorgan Chase, a leading global financial institution, is facing an increase in fraudulent transactions due to sophisticated cyber threats and financial fraud schemes. The bank processes millions of transactions daily, making it crucial to detect and prevent fraudulent activities in real time.

Your task is to build a **Fraud Detection System** that leverages **real-time transaction data** and **machine learning models** to detect anomalies. The system should alert the security team immediately when suspicious transactions are detected to **enhance security and minimize financial losses**.

---

## Project Objectives

1. **Data Ingestion & Preprocessing**
   - Process transaction data from various sources (**CSV, SQL, API, NoSQL**)
   - Clean and normalize data to handle missing values, inconsistencies, and outliers.
   - Store the processed data in a structured format for further analysis.
2. **Fraud Detection Model Development**
   - Implement **Supervised Learning Models** (e.g., Logistic Regression, Random Forest) using historical fraud-labeled data.
   - Use **Unsupervised Anomaly Detection** techniques (e.g., Isolation Forest, Autoencoders) for identifying new fraud patterns.
   - Perform **Feature Engineering** to extract meaningful transaction patterns.
3. **Real-Time Anomaly Detection & Alerts**
   - Develop an **automated pipeline** to process real-time transactions.
   - Detect fraudulent activities based on historical and live data.
   - Generate **alerts for security teams** to take action on flagged transactions.
4. **Visualization & Reporting**
   - Build **interactive dashboards in Power BI/Tableau** to monitor fraudulent activities.
   - Provide **detailed reports** on fraudulent patterns, transaction trends, and high-risk regions.
5. **Model Evaluation & Optimization**
   - Compare multiple fraud detection models and select the best-performing one.
   - Tune hyperparameters improve model accuracy and reduce false positives.
   - Deploy the model for real-time fraud monitoring.

---

## Data Dictionary

| Column Name | Data Type | Description | Possible Values / Example Data |
|---|---|---|---|
| **Transaction_Id** | String | Unique identifier for each transaction | "PAY-BILL-3589", "WITHDRAWAL-3591" |
| **Sender_Id** | String | Unique ID assigned to the customer initiating the transaction | "CLIENT-3566", "CLIENT-3272" |
| **Sender_Account** | String | Bank account number of the sender | "ACCOUNT-3578", "ACCOUNT-3284" |
| **Sender_Country** | String | Country where the sender's account is registered | "USA", "Germany", "Canada" |
| **Sender_Sector** | Integer | Industry sector of the sender | 21264 (Banking), 4809 (Retail) |
| **Sender_Job** | String | Job type of the sender | "CCB" (Corporate Banking), "IND" (Individual) |
| **Bene_Id** | String | Unique ID assigned to the recipient (beneficiary) | "COMPANY-3574", "CLIENT-3333" |
| **Bene_Account** | String | Bank account number of the recipient | "ACCOUNT-3587", "ACCOUNT-3338" |
| **Bene_Country** | String | Country where the recipient's account is registered | "Germany", "USA", "Canada" |
| **USD_Amount** | Float | Transaction amount in US dollars | 492.67, 388.92, 730.69 |

| | | | |
|---|---|---|---|
| **Transaction_Type** | String | Type of transaction | "MAKE-PAYMENT", "WITHDRAWAL", "MOVE-FUNDS", "DEPOSIT-CASH", "QUICK-PAYMENT" |
| **Transaction_Mode** | String | Mode of transaction execution | "Online", "ATM", "Mobile Banking", "Bank Transfer" |
| **Transaction_Status** | String | Status of the transaction | "Successful", "Pending", "Failed" |
| **Time_Stamp** | DateTime | Timestamp of when the transaction was executed | "2024-03-05 14:23:10" |
| **Device_Type** | String | Type of device used for transaction | "Mobile", "Desktop", "ATM", "POS" |
| **IP_Address** | String | IP address from which the transaction was initiated | "192.168.1.1", "203.45.23.101" |
| **Fraud_Flag** | Integer | Whether the transaction is fraudulent (1) or legitimate (0) | 0 (Legit), 1 (Fraud) |
| **Fraud_Risk_Score** | Float | Probability of the transaction being fraudulent (computed by ML Model) | 0.12, 0.85, 0.97 |
| **Alert_Status** | String | Status of fraud alert triggered | "No Alert", "Review Needed", "Blocked" |

## Additional Notes

- **Transaction_Type**: Helps in identifying common fraudulent transaction patterns (e.g., frequent small withdrawals may indicate money laundering).
- **Sender & Beneficiary Country**: Cross-border transactions are often higher-risk.
- **Transaction_Mode & Device_Type**: Fraudsters often use **new/unrecognized devices**.
- **IP_Address**: Useful for detecting **unusual geographic locations**.

- **Fraud_Risk_Score**: Calculated using **Machine Learning models** to provide a probability estimate.
- **Alert_Status**: Transactions above a certain risk score (e.g., >0.90) may be **automatically blocked**.

---

| Deliverable | Marks |
|---|---|
| **Cleaned and Processed Dataset** - A structured dataset ready for fraud detection analysis. | **15 Marks** |
| **Exploratory Data Analysis (EDA) Report** - Insights into fraudulent transactions and trends. | **15 Marks** |
| **Fraud Detection Model** - Machine learning model trained for fraud classification. | **25 Marks** |
| **Real-Time Fraud Monitoring System** - A deployed model with fraud alert triggers. | **25 Marks** |
| **Fraud Investigation Dashboard** - Power BI/Tableau dashboard for fraud monitoring. | **15 Marks** |
| **Final Report & Presentation** - Summary of findings, model performance, and business recommendations. | **5 Marks** |