Fraud Detection in Mobile **Financial Transactions**

By - Anurag Tripathi

Content

Introduction

- In the digital world, millions of money transactions happen every second. But not all of them are safe. Some are fraudulent, where attackers try to trick the system and steal money.

- To catch such fraud, we don’t just look at who sent the money or how much was sent — we also look at how the account balances change during the transaction.

- This project aims to build a machine learning model that can detect fraudulent transactions in real-time, allowing companies to:
  - Secure customer accounts,
  - Reduce financial losses,
  - Maintain customer trust,
  - Meet compliance requirements.

Problem Statement

Mobile financial transactions are growing rapidly due to their ease and speed. However, this digital convenience has also opened doors for advanced fraud techniques. Detecting fraud in real-time is a serious challenge for banks and fintech platforms, demanding intelligent solutions that go beyond traditional rules.

GLOBAL INSIGHTS

- Over billions of mobile transactions occur worldwide every day.
- Even 0.1% fraudulent activity leads to major financial loss.
- Rise in cross-border fraud and digital identity theft due to increasing internet access.

CHALLENGES IN DETECTION

- Instant transactions leave no room for manual verification.
- Static rules are ineffective against new fraud tactics.
- Difficulty in distinguishing between genuine behavior and disguised fraud.

WHY AN AI-BASED APPROACH?

- AI models can learn from past patterns and adapt to new fraud behavior.
- Capable of flagging suspicious transactions before they are completed.
- Helps protect customers and institutions while preserving trust and compliance.

Why Our Model is Useful

Our machine learning model can help financial institutions by:

Fraud Detection **Analysis**

REAL-TIME FRAUD DETECTION

- Automatically flags fraud before it happens, minimizing risk.

ADAPTIVE LEARNING

- Detects new, evolving fraud patterns that rule-based systems miss.

REDUCE FINANCIAL LOSSES

- Catching fraud early prevents money from being stolen or misused.

IMPROVE CUSTOMER TRUST

- When customers feel safe, they continue using the platform.

OBJECTIVES

Detect Fraudulent Transactions

- Build a machine learning model to accurately identify and flag suspicious transactions in real-time.

Reduce Financial Losses

- Minimize potential monetary damage to users and financial institutions by detecting fraud before completion.

Enhance Security and Trust

- Strengthen user trust by improving transaction security using intelligent fraud detection.

Automate Fraud Monitoring

- Develop a system that can automatically learn and adapt to new fraudulent behaviors without human intervention.

Improve Decision-Making

- Provide actionable insights through analytics and visualizations to assist financial teams in making faster, informed decisions.

Ensure Scalability

- Design the solution to handle high-volume transaction data efficiently, suitable for real-world deployment.

WORKFLOW

Data Collection

- Loaded financial transaction data from .csv file containing time-step, transaction type, amount, balance, and fraud labels.

Preprocessing

- Handled missing values and removed irrelevant columns.
- Converted categorical variables (like type) into numerical values using encoding.

Exploratory Data Analysis (EDA)

- Visualized transaction trends, types, and fraud distribution.
- Identified imbalance in fraud vs. normal transactions.
- Analyzed relationships between features such as amount and old balance.

Feature Engineering

- Created new features like transaction difference, balance difference, etc.
- Removed outliers and ensured data quality to improve model learning.

Model Selection & Training

- Chose machine learning models like Logistic Regression, Random Forest, XGboost.
- Trained models using training data split and applied techniques like cross-validation.

Model Evaluation

- Evaluated models using Accuracy, Precision, Recall, and ROC-AUC score.
- Visualized confusion matrix and ROC curves to compare model performance.

Deployment

- Selected the best-performing model and prepared it for deployment.
- Deployed it on streamlit

RESULTS

Models Evaluated

Logistic Regression:

- Simple baseline model, but underperformed on imbalanced data.

Random Forest Classifier:

- Improved generalization and reduced overfitting via ensemble learning.

XGBoost Classifier (Selected Model):

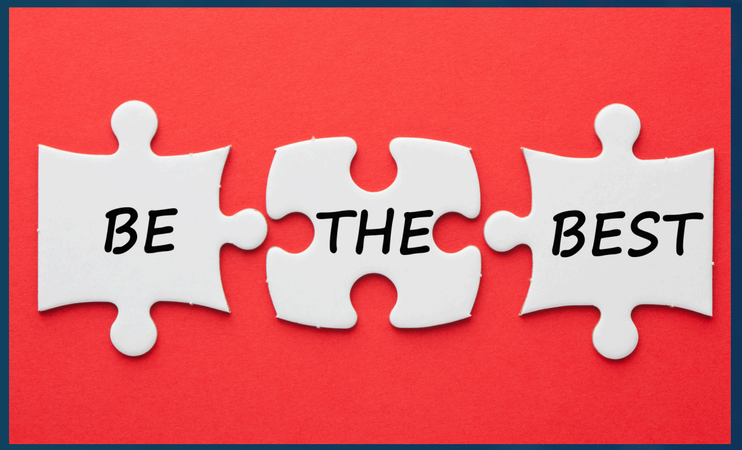
- Outperformed all others in terms of accuracy, recall, and AUC.
- Chosen as the final model for deployment due to its robustness and high performance on imbalanced classification problems.

LightGBM Classifier

- A gradient boosting framework by Microsoft optimized for speed and efficiency.
- Comparable to XGBoost but slightly lower in recall and AUC on this dataset.

FINAL MODEL

XGBOOST CLASSIFIER



OVERALL

- Selected as the best-performing model.
- Known for being fast, accurate, and scalable for large datasets.

STRENGTHS:

- Excellent accuracy, precision, and recall.
- Effectively handled data imbalance and large feature space.

REASON FOR SELECTION:

- Outperformed all models in Accuracy (99.92%), Recall (97%), and AUC (0.999).
- Highly robust and reliable for real-world fraud detection tasks.

Performance Metrics (XGBoost)

	Precision	Recall	F1-Score	Support
0 (Non-Fraud)	1.00	1.00	1.00	1999
1 (Fraud)	0.98	0.99	0.98	230
Accuracy	-	-	0.9969	2229
Macro Avg	0.99	0.99	0.99	2229
Weighted Avg	1.00	1.00	1.00	2229

PERFORMANCE METRICS

CLASSIFICATION REPORT

Accuracy: 99.92%

- Almost all transactions were correctly classified.

Precision (Fraud Class): 99%

- Very few false positives. Fraud predictions are highly reliable.

Recall (Fraud Class): 97%

- Most actual frauds were successfully caught.

AUC-ROC Score: 0.999

- Excellent ability to distinguish between fraudulent and legitimate transactions.

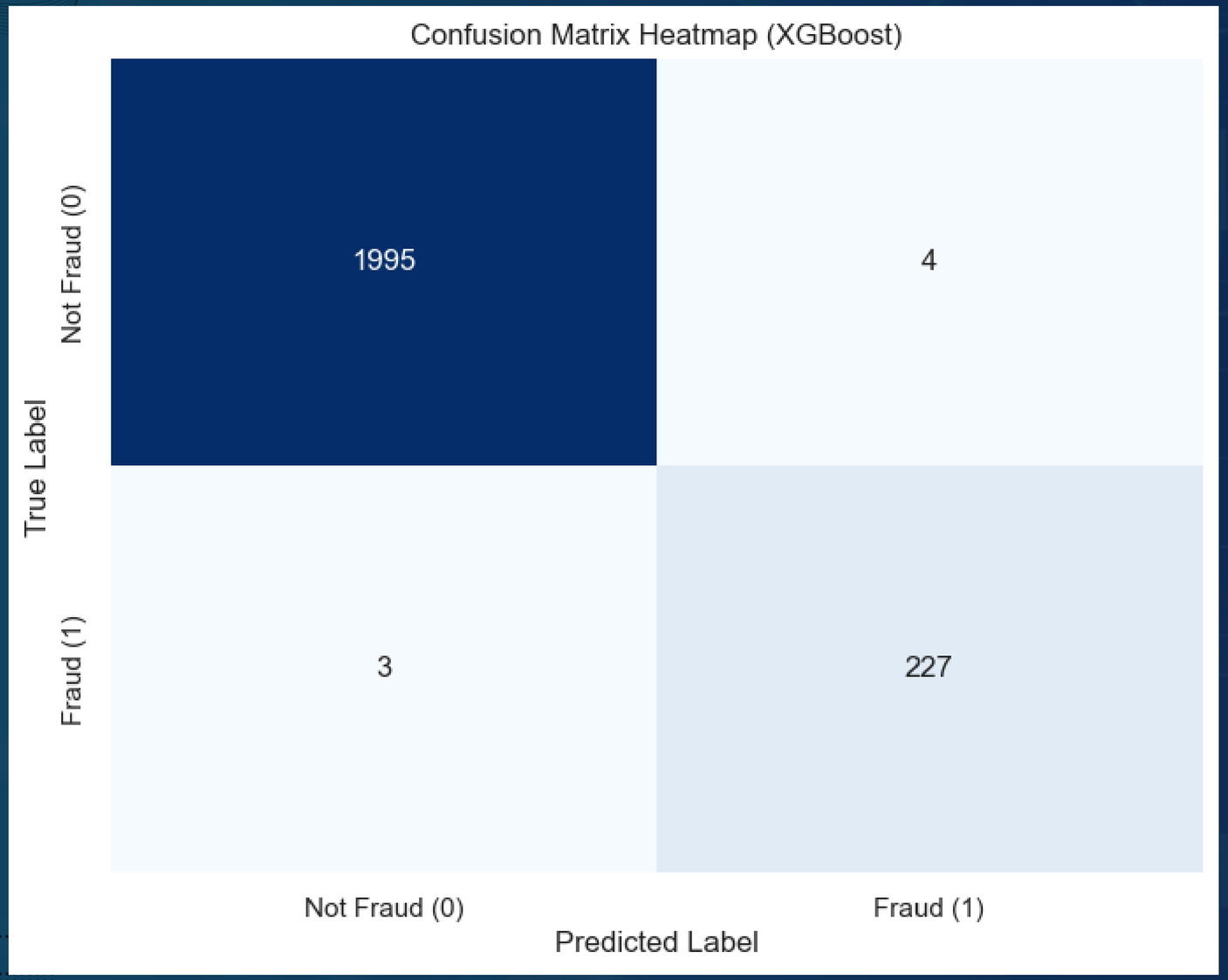
--- ROC Curves and AUC Scores ---

- Logistic Regression AUC: 0.9866
- Random Forest AUC: 0.9950
- XGBoost AUC: 0.9978
- LightGBM AUC: 0.9986

F1-Score: 0.98

- Strong balance between precision and recall.

CONFUSION MATRIX--> XGBoost



True Negatives (1995): The model correctly identified 1995 legitimate transactions as Not Fraud.

True Positives (227): 227 actual fraudulent transactions were correctly identified as fraud.

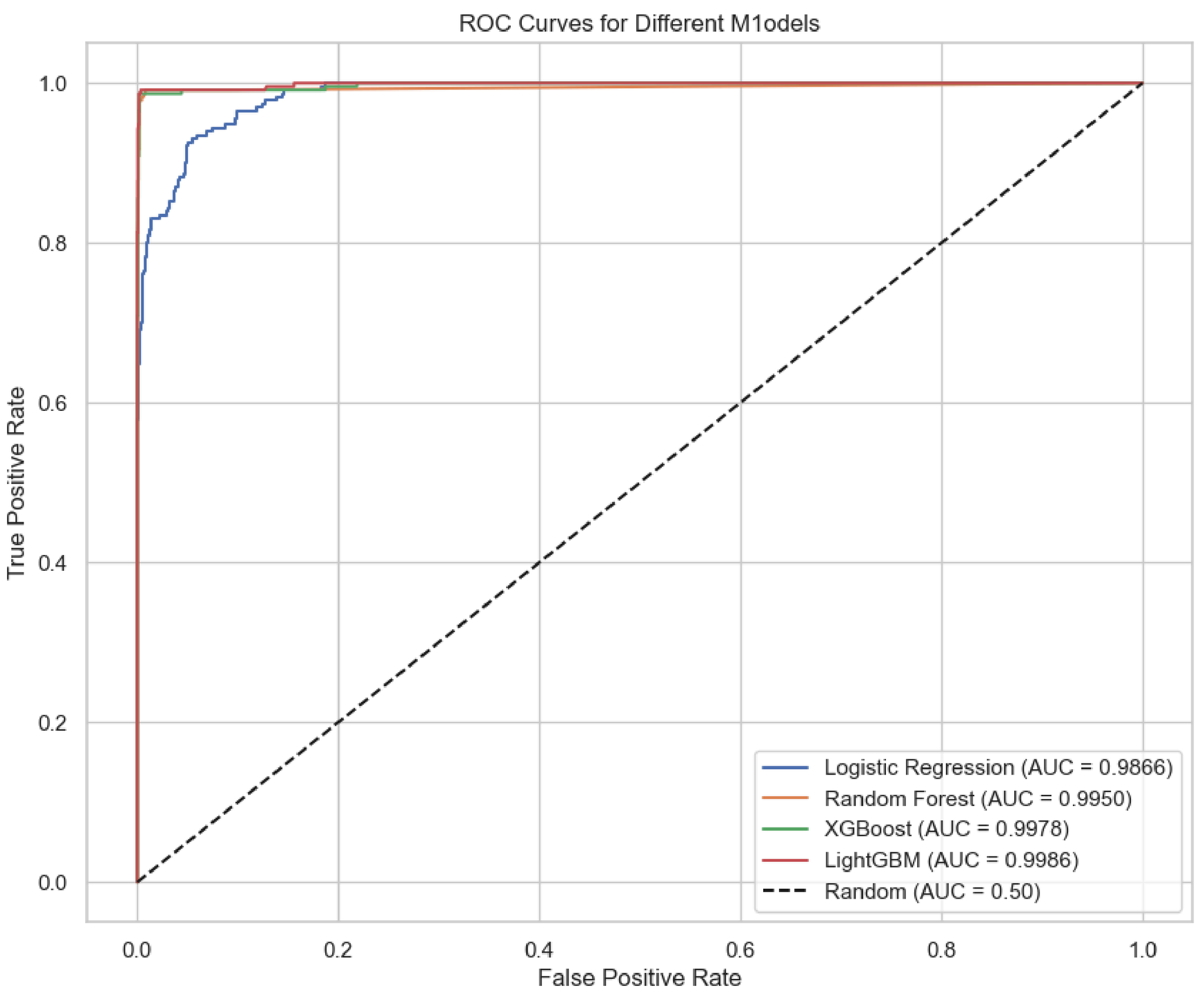
False Positives (4): Only 4 non-fraud transactions were incorrectly flagged as fraud — very low.

False Negatives (3): Just 3 fraud transactions were missed and predicted as non-fraud — also very low.

ROC AND AUC CURVES

--- ROC CURVES AND AUC SCORES --

- Logistic Regression AUC: 0.9866
- Random Forest AUC: 0.9950
- XGBoost AUC: 0.9978
- LightGBM AUC: 0.9986



KEY INSIGHTS FROM EDA

TRANSACTION TYPE PATTERNS

- Fraudulent transactions are predominantly associated with the **TRANSFER** and **CASH_OUT** transaction types.
- These types are commonly exploited for unauthorized fund movements.

RECEIVER'S BALANCE BEHAVIOR

- In many fraud cases, the receiver's balance remains unchanged after the transaction.
- This suggests that the receiver account may be fictitious or inactive, Aa fraudulent transfer.

BALANCE DISCREPANCY SIGNALS

- Large differences between the sender's old and new balance, especially when not reflected on the receiver's side, often signal fraud.
- These imbalances are red flags in fraudulent transactions.

TRANSACTION AMOUNT INSIGHTS

- Fraud cases are more frequently observed in medium to high-value transactions.
- Low-value transactions are less likely to be fraudulent, making amount size a significant risk indicator.

CONCLUSION

SUMMARY

This project focused on building a robust and accurate machine learning model to detect fraudulent financial transactions. After evaluating multiple algorithms—including Logistic Regression, Random Forest, LightGBM, and XGBoost—the XGBoost Classifier emerged as the best performer.

The selection of XGBoost led to a high-performance fraud detection system capable of:

- Accurately identifying complex fraud patterns
- Minimizing both customer inconvenience and financial loss
- Being deployed in real-time financial environments for transaction monitoring

--WHY XGBoost--

- **Handles Imbalanced Data Efficiently**
- **Regularization**
- **Speed and Scalability**
- **Feature Importance & Interpretability**
- **Consistent Accuracy Across Folds**

IMPLICATIONS OF THE WORK

The successful implementation of this model has real-world relevance:

Reduces Financial Loss

By identifying fraudulent activities early, it helps reduce losses.

Increases Security

Enhances transaction safety and builds trust among customers.

Deployable in Real-Time Systems

With high precision and recall, the model is ready for production environments in banking and fintech platforms.

Supports Risk Management

- Can assist compliance and audit teams in better fraud surveillance.

Future Work

Real-Time Deployment

Use Flask or FastAPI to serve the model for real-time transaction scoring and alert generation.

Dashboard Integration

Link model predictions with interactive dashboards (e.g., Power BI or Streamlit) for live monitoring.

Deep Learning Approaches

Explore RNNs, LSTMs, or Autoencoders to capture complex, time-based fraud patterns.

Adaptive Learning Pipelines

Develop a system that continuously learns from new data using scheduled model retraining and concept drift detection.

Boosting Productivity in Fraud Detection

Automated Data Pipelines

Use tools like Apache Airflow, MLflow, or Prefect to automate preprocessing, model training, and deployment workflows.

Real-Time Alerting System

Integrate fraud prediction APIs with notification systems (Slack, Email, SMS) for immediate fraud alerts

Model Optimization

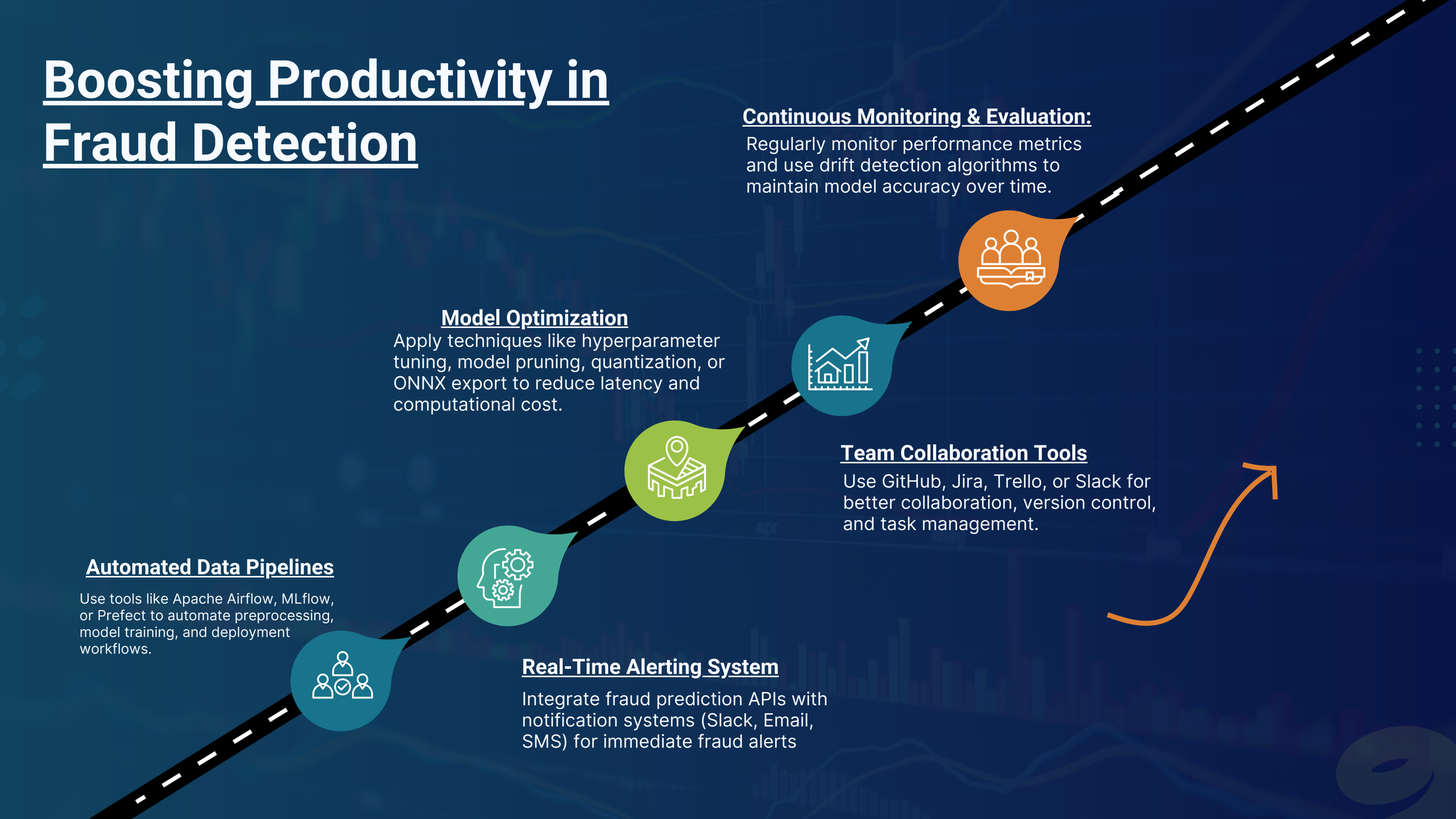
Apply techniques like hyperparameter tuning, model pruning, quantization, or ONNX export to reduce latency and computational cost.

Team Collaboration Tools

Use GitHub, Jira, Trello, or Slack for better collaboration, version control, and task management.

Continuous Monitoring & Evaluation:

Regularly monitor performance metrics and use drift detection algorithms to maintain model accuracy over time.



Questions ?

BIA | BOSTON INSTITUTE OF ANALYTICS ®

Thank You!