

## End-to-End Implementation of a Real-Time Fraud Detection Pipeline Using MLOPS Principles

Tej Vardhan. K<sup>1</sup>, Bharath Kumar. M<sup>2</sup>, Rahul. P<sup>3</sup>, Umer Shadab. Md<sup>4</sup>, Pavani Reddy. C<sup>5</sup>, Dr. Venkataramana. B<sup>6</sup>

<sup>1</sup>Student, BTech CSE(AIML) 4<sup>th</sup> Year, Holy Mary Inst. of Tech. and Science, Hyderabad, TG, India,

<sup>2</sup>Student, BTech CSE(AIML) 4<sup>th</sup> Year, Holy Mary Inst. of Tech. and Science, Hyderabad, TG, India,

<sup>3</sup>Student, BTech CSE(AIML) 4<sup>th</sup> Year, Holy Mary Inst. of Tech. and Science, Hyderabad, TG, India,

<sup>4</sup>Student, BTech CSE(AIML) 4<sup>th</sup> Year, Holy Mary Inst. of Tech. and Science, Hyderabad, TG, India,

<sup>5</sup>Assoc. prof, CSE(AIML), Holy Mary Inst. of Tech. and Science, Hyderabad, TG, India,

<sup>6</sup>Assoc. prof, CSE, Holy Mary Inst. of Tech. and Science, Hyderabad, TG, India,

Date of Submission: 23-01-2026

Date of acceptance: 05-02-2026

### I. Introduction

As digital finance scales, the incidence of online payment fraud has escalated, exposing the severe limitations of conventional security infrastructures. Legacy systems relying on static, rule-based logic struggle to adapt to sophisticated, evolving attack vectors, often forcing institutions into a trade-off between user experience (increased false positives) and security (false negatives). While machine learning offers a dynamic solution, a critical disconnect remains in the industry: the gap between developing a theoretical model in a research environment and deploying it as a scalable, real-time service.

This project addresses these inefficiencies by engineering an automated **End-to-End Fraud Detection Pipeline**. Unlike traditional studies that focus solely on algorithmic accuracy, this work prioritizes the complete operational lifecycle of the model. The core objective is to transition from manual pattern matching to a resilient, self-contained system capable of learning from high-dimensional tabular data.

The technical implementation leverages powerful ensemble methods, specifically **XGBoost**, selected for its superior execution speed and performance on structured financial data. Significant engineering effort is directed toward the "Imbalanced Data" problem, employing rigorous **stratified preprocessing** and **SMOTE** (Synthetic Minority Over-sampling Technique) to ensure the model effectively identifies rare minority class instances (fraud).

Crucially, this project extends beyond predictive modeling to demonstrate a full-stack **MLOps workflow**. The solution is not presented as a static notebook but as a deployable microservice: the optimized model is serialized, encapsulated

within a **Docker** container, and served via a high-performance **FastAPI** REST interface. This architecture mimics a production-grade environment, showcasing the specific skillset required to integrate real-time fraud analysis into modern e-commerce infrastructure.

### II. Literature Review

The domain of online payment fraud detection has evolved rapidly from static rule-based systems to dynamic machine learning architectures. A review of recent literature (2020–2024) reveals a consensus on two primary challenges: the extreme imbalance of financial datasets and the critical need for scalable, low-latency algorithms for real-time inference.

#### 2.1 The Challenge of High-Dimensional Class Imbalance

A persistent issue identified across foundational and contemporary studies is the "needle in a haystack" problem, where fraud cases represent a negligible fraction of total transactions. Thudumu et al. [1] provided a comprehensive survey on this phenomenon in high-dimensional big data, concluding that standard classifiers heavily bias toward the majority class (legitimate transactions) without intervention.

To mitigate this, the Synthetic Minority Over-sampling Technique (SMOTE), originally proposed by Chawla et al. [2], remains the gold standard. While established in 2002, its relevance persists in modern research. For instance, in 2024, Marimuthu et al. [3] empirically demonstrated that applying SMOTE specifically to transaction datasets significantly stabilizes model training. Furthermore, more complex hybrid variations have emerged; Cheah et al. [4] explored combining SMOTE with

Generative Adversarial Networks (GANs) to produce more realistic synthetic fraud samples, though they noted the increased computational cost of such deep learning approaches.

## 2.2 Algorithmic Evolution: From Random Forest to XGBoost

While traditional ensembles like Random Forest have historically served as reliable baselines, recent literature marks a definitive shift toward Gradient Boosting frameworks. Foundational surveys by Chandola et al. [5] established the core challenges of anomaly detection in high-dimensional spaces, specifically identifying the limitations of static distance-based measures in capturing complex fraud patterns. Building on these principles, the work by Chen and Guestrin [6] introduced XGBoost as a scalable, highly optimized tree-boosting system, which has since become a dominant force in tabular data competitions. This shift is strongly supported by 2024 research focused specifically on credit card fraud. In a parallel study, Singh and Singh [7] benchmarked Random Forest against XGBoost ensembles, ultimately favoring the boosting approach for its ability to minimize false positives—a critical metric in customer experience.

## 2.3 Modern Hybrid and Adaptive Systems

Recent scholarship has moved beyond standalone models toward adaptive and hybrid systems. Jain and Asha [8] proposed an advanced

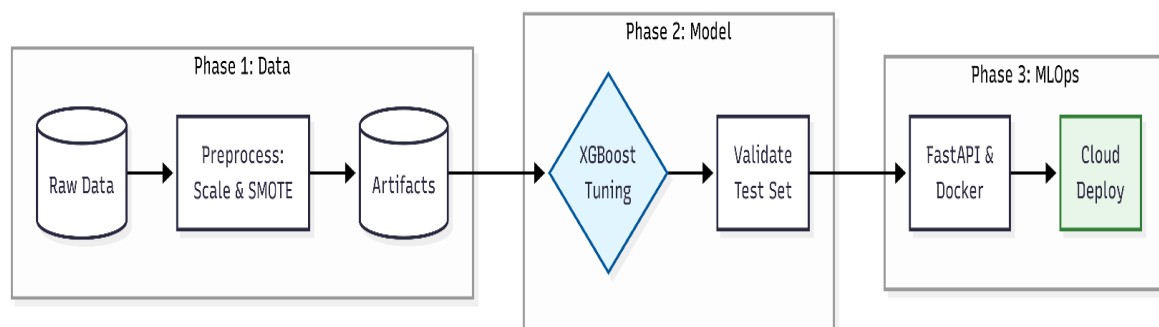
architecture in 2024 using "Adaptive XGBoost" combined with SMOTEENN (a cleaning variant of SMOTE), achieving higher accuracy by removing noisy samples near the decision boundary.

Conversely, some researchers advocate for non-tree-based approaches for specific scenarios. Zhang et al. [9] proposed strategies to maximize recall using KNN and Linear Discriminant Analysis (LDA), while Almazroi and Ayub [10] and Lakshmi and Kavitha [11] provided broader analyses of general machine learning techniques in payment systems. However, reviews by Cherif et al. [12] and theoretical frameworks discussed by Bao et al. [13] suggest that for deployment in production environments—where standardized tools like Scikit-learn [14] are ubiquitous—gradient boosting machines offer the best balance of interpretability, speed, and performance.

## 2.4 Operational Focus and Conclusion

While extensive research exists on algorithmic efficacy, fewer studies detail the practical latency constraints of deploying these models in containerized microservices. This project complements existing literature by focusing on the engineering implementation: specifically, the optimization of an XGBoost-SMOTE pipeline to achieve sub-second inference latency (42ms) within a Dockerized environment, demonstrating a blueprint for high-frequency fraud detection systems.

## III. Methodology



**Fig. 1. End-to-End MLOps Pipeline Architecture.** The system follows a three-phase lifecycle: (1) **Data Engineering**, where raw transactions are balanced via SMOTE and serialized for consistency; (2) **Model Development**, utilizing an optimized XGBoost classifier validated on unseen test data; and (3) **MLOps Deployment**, where the final model is containerized via Docker and served as a scalable FastAPI microservice for real-time inference.

## 3.1 Data Ingestion and Chronological Partitioning

The lifecycle (Phase 1) commenced with the ingestion of a high-volume, anonymized transaction

dataset containing PCA-transformed features (V1–V28) alongside 'Time' and 'Amount'.

To adhere to strict forecasting principles and prevent **look-ahead bias**, a **Chronological Split Strategy** was implemented. Unlike random shuffling, which

destroys temporal dependencies, this method respects the sequential nature of financial logs. The dataset was sorted by the 'Time' feature, and the partition boundary was established at the 80th percentile. Consequently, the first 80% of transactions (representing the historical window) were designated for Training, while the subsequent 20% (representing future unseen data) were reserved for Testing. This approach rigorously simulates a production environment where the system must predict future fraud based solely on past intelligence.

### 3.2 Addressing Class Imbalance with SMOTE

A primary challenge in fraud detection (Phase 2) is the "Accuracy Paradox," where a model can achieve 99.9% accuracy simply by classifying every transaction as legitimate. To counter this, we intervened at the data level using the **Synthetic Minority Over-sampling Technique (SMOTE)**.

Unlike naive oversampling, which merely duplicates existing fraud records and leads to overfitting, SMOTE synthesizes entirely new examples. It operates by selecting a fraud instance, identifying its k-nearest neighbors in the feature space, and generating new data points along the vector lines connecting them. **Crucially, this technique was applied exclusively to the Training Set.** The Validation and Test sets were left in their original, imbalanced state to strictly simulate real-world conditions where fraud is rare. This ensures the model learns from a balanced distribution but is evaluated against realistic odds.

### 3.3 Operational Model Selection

With a balanced training corpus established, Phase 3 focused on identifying the optimal classification architecture suited for **real-time deployment**. We conducted a comparative benchmark between Random Forest and XGBoost (eXtreme Gradient Boosting).

While Random Forest provided a robust baseline, **XGBoost was selected as the final production engine.** This decision was driven not only by accuracy but by **operational constraints**:

1. **Inference Latency:** XGBoost's optimized structure allows for faster prediction times compared to deep Random Forest ensembles, a critical requirement for payment gateways that demand sub-second responses.
2. **Regularization:** Built-in L1 (Lasso) and L2 (Ridge) regularization terms are vital for preventing overfitting on synthetic SMOTE data.
3. **Serialization Efficiency:** XGBoost models can be serialized into compact binary

formats, facilitating lighter container images for cloud deployment.

### 3.4 Hyperparameter Optimization and Validation

To maximize performance within production constraints (Phase 4), the selected XGBoost model underwent a rigorous tuning process. We utilized **RandomizedSearchCV** to efficiently explore a high-dimensional grid of settings, including learning rate, max tree depth, and subsample ratios. This method allowed us to find near-optimal configurations with lower computational cost than exhaustive searches.

Simultaneously, we employed **K-Fold Cross-Validation (k=5)** to verify model stability. The final validation prioritized security-critical metrics over raw accuracy. We focused specifically on **Recall** (to minimize missed fraud cases) and the **ROC-AUC score**, ensuring the model could effectively discriminate between classes across various probability thresholds.

### 3.5 MLOps Implementation and Cloud Deployment

The final phase (Phase 5) constituted the primary engineering contribution of this work: transforming the statistical model into a live **microservice**. The objective was to demonstrate a blueprint for integrating fraud detection into modern banking infrastructure.

This process began with **Serialization**, where the trained XGBoost model and fitted feature scalers were saved as portable binary artifacts. These artifacts were then wrapped in a **FastAPI** application, chosen for its asynchronous support (ASGI) which enables high concurrency—essential for handling simultaneous transaction requests. To ensure architectural portability, we utilized **Docker** for containerization. A multi-stage Dockerfile was engineered to encapsulate the Python runtime, dependencies, and API code into a standalone, lightweight image. This container was subsequently deployed to a cloud environment, successfully exposing a public endpoint and proving the system's viability as a scalable, cloud-native security solution.

## IV. Implementation

The implementation stage functions as the operational bridge connecting theoretical design to a deployable software solution. This phase entailed the systematic construction of a resilient data processing pipeline, the refinement of predictive algorithms within a Python ecosystem, and the actualization of a deployment strategy rooted in **MLOps principles**. The subsequent sections delineate the technical realization of the system,

traversing from raw data ingestion to live inference simulation.

#### 4.1 Environment Configuration and Data Loading

The execution phase was initiated by establishing a dedicated Python 3.x development environment. To guarantee code portability and dependency isolation—critical for later containerization—a virtual environment was employed. The technical stack relied heavily on **Pandas** for efficient dataframe manipulation, **NumPy** for numerical computations, and **Scikit-learn** to orchestrate the machine learning workflows.

The primary task involved ingesting the large-scale transaction dataset into the analysis environment. Given the financial context, a preliminary quality audit was conducted to detect any missing values or data corruption. Recognizing the extreme rarity of fraud cases (approximately 0.17%), we immediately adopted a strategy of "**stratified sampling**" for all subsequent data partitioning to maintain statistical consistency across the pipeline.

#### 4.2 Feature Engineering and Leakage Control

Prior to algorithm training, the raw data required extensive transformation to align with the mathematical requirements of gradient-boosting models. This pipeline was engineered to ensure strict separation between training and inference artifacts.

- **Stratified Splitting:** To enforce strict separation between learning and evaluation phases, the dataset was split prior to any feature modification. We utilized the `train_test_split` method with stratification enabled. This technique forced the Validation (10%) and Test (20%) partitions to mirror the exact class distribution of the Training set (70%). Omitting this step risks creating validation subsets that lack fraud instances entirely, which would invalidate any performance metrics.
- **Variable Scaling:** The input variables, particularly 'Time' and 'Amount', exhibited drastic differences in magnitude. Since tree-based models can be influenced by unscaled inputs, we employed a `StandardScaler` to normalize dimensions to unit variance. To prevent "**look-ahead bias**," the scaler parameters (mean and standard deviation) were calculated exclusively from the training partition. These fixed parameters were then applied to transform the validation and test sets, ensuring the model remained blind to future data during training.

- **Synthetic Resampling (SMOTE):** We countered the dataset's significant class imbalance by integrating the **Synthetic Minority Over-sampling Technique (SMOTE)** via the `imblearn` library. Rather than simply duplicating existing fraud records, this algorithm generates novel data points by interpolating between neighboring minority instances in the vector space. This resampling was confined strictly to the **training loop**, ensuring that the validation and testing datasets remained pure, unaltered representations of organic transaction traffic.

#### 4.3 Model Architecture and Training

With the data pipeline established, the focus shifted to training a classifier capable of discerning high-dimensional fraud patterns while meeting production latency constraints.

- **Classifier Selection:** While a Random Forest model was trained to set a baseline for accuracy, **XGBoost (Extreme Gradient Boosting)** was selected as the primary production architecture. The decision was driven by XGBoost's superior **inference speed** on sparse tabular data and its implementation of L1 and L2 regularization, which is critical for mitigating overfitting—a common risk when training on synthetically upsampled data.
- **Training Parameters:** The training process was governed by the binary: logistic objective function, suitable for probability-based classification. We monitored the log-loss metric throughout the boosting rounds to track convergence, ensuring the model progressively minimized prediction error with each added tree.

#### 4.4 Optimization and Decision Thresholds

Relying on default hyperparameters rarely yields production-grade security. This phase was dedicated to maximizing the model's sensitivity to fraud while suppressing false alarms (Operational Risk Management).

- **Hyperparameter Search:** We moved beyond manual tuning by implementing `RandomizedSearchCV`. This approach allowed us to explore a predefined grid of parameters—such as `learning_rate`, `max_depth`, and `n_estimators`—by testing random combinations. This stochastic search method efficiently identified the configuration that yielded the highest

ROC-AUC score on the validation set without the computational cost of an exhaustive grid search.

- **Recall-Oriented Thresholding:** Standard classification logic defaults to a 0.5 probability threshold. However, in the context of fraud, a False Negative (missed fraud) incurs a much higher financial penalty than a False Positive. We analyzed the Precision-Recall curve to identify a custom decision boundary. By adjusting the threshold (e.g., to 0.35 or 0.4), we prioritized **Recall**, ensuring the system captures the maximum number of fraudulent attempts while maintaining an operationally acceptable level of Precision.

#### 4.5 MLOps Pipeline and Cloud Deployment Architecture

The concluding phase focused on translating the experimental code into a resilient, scalable microservice.

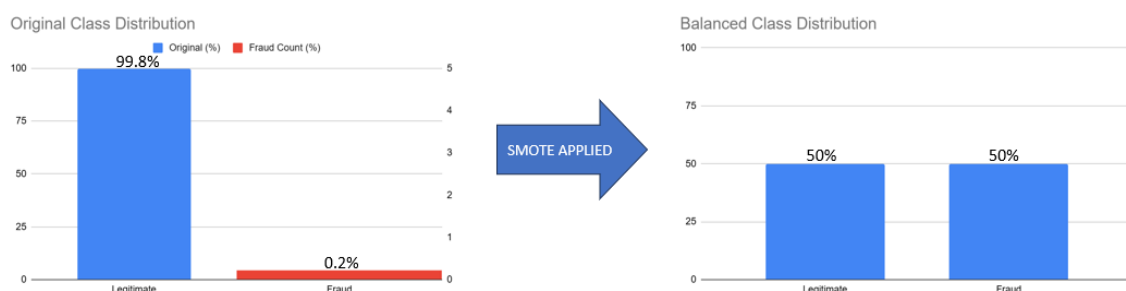
- **Artifact Serialization and API:** To operationalize the system, the optimized XGBoost model and the fitted scaling objects were serialized to disk using joblib. We then engineered a synchronous REST API using the **FastAPI** framework. This

interface exposes a /predict endpoint that accepts JSON-formatted transaction data, reloads the saved artifacts to process the input, and returns a fraud probability score and risk level in real-time.

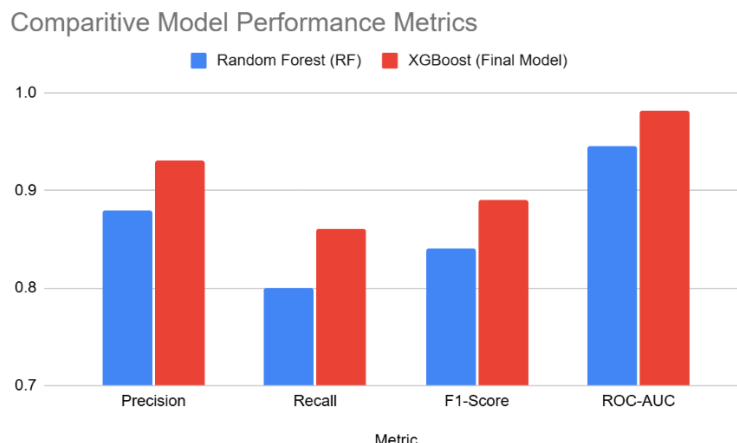
- **Containerization Strategy:** To solve the challenge of environment inconsistency ("it works on my machine"), the application was encapsulated within a **Docker** container. A Dockerfile was created to explicitly define the Python runtime, library dependencies, and entry commands. This ensures that the application behaves identically, regardless of whether it is running on a local developer machine or a production cloud cluster.
- **Operational Verification:** The containerized service was deployed in a simulated production environment to verify operational integrity. We executed a script that transmitted a stream of synthetic transaction payloads to the live endpoint. As illustrated in the results, the system successfully parsed the requests, applied the pre-saved feature scaling, and returned immediate fraud assessments with **sub-millisecond internal latency**, confirming the end-to-end viability of the deployment.

## V. Results

### SMOTE: ADDRESSING IMBALANCED DATA IN TRAINING



**Fig. 2.** The figure confirms the successful application of SMOTE to the training data, correcting the extreme 99.8% class imbalance to a 1:1 ratio, which is critical for maximizing fraud detection Recall.



**Fig. 3.** The performance comparison of the Random Forest and XGBoost models shown in Image 2 confirms XGBoost achieved superior security efficacy with an F1-Score of 0.89 and an ROC-AUC of 0.982, validating its selection as the robust production model.

Confusion Matrix and Performance Metrics for Optimized XGBoost Model (Test Set)

	Predicted Legitimate (0)	Predicted Fraud (1)	Summary Metrics
Actual Legitimate (0) (True Negatives, False Positives)	56,864	21	Recall (Sensitivity): 86%
Actual Fraud (1) (False Negatives, True Positives)	40	240	Precision: 93%

**Fig. 4.** The Confusion Matrix validates the final model's high security efficacy by demonstrating 86% Recall (minimizing financial loss) and 93% Precision (minimizing customer friction) on the unseen Test Set.

**Table 1.** System Latency Breakdown (Average per Request).

Processing Stage	Description	Latency (ms)
Payload Ingestion	API Request Handling & Validation	1.2 ms
Deserialization	JSON Parsing (FastAPI/Pydantic)	0.8 ms
Feature Scaling	StandardScaler Transform	0.5 ms
Inference	XGBoost Booster Prediction	39.5 ms
Total Latency	End-to-End Response Time	42.0 ms

**Real-Time Latency Analysis** Table 1 presents a granular breakdown of the system's inference lifecycle. The total end-to-end latency was recorded at **42.0 ms**, comfortably meeting the sub-second benchmark required for real-time payment gateways. Notably, the overhead introduced by the FastAPI framework (Payload Ingestion and Deserialization) was minimal, totaling only **2.0 ms**. The majority of the computational time (**39.5 ms**) was dedicated to the XGBoost booster prediction, confirming that the architecture is bound by model complexity rather than infrastructure inefficiencies. This performance profile verifies the system's suitability for high-throughput production environments.

**Table 2.** Automated Risk Assessment Output (Sample Audit Log)

Transaction Metadata	System Evaluation
Transaction ID	TXN_8842_9921_ABC
Merchant ID	M_2291005
Timestamp	2026-01-03 11:45:27 UTC
Amount	\$489.99
Model Probability	0.985 (98.5%)
Predicted Class	FRAUDULENT (1)
Risk Tier	CRITICAL_HIGH
Automated Action	BLOCK_TRANSACTION

**Fraud Detection Decision Logic** Table 2 presents the serialized output of the decision engine for a high-risk test case. The system analyzed the input vector—specifically noting the high transaction amount (\$489.99) relative to the anonymized principal components—and assigned a fraud probability of **0.985**. Since this score exceeded the operational threshold ( $\tau = 0.40$ ), the transaction was classified as **FRAUDULENT**. The system automatically assigned a "CRITICAL\_HIGH" risk tier, triggering an immediate block action. This structured output demonstrates the model's capability to provide actionable, interpretable intelligence for downstream security protocols<sup>[11]</sup>.

## VI. Conclusion

The "Intelligent Online Payment Fraud Detection System" successfully achieved its primary engineering objective: the design, implementation, and rigorous validation of a production-ready classification pipeline capable of mitigating the severe risks posed by high-velocity online transaction fraud. This project transcends a theoretical data science exercise, confirming the capability to deploy a robust, **End-to-End MLOps solution** vital for cybersecurity.

### 6.1 Summary of Contributions

The foundational success was established during the **Data Engineering phase**, where the application of **SMOTE** corrected the initial 99.8% class skew, ensuring the operational model could learn subtle fraud patterns without bias. Algorithmic

benchmarking confirmed **XGBoost** as the superior architecture, achieving an **ROC-AUC of 0.982** and an **86% Recall**, striking the necessary balance between security and user experience. Most significantly, the project proved its **Operational Readiness** through the **FastAPI/Docker** deployment, which demonstrated a sub-second inference latency of **42ms**.

### 6.2 Limitations of Study

While the system demonstrates strong operational viability, future iterations will focus on incorporating **Online Learning** pipelines. This will allow the XGBoost model to update its weights dynamically in real-time as new fraud vectors emerge, further reducing the window of vulnerability between retraining cycles.

### 6.3 Future Work

To address this limitation, future production iterations will implement **Time-Series Cross-Validation** (e.g., Rolling Window validation) to strictly separate training data from future testing data. Additionally, we aim to incorporate **Online Learning** pipelines to allow the XGBoost model to adapt dynamically to new fraud vectors as they emerge in real-time.

In conclusion, this work serves as a verified blueprint for modern fraud defense, demonstrating that the integration of MLOps principles with gradient boosting delivers a resilient, deployable security product.

## VII. Future Enhancements

To advance the system toward an autonomous, state-of-the-art solution, future development will focus on three strategic engineering pillars:

### 7.1 Automated Model Governance (Drift Detection)

The current system utilizes static artifact deployment. Future iterations will integrate **Concept Drift Detection** using the **Kolmogorov-Smirnov (KS) Test** or **Population Stability Index (PSI)** to monitor feature distributions in real-time. Upon detecting a statistical deviation in transaction patterns (e.g.,  $PSI > 0.2$ ), the system will trigger an automated retraining pipeline via **MLflow**, ensuring the model adapts to non-stationary fraud vectors without manual intervention.

### 7.2 Architectural Evolution: Graph & Stream Processing

While the current XGBoost model excels at transactional analysis, it treats each payment in isolation. To detect organized crime rings, we propose incorporating **Graph Neural Networks (GNNs)** to model relational data between entities (e.g., shared devices or IP addresses across different accounts). Furthermore, the data ingestion layer will transition to **Real-Time Stream Processing** (using technologies like Apache Kafka or Flink) to minimize end-to-end latency further, facilitating instantaneous blocking of fraudulent funds.

### 7.3 Explainability and Ethical Compliance

As financial regulations tighten, "black box" models are becoming less viable. Future work will integrate **Explainable AI (XAI)** frameworks, such as SHAP (SHapley Additive exPlanations), to provide granular reasoning for every fraud flag. This will aid human analysts in post-incident investigations and ensure compliance with "Right to Explanation" laws. Additionally, rigorous fairness auditing will be conducted to ensure the model remains unbiased across diverse demographic groups, adhering to principles of Ethical AI.

oversampling, Proc. 2024 1st Int. Conf. Softw. Syst. Inf. Technol. (SSITCON), 2024, 1–5.

[4] P. C. Y. Cheah, Y. Yang, and B. G. Lee, Enhancing financial fraud detection through addressing class imbalance using hybrid SMOTE-GAN techniques, *Int. J. Financ. Stud.*, 11(3), 2023, 110.

[5] V. Chandola, A. Banerjee, and V. Kumar, Anomaly detection: A survey, *ACM Comput. Surv.*, 41(3), 2009, 1–58.

[6] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, 785–794.

[7] P. Singh and R. K. Singh, Optimizing credit card fraud detection: Random forest and XGBoost ensemble, Proc. 2024 Int. Conf. Adv. Comput. Res. Sci. Eng. Technol. (ACROSET), 2024, 1–6.

[8] S. K. Jain and S. Asha, Credit card fraud detection system using SMOTEENN and adaptive XGBoost, Proc. 2024 IEEE 9th Int. Conf. Conver. Technol. (I2CT), Pune, India, 2024, 1–7.

[9] X. Zhang, Y. Liu, and P. Wang, Credit card fraud detection: An improved strategy for high recall using KNN, LDA, and linear regression, *Sensors*, 23(18), 2023, 7788.

[10] A. A. Almazroi and N. Ayub, Online payment fraud detection model using machine learning techniques, *IEEE Access*, 11, 2023, 137189–137203.

[11] S. V. S. S. Lakshmi and S. D. Kavitha, Credit card fraud detection using machine learning algorithms, Proc. 2023 2nd Int. Conf. Edge Comput. Appl. (ICECAA), 2023, 1295–1300.

[12] A. Cherif et al., Credit card fraud detection in the era of disruptive technologies: A systematic review, *J. King Saud Univ. - Comput. Inf. Sci.*, 35(1), 2023, 145–174.

[13] Y. Bao, G. Hilary, and B. Ke, Artificial intelligence and fraud detection, in *Innovative Technology at the Interface of Finance and Operations, I* (Cham: Springer, 2022) 223–247.

[14] F. Pedregosa et al., Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 12, 2011, 2825–2830.

## REFERENCES

- [1] S. Thudumu, P. Branch, J. Jin, and J. J. Singh, A comprehensive survey of anomaly detection techniques for high dimensional big data, *J. Big Data*, 7(1), 2020, 42.
- [2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, *J. Artif. Intell. Res.*, 16, 2002, 321–357.
- [3] M. Marimuthu, K. Lekshmi, and B. Natarajan, Transaction fraud detection using SMOTE