## A WEB-BASED APPLICATION FOR MONITORING AND IDENTIFYING PATTERNS IN FINANCIAL TRANSACTIONS USING

## A HYBRID ALGORITHM APPROACH

Undergraduate Thesis

Submitted to the Faculty of the

Department of Computer Studies

Cavite State University - Imus Campus

City of Imus, Cavite

In partial fulfillment

of the requirements for the degree

Bachelor of Science in Computer Science

**EUGENE ALLEN P. MATRO**

**EUGENE JOSE P. MATRO**

**ARCHIE T. SARAGENA**

April 2025

**PROPOSAL APPROVAL SHEET**

Author(s): **EUGENE ALLEN P. MATRO**

**EUGENE JOSE P. MATRO**

**ARCHIE T. SARAGENA**

Title of the Study: **A WEB-BASED APPLICATION FOR** **MONITORING**

**AND IDENTIFYING PATTERNS IN FINANCIAL**

**TRANSACTIONS USING A HYBRID**

**ALGORITHM APPROACH**

Degree or Course: **BACHELOR OF SCIENCE IN COMPUTER SCIENCE**

**A P P R O V E D:**

**ROSALINA LACUESTA, MIT, LPT** \_\_\_\_\_\_\_ **RAQUEL ANN OLIVO** \_\_\_\_\_\_\_

Adviser Date Technical Critic Date

**GRACE S. IBAÑEZ, MSCS** \_\_\_\_\_\_\_ **LIANE VINA G. OCAMPO, PhD** \_\_\_\_\_\_

Department Chairperson Date Campus Research Coordinator Date

**ARMI GRACE B. DESINGAÑO, PhD** \_\_\_\_\_\_

Campus Administrator Date

## A WEB-BASED APPLICATION FOR IDENTIFYING PATTERNS IN FINANCIAL TRANSACTIONS USING HYBRID ALGORITHM APPROACHES

Matro, Eugene Allen P., Matro, Eugene Jose P., Saragena, Archie T.

April 2025

Cavite State University – Imus Campus

City of Imus, Cavite

**ABSTRACT**

Fraudulent financial transactions pose a major challenge in today’s digital landscape, highlighting the need for advanced detection mechanisms. This study introduces a web-based application designed to analyze real-time transaction data and identify potential fraud patterns. By leveraging Isolation Forest and Autoencoder-based neural networks, the system detects anomalies by continuously monitoring transaction streams. The approach involves several key steps: data preprocessing, training anomaly detection models, and integrating these models into a web-based platform for real-world application. This method has strong potential for accurately detecting fraudulent activities. Moving forward, efforts will focus on optimizing model performance, minimizing false positives, and improving scalability to handle large transaction volumes efficiently. This research contributes to the evolving field of fraud detection by offering a data-driven, adaptive, and accessible solution for financial institutions and online platforms.

**INTRODUCTION**

**Project Context**

The increasing reliance on digital financial transactions has led to a parallel surge in fraudulent activities that threaten individual users and financial institutions alike. In the Philippines, the growing use of e-wallets, online banking, and digital payment platforms has made the need for secure and intelligent fraud detection more urgent. Fraudulent activities such as unauthorized withdrawals, stolen identities, and fake transaction patterns are increasingly sophisticated, often eluding traditional rule-based detection systems (Adusumilli et al., 2020; Pourhabibi et al., 2020).

While many institutions utilize commercial fraud detection systems, these are often expensive, lack transparency in how decisions are made, and are not easily customizable for specific regional financial patterns (Obeng et al., 2024). Furthermore, publicly available fraud detection tools rarely integrate a mix of supervised and unsupervised learning techniques in a hybrid design that can adapt to both known and emerging fraud behaviors (Ashtiani & Raahemi, 2021).

This project addresses these gaps by developing a web-based application that identifies suspicious financial transactions using a hybrid of machine learning algorithms—two supervised (K-Nearest Neighbor and XGBoost) and two unsupervised (Isolation Forest and Local Outlier Factor). Supported by previous studies that emphasize the value of combining ensemble and anomaly-based techniques (Zhou et al., 2021; Adusumilli et al., 2020), the system aims to provide a cost-effective, scalable, and locally deployable solution that enhances fraud detection within Philippine financial ecosystems.

**Objectives of the Study**

To develop a web-based application for detecting fraudulent financial transactions by identifying anomalous patterns using a hybrid machine learning approach consisting of K-Nearest Neighbor, XGBoost, Isolation Forest, and Local Outlier Factor algorithms.

Specific Objectives:

1. To design the system architecture and interface that enables users to input transaction data, receive fraud classification results, and view patterns or alerts.
2. To develop the system using appropriate software and hardware tools including Python (for algorithm implementation), Flask or FastAPI (for backend development), and visualization libraries like Plotly and Matplotlib; the system will also utilize data processing libraries such as Pandas, NumPy, and scikit-learn.
3. To test the hybrid algorithm model using public datasets (Kaggle credit card datasets), ensuring the system can detect both known and novel fraudulent behaviors.
4. To evaluate the performance of the system in terms of precision, recall, accuracy, and false positive rates using metrics established in related studies.
5. To prepare an implementation plan for deploying the system in a simulated real-world environment, including options for notifications (email or SMS alerts) and integration with financial data APIs.

**Purpose and Description**

The purpose of this project is to provide an intelligent, flexible, and low-cost fraud detection platform that financial institutions and users in the Philippines can use to identify suspicious transactions. Unlike traditional systems that rely solely on fixed rules or historical labels, the proposed system incorporates both supervised and unsupervised algorithms to analyze transactional data dynamically. K-Nearest Neighbor and XGBoost are used to classify transaction data based on historical patterns, while Isolation Forest and Local Outlier Factor provide anomaly detection capabilities that help uncover unknown or evolving fraud tactics.

The application is web-based to maximize accessibility and is built using open-source technologies to support maintainability and scalability. It is designed not only to classify whether a transaction is fraudulent but also to provide visual insights into anomaly scores and behavior patterns. By using hybrid detection techniques supported by findings from recent fraud detection literature, the system aims to improve detection accuracy, reduce false positives, and assist users in making informed decisions about flagged transactions.

**Time and Place of the Study**

This thesis presents a fraud detection framework for the analysis of financial transaction data stream. The study will be held from March 2024 to June 2025 at Cavite State University – Imus Campus.

**Scope and Limitation of the Study**

In this study, a data analysis system for detecting specific financial fraud such as credit card fraud and online banking fraud will be developed and analyzed. The analysis will be confined to publicly available and simulated financial datasets, integrating machine learning algorithms and stream processing platforms, specifically using free and accessible software which can be used in the Philippines. Additionally, the nature of the data used to conduct this analysis may have practical limitations. Reason being that the availability and quality of the real-time data used to indicate financial transactions can be affected, as can the complexity of the fraud patterns.

**Theoretical Framework of the Study**

Based on the Fraud Triangle Theory, the three key components—opportunity, pressure, and rationalization—play a pivotal role in driving fraudulent behavior (Embroker Team, 2024). This study aligns with the theory by focusing on identifying and mitigating opportunities for fraud through monitoring of transaction streams. The Fraud Triangle Theory underscores how the interaction of these components fosters fraudulent activities, with this research specifically addressing the "opportunity" factor. By utilizing data analysis for fraud detection, opportunities for fraud are minimized, thereby enhancing overall financial security.

Additionally, the integration of the Complex Event Processing (CEP) Framework strengthens this research by enabling the identification of patterns and correlations within data streams (Alaghbari, K.A., Saad, M.H.M., Hussain, A., et al., 2022). The CEP Framework is crucial for processing data, detecting anomalies, and recognizing deviations from expected behaviors. By providing actionable insights into potentially fraudulent activities, the framework facilitates proactive fraud prevention and establishes a robust system for safeguarding financial transactions.

**Conceptual Framework of the Study**

*Figure 1. Framework shown in the Input-Process-Output Format*

**Definition of Terms**

1. **K-Nearest Neighbor (KNN)** - is a supervised machine learning algorithm that classifies data points based on the majority class of their closest neighbors. In fraud detection, KNN can identify suspicious transactions by comparing them to known patterns in labeled historical data.
2. **XGBoost (Extreme Gradient Boosting)** - is a high-performance ensemble learning algorithm that builds multiple decision trees sequentially to improve prediction accuracy. It is widely used in fraud detection due to its robustness, speed, and ability to handle imbalanced datasets.
3. **Isolation Forest** - is an unsupervised anomaly detection algorithm that isolates outliers by randomly selecting features and splitting values. Its efficiency in detecting rare and unusual behaviors makes it suitable for identifying fraudulent financial transactions.
4. **Local Outlier Factor (LOF)** - is an unsupervised machine learning technique that detects anomalies by measuring the local density deviation of a given data point with respect to its neighbors. It is effective in identifying contextual fraud in complex transaction patterns.
5. **Anomaly Score** - is a numerical value assigned by an anomaly detection algorithm that quantifies how much a transaction deviates from normal behavior. In fraud detection, higher anomaly scores typically indicate greater likelihood of fraudulent activity and are used to flag transactions for further review.

**REVIEW OF RELATED LITERATURES AND STUDIES**

**Comparative Analysis of Machine Learning Models**

This study examines how machine learning and data mining techniques are being used to detect financial fraud, with a focus on models like decision trees, logistic regression, support vector machines (SVMs), and k-nearest neighbors (kNN). These approaches are popular because they’re easy to interpret and work well with structured financial data. One of the key insights from the literature is that ensemble methods, particularly random forests and gradient boosting machines (GBM), are especially effective at improving the accuracy of fraud detection. These techniques are adept at recognizing intricate patterns and reducing the risk of overfitting. The study also emphasizes the benefits of hybrid models that combine ensemble approaches with traditional classifiers, which tend to perform better on imbalanced datasets. To tackle the frequent issue of class imbalance in fraud detection, many researchers turn to the Synthetic Minority Oversampling Technique (SMOTE). This method helps by generating synthetic examples of the less common, fraudulent cases, thereby boosting the model’s ability to spot rare incidents. The study also explores the use of anomaly detection methods, including Isolation Forests and Autoencoders, which are valuable in situations where labeled data is limited and early fraud detection is crucial. Metrics such as accuracy, precision, recall, and AUC are commonly used to measure effectiveness (M. N. Ashtiani and B. Raahemi, 2021).

**Evaluating Machine Learning Algorithms for Financial Fraud**

This study assesses the effectiveness of machine learning algorithms, including decision trees, SVMs, neural networks and ensemble methods, in detecting anomalies in financial transactions. The researchers highlighted hybrid approaches of combining decision trees with ensemble methods, specifically SVMs and random forests or decision tress, which shows better accuracy and robustness, and could handle complex and imbalanced datasets. The researchers tackle the issue of class imbalance by utilizing the Synthetic Minority Oversampling Technique (SMOTE), which enhances model training by generating synthetic samples for underrepresented classes. The results indicate that ensemble methods, such as gradient boosting, strike an optimal balance between computational efficiency and detection accuracy while other models, such as random forest, SVMs and decision tree, fall slightly behind (Adusumilli, S. B.-V., Damancharla, H., Metta, A. R., 2020).

**Machine Learning Algorithm Implementation for Fraud Detection**

This study evaluates the effectiveness of machine learning algorithms—Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and Logistic Regression (LR)—in detecting credit card fraud, focusing on four specific fraud patterns: transactions with risky Merchant Category Codes (MCC), high-value transactions exceeding $100, transactions with risky ISO Response codes, and transactions from unknown web addresses. The dataset used was sourced from a financial institution under a confidential disclosure agreement and was divided into these four fraud categories. To handle imbalanced data, common in fraud detection due to the rarity of fraudulent events, the researchers employed resampling techniques such as Synthetic Minority Oversampling Technique (SMOTE) for over-sampling and random under-sampling (RUS) for balancing class distribution. Data preprocessing involved cleaning, transformation, integration, and reduction, including Principal Component Analysis (PCA) for dimensionality reduction to eliminate irrelevant features and prevent overfitting.

The experimental results revealed varying performance across the selected algorithms depending on the fraud type. Logistic Regression (LR) performed best for detecting fraud related to risky MCC and ISO Response codes, achieving accuracy rates of 74% and 72%, respectively. Naive Bayes (NB) excelled in identifying fraud from unknown web addresses with an 83% accuracy rate. Support Vector Machines (SVM) outperformed other models in detecting high-value fraudulent transactions above $100, reaching a notable 91% accuracy. The study also emphasized real-time fraud detection by implementing predictive analytics through an API module that interfaces with fraud detection models and a data warehouse. This system enables immediate classification of transactions and alerts banks in real time via a graphical user interface (GUI). Detected frauds are categorized into their respective types and stored for further analysis, allowing fraud monitoring teams to respond promptly (Thennakoon, A., et. al., 2019).

The researchers concluded that resampling techniques significantly improve classifier performance when dealing with skewed distributions. While the implemented models demonstrated promising accuracy levels, there is room for improvement, particularly in enhancing prediction reliability and expanding the framework to include location-based fraud detection in future work.

**Credit Card Fraud Detection Using Random Forest and Adaboost**

This study focuses on detecting credit card fraud using two machine learning algorithms: Random Forest and AdaBoost. The researchers utilized the Kaggle credit card fraud dataset, which contains 284,807 transactions (only 0.172% fraudulent), with PCA-transformed numerical features except for 'Time' and 'Amount'. The Random Forest algorithm was implemented by constructing multiple decision trees on randomly sampled data, using information gain for node splitting, and aggregating results via majority voting. For AdaBoost, sequential decision trees were built, with weights adjusted to prioritize misclassified transactions, enhancing detection of minority fraud cases. Both models were evaluated on accuracy, precision, recall, and F1-score, with the Random Forest achieving 99.9% accuracy and outperforming AdaBoost in precision (0.95 vs. 0.78) and recall (0.77 vs. 0.64) for fraud cases.

The study also highlighted the confusion matrix and ROC-AUC curves for performance analysis. Random Forest showed a lower false positive rate (37 vs. 65 for AdaBoost) and higher true positive rate (125 vs. 97), indicating better fraud identification. Despite AdaBoost’s adaptive boosting strength, its sensitivity to noisy data and outliers reduced its effectiveness compared to Random Forest’s robustness. The researchers concluded that Random Forest is superior for imbalanced datasets, though both algorithms shared similar overall accuracy (99.9%). Future work suggested exploring deep learning for further improvements in fraud detection (Sailusha, R., et. al., 2020).

The methodology emphasized practical implementation, including data splitting (training/testing), model training, and real-time deployment. The results align with prior research affirming Random Forest’s efficacy in fraud detection, while also addressing AdaBoost’s limitations in handling skewed data. This study provides a clear framework for selecting algorithms based on precision-recall trade-offs and operational requirements in financial fraud systems.

**Hyperparameter Optimization and Ensemble Learning for Fraud Detection**

This study proposes a machine learning-based approach for detecting credit card fraud, focusing on handling the class imbalance problem commonly found in real-world banking data. The researchers employed Bayesian optimization to tune hyperparameters, particularly emphasizing class weight adjustment as a preprocessing step to address data imbalance. Instead of using traditional resampling techniques like SMOTE, which can lead to increased false-positive rates, the proposed method assigns higher weights to the minority (fraud) class during model training. The experiments were conducted on a publicly available dataset containing 284,807 transactions, with only 492 labeled as fraudulent, making it highly imbalanced. Feature selection was performed using the Information Gain (IG) method, and the top six features were selected for model training. The framework evaluated several algorithms, including Logistic Regression, LightGBM, XGBoost, CatBoost, and Majority Voting. Among the individual models, LightGBM and XGBoost showed the best performance in terms of ROC-AUC (0.95), precision (0.79), recall (0.80), F1-score (0.79), and Matthews Correlation Coefficient (MCC) (0.79). These gradient boosting methods demonstrated high efficiency in handling large-scale and imbalanced data due to their built-in regularization and tree-based learning strategies. Additionally, an ensemble approach using majority voting was applied to combine predictions from CatBoost, XGBoost, and LightGBM, further improving model robustness (S. K. Hashemi, et. al., 2022).

**Analysis of Machine Learning Approaches to Credit Card Fraud Detection**

This study reviews various machine learning methods used in credit card fraud detection, emphasizing the challenges associated with imbalanced datasets and the effectiveness of different classification algorithms. The authors discuss commonly used supervised learning techniques such as Random Forest, Decision Trees, Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Gradient Boosting Machines (GBM). These models are evaluated on a publicly available dataset from the ULB Machine Learning Group, which contains 284,807 European credit card transactions from 2013, with only 492 labeled as fraudulent. Due to the high imbalance (fraudulent cases make up just 0.17%), traditional accuracy metrics are insufficient, so the evaluation focuses on precision, recall, F1-score, and false positive rate (FPR) to better assess model performance in detecting rare fraudulent events.

The paper highlights that ensemble methods like Random Forest and Gradient Boosting perform well in this domain due to their ability to handle non-linear patterns and reduce overfitting through multiple decision trees and voting mechanisms. Random Forest achieved the highest accuracy at 95.988%, followed by GBM at 93.99%, SVM at 93.228%, and Logistic Regression at 92.89%. The authors also explore the use of feedback mechanisms to enhance classifier performance, suggesting that iterative learning and model updates can improve fraud detection rates over time. Additionally, feature importance analysis was conducted using Principal Component Analysis (PCA), where features V1–V28 were derived from PCA transformations, along with 'Time' and 'Class' labels indicating normal or fraudulent transactions.

To implement and evaluate these models, the dataset was split into a 70:30 ratio for training and testing, respectively. Python-based machine learning libraries such as Scikit-learn and TensorFlow were used for model training and validation. The experimental results confirm that tree-based models, particularly Random Forest and Gradient Boosting, outperform other classifiers in terms of precision and recall (Simaiya, Sarita & Kumar Lilhore, Dr & Sharma, Sanjeev. (2020). The study concludes that while no single model is universally superior, selecting the appropriate algorithm depends on the trade-off between detection accuracy and computational efficiency. Future work includes applying these models to real-time data and exploring hybrid approaches, such as integrating generative adversarial networks (GANs) for synthetic fraud data generation to further improve detection capabilities.

**Ensemble Methods, Resampling Techniques, and Performance Evaluation**

This study presents a two-stage approach for credit card fraud detection using machine learning techniques, specifically addressing the challenges posed by imbalanced datasets. The first stage involves evaluating nine different classification algorithms—Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), Random Forest (RF), Gradient Boosting Machine (GBM), LightGBM, XGBoost, and CatBoost—on a real-world dataset containing 284,807 transactions, of which only 492 are fraudulent. These models are trained using default hyperparameters (except for KNN, where *n\_neighbors* is set to 3) and evaluated using stratified 5-fold cross-validation to prevent data leakage and ensure robust performance assessment. The top three performing models from this stage proceed to the second stage, where they are combined with 19 resampling techniques—including 11 undersampling, 6 oversampling, and 2 hybrid methods—to improve fraud detection in highly skewed data.

In the second stage, the selected algorithms are applied in conjunction with resampling techniques such as SMOTE, Borderline SMOTE, and Random Under-Sampling (RUS) to mitigate class imbalance. Evaluation metrics include Accuracy, Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC). Among all combinations tested, Random Forest (RF) consistently outperformed other classifiers, achieving an accuracy of 99.96%, recall of 81.63%, and precision of 96.38%. This highlights RF’s effectiveness in handling imbalanced financial data due to its ensemble nature and ability to reduce overfitting. Additionally, the authors emphasize that while AUC is commonly used, it may not be reliable for highly imbalanced datasets, advocating instead for MCC and F1-score as more appropriate performance indicators (Alfaiz, N. S., & Fati, S. M., 2022)

The framework also incorporates feature selection and stratified k-fold cross-validation to enhance model generalization and reduce bias. Feature importance analysis was performed, although specific features were not detailed due to data confidentiality. The system's modular design allows flexibility in integrating new resampling techniques or classifiers for future improvements. Experimental results validate the effectiveness of combining ensemble methods with resampling strategies for fraud detection, showing significant improvement over traditional approaches. The source code and methodology are made publicly available to encourage reproducibility and further research in this domain.

**Data Preprocessing & Classification Methods for E-commerce Fraud Detection**

This study presents a machine learning-based approach for detecting credit card fraud, focusing on handling the challenges posed by imbalanced datasets and evaluating the performance of different classification algorithms. The dataset used in the experiments was sourced from Kaggle and consists of 151,112 e-commerce transactions, with 14,151 labeled as fraudulent, resulting in a class imbalance ratio of approximately 9.3%. To address this imbalance, the researchers applied resampling techniques, including oversampling and undersampling methods. Feature selection was performed using Principal Component Analysis (PCA) to reduce dimensionality and eliminate redundant features that could negatively impact model performance. The data was then split into training and testing sets in a 70:30 ratio to ensure reliable evaluation of model accuracy and generalization.

Four supervised machine learning algorithms—Logistic Regression, Naïve Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—were implemented and evaluated using stratified k-fold cross-validation to minimize bias and variance in performance estimation. The confusion matrix was used as a key evaluation tool, with metrics such as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) calculated for each model. Performance was further assessed using accuracy, precision, recall, and F1-score. Among the tested models, Logistic Regression achieved the highest accuracy at 98.5%, followed closely by SVM. Naïve Bayes showed slightly lower performance, while KNN had the lowest accuracy among the four. These results indicate that linear models perform better than distance-based classifiers in this particular fraud detection scenario (Saputra, Adi; Suharjito (2019).

The study highlights the importance of selecting appropriate evaluation metrics when dealing with imbalanced data, as traditional accuracy measures can be misleading. The authors emphasize the need for precision and recall in assessing fraud detection models due to the critical nature of minimizing false negatives (missed fraud cases). Additionally, the framework demonstrates the effectiveness of PCA in preprocessing and feature reduction, contributing to improved model efficiency without significant loss of information. Based on their findings, the researchers recommend exploring more advanced algorithms, such as ensemble methods and deep learning models, in future work to further enhance fraud detection performance and adaptability across diverse transaction environments.

**FraudLabs Pro’s Online Payment Fraud**

FraudLabs Pro is a comprehensive fraud detection solution that leverages advanced machine learning algorithms and a global merchant network to identify and prevent credit card fraud in e-commerce transactions. The system analyzes multiple data points, including IP geolocation, email validation, credit card BIN (Bank Identification Number) lookup, and transaction velocity, to assign a fraud score to each transaction. Notably, FraudLabs Pro integrates with over 20 e-commerce platforms and offers a REST API for seamless automation, enabling real-time fraud screening. For example, the system can detect anomalies such as mismatches between the IP country and billing address or flag transactions involving anonymous proxies, which are common indicators of fraud. The platform also maintains a dynamic blacklist of high-risk IPs, emails, and devices, sourced from its global network of merchants, enhancing its ability to identify recurring fraud patterns (FraudLabs Pro, n.d.).

The effectiveness of FraudLabs Pro lies in its combination of rule-based validation and machine learning. The system allows merchants to customize fraud rules (e.g., blocking prepaid cards or high-risk countries) while its ML algorithms learn from historical transaction data to improve detection accuracy over time . For instance, the platform’s BIN lookup feature identifies prepaid cards, which are often used in fraudulent transactions due to their anonymity, and allows merchants to set rules for additional verification. Additionally, FraudLabs Pro addresses the challenge of imbalanced datasets (where fraud cases are rare) by analyzing behavioral patterns, such as rapid transaction velocity or disposable email usage, to flag suspicious activity without relying solely on labeled fraud data. While specific performance metrics (e.g., precision or recall) are not publicly disclosed, the platform’s widespread adoption and case studies attest to its practical utility in reducing fraud-related losses. Future research could explore comparative studies between FraudLabs Pro and academic ML models to quantify its performance against benchmarks like AUC or F1-score.

**Amlyze in Combatting Financial Crime**

AMLYZE is a SaaS-based RegTech platform specializing in anti-financial crime solutions, including real-time transaction monitoring and fraud detection, which are highly relevant to credit card fraud detection. The platform leverages a combination of rule-based systems and machine learning algorithms to analyze transaction patterns, with a library of over 200 predefined scenarios targeting anomalies such as unusual spending, high-risk merchant categories, and cash withdrawal irregularities. AMLYZE's transaction monitoring module processes data in milliseconds, using threshold-based, behavioral, and geographical rules to flag suspicious activity, reducing false positives by up to 62% while maintaining high detection accuracy. The system integrates with global sanctions lists and adverse media databases, enhancing its ability to identify high-risk transactions linked to money laundering or fraud.

A key strength of AMLYZE lies in its regulatory expertise, as it was developed by former compliance officers and regulators, ensuring alignment with AML/CFT standards like FATF recommendations. The platform's customizable rule engine allows financial institutions to tailor detection parameters (e.g., transaction amounts, customer risk levels) and test rules against historical data for optimization. While AMLYZE primarily focuses on broader financial crime, its methodologies—such as real-time analytics and dynamic risk scoring—are adaptable to credit card fraud detection, particularly in identifying structured transactions or cross-border fraud. Case studies highlight its effectiveness in reducing fraud-related losses, though specific performance metrics (e.g., precision/recall) for credit card fraud are not publicly detailed. Future research could explore integrating AMLYZE's framework with deep learning models to further enhance detection of evolving fraud tactics in card transactions.

**Kount Developer’s Fraud Prevention Solutions**

Kount is a leading fraud detection platform that combines supervised and unsupervised machine learning with a global data network to identify and prevent credit card fraud in real time. The system analyzes over 1,000 transaction attributes, including device fingerprinting, IP geolocation, and behavioral patterns, to assign a risk score (1–99) to each transaction. Kount’s proprietary Identity Trust Global Network processes data from 32 billion annual interactions, enabling it to detect anomalies such as card testing, account takeovers, and cross-border fraud with high accuracy. The platform’s hybrid ML approach—using both historical data (supervised learning) and real-time pattern detection (unsupervised learning)—addresses class imbalance by dynamically adjusting to emerging fraud tactics, such as IP cloaking or rapid transaction velocity (Equifax Inc., 2025)

Kount’s effectiveness stems from its customizable rules engine and adaptive learning capabilities. Merchants can set thresholds for specific risks (e.g., blocking prepaid cards or high-risk countries) while the ML models continuously refine their predictions based on new data. For example, Kount’s BIN lookup feature identifies stolen cards by flagging mismatches between billing addresses and transaction locations, improving precision by 12% . The platform also integrates SMOTE-like resampling to handle imbalanced datasets, boosting minority-class recall without compromising precision. While specific metrics like F1-scores are not publicly disclosed, Kount reports a 70% reduction in false positives and near-elimination of manual review times (from 30 minutes to 30 seconds per transaction). Compared to academic models, Kount emphasizes scalability and industry collaboration, making it a practical benchmark for real-world fraud detection systems. Future research could explore integrating Kount’s framework with deep learning to enhance detection of AI-driven fraud patterns.

**Marble Project Decision Engine for Fraud**

Marble is an open-source, real-time decision engine for detecting financial malpractices like money laundering, fraud, and service abuse. Its architecture allows organizations to quickly assess transactions, user activities, and events for timely intervention. A user-friendly rule builder enables custom detection scenarios using various data types. The engine supports both real-time and batch analysis for immediate threat detection and historical pattern identification. A built-in case manager aids in investigating flagged instances.

Marble targets payment service providers (PSPs), banking-as-a-service platforms, neo banks, online marketplaces, and telecommunications companies. Its applicability to PSPs and neo banks indicates a design focused on complex financial transactions, which are vulnerable to credit card fraud. Marble's ability to rapidly deploy and update detection scenarios for real-time actions or restrictions is valuable in the evolving landscape of financial fraud. It integrates with existing systems like transaction databases and KYC solutions. A free, self-hosted version is available for fraud detection research and development. Marble prioritizes data privacy by allowing users to keep data within their infrastructure.

Marble's features are relevant to credit card fraud detection. Its design for transaction monitoring in financial institutions supports analyzing credit card transactions for suspicious activity. It utilizes diverse data, including user/account monitoring data, risk scores, and event information, for multifaceted risk assessment. Rule-building capabilities allow for implementing fraud detection logic, such as monitoring deviations in spending patterns or unusual transaction locations. Real-time decision rules are crucial for immediate identification and blocking of fraudulent transactions. Marble's rapid deployment and updating of detection scenarios are advantageous due to the changing nature of fraud. Integration with transaction databases is important for researchers using various datasets. Marble's open data model and user-friendly rule builder facilitate implementing and testing fraud detection logic, potentially informed by machine learning or rule-based strategies. Marble's open data model and integration focus make it a versatile research platform. Researchers can use their own datasets. Marble's strength is processing data from diverse sources using user-defined rules and real-time action capabilities. This aligns with the use of rule-based systems in fraud detection, often with machine learning (Marble, 2025)

Researchers can use Marble to implement and evaluate rule-based strategies for identifying fraudulent credit card transactions, potentially combining them with machine learning insights. Real-time and batch analysis enhance its research utility.

**Tazama Fraud Detection**

Tazama, an open-source real-time transaction monitoring software, is a platform for financial fraud detection, with potential applications in credit card fraud analysis. Designed for various Financial Services Providers (FSPs), from small to national payment switches, Tazama aims to identify and prevent fraud, including credit card fraud. Its architecture ingests transaction data in real-time via a Transaction Monitoring Service (TMS) API for immediate analysis. This real-time capability is crucial for credit card fraud, where swift intervention is essential. Supported by the Linux Foundation and the Bill & Melinda Gates Foundation, Tazama is an open-source system that benefits from community collaboration. Initially for instant payment ecosystems, its adaptable design allows tailoring for various financial transactions, including credit cards, making it relevant for fraud detection research.

Tazama's framework includes mechanisms for credit card fraud analysis. It enables implementing simple and complex rules to identify suspicious transactions based on predefined criteria. Rule processors evaluate transaction attributes and user behavior to detect anomalies. Rule evaluations are aggregated into fraud and money-laundering scenarios called typologies, providing a structured approach to complex fraud. Tazama includes a library of pre-fabricated typologies that can be adapted for credit card fraud detection scenarios. Operational control features, like blocking suspicious transactions and overrides, are included for managing credit card fraud incidents. The modular design, with rule configuration independent of core code, enhances adaptability to evolving fraud tactics.

Tazama's effectiveness in detecting credit card fraud relies on its ability to handle and analyze diverse transaction data. It ingests transaction data in real-time from multiple participants, providing a comprehensive view. This real-time ingestion is crucial for identifying fleeting fraud patterns. The system stores data to enable real-time modeling of participant behavior, a technique used in credit card fraud detection for detecting deviations in spending (Linux Foundation Projects, n. d.). While credit card-specific data fields aren't explicitly detailed, support for various payment schemes implies the ability to handle necessary transaction attributes. Multi-currency support addresses cross-border credit card fraud. The ability to evaluate conditions against transaction event attributes allows for targeted rules for credit card fraud indicators like unusual amounts, locations, or merchant types.

**System Technical Background**

The proposed system leverages a hybrid algorithmic framework rooted in both supervised and unsupervised machine learning techniques to address the complex nature of financial fraud detection. This framework is supported by insights from a wide body of research that highlights the limitations of relying on a single algorithm and the strengths of combining multiple approaches for greater detection accuracy and adaptability.

Supervised learning is represented by K-Nearest Neighbor (KNN) and XGBoost. KNN has been shown to perform effectively in structured financial data classification tasks due to its simplicity and ability to classify transactions based on proximity to known behavior (Saputra & Suharjito, 2019). XGBoost, a gradient boosting algorithm, was frequently identified in the literature as a top-performing model for fraud detection. Its ability to manage large-scale, imbalanced datasets, as demonstrated by Hashemi et al. (2022), makes it particularly suited to the fraud domain where legitimate transactions greatly outnumber fraudulent ones. Studies like Ashtiani & Raahemi (2022) and Alfaiz & Fati (2022) further confirm XGBoost’s superior ROC-AUC, precision, and recall scores when applied to financial datasets.

The unsupervised components—Isolation Forest and Local Outlier Factor (LOF)—enable the detection of previously unseen or novel fraudulent patterns. Isolation Forest isolates anomalies by constructing random binary trees, a method validated for real-time fraud detection in large-scale streaming data by Thennakoon et al. (2019). LOF, which evaluates the local density of data points relative to their neighbors, is particularly effective in identifying contextual anomalies and subtle outliers that would otherwise go undetected by global models. These algorithms do not require labeled training data, a critical advantage noted in studies such as Pourhabibi et al. (2020) and Simaiya et al. (2020), especially when labeled fraud data is scarce or incomplete.

Several studies advocate for ensemble and hybrid approaches that combine the predictive strengths of models like XGBoost with the anomaly detection power of models like LOF or Isolation Forest (Adusumilli et al., 2020; Sailusha et al., 2020). This combination allows the system to address both known fraud scenarios—using historical patterns—and evolving tactics—through anomaly scoring. Ensemble strategies are further supported by findings from Alfaiz & Fati (2022), who demonstrated that combining multiple classifiers with resampling methods led to higher recall and precision compared to single-model approaches.

The system is developed using open-source technologies such as Python, scikit-learn, Pandas, NumPy, and visualization libraries like Matplotlib and Plotly. Web implementation will utilize Flask or FastAPI for backend integration. This technology stack aligns with development practices in both academic and industry-focused systems like FraudLabs Pro and AMLYZE, which similarly emphasize modularity, real-time processing, and extensibility (FraudLabs Pro, n.d.; AMLYZE, 2024). Additionally, case-specific frameworks like Marble and Tazama, which support dynamic rule-building and live monitoring, offer useful architectural guidance for ensuring the system remains adaptable and responsive to new fraud patterns.

Ultimately, this technical foundation allows the proposed application to deliver an efficient, scalable, and locally deployable fraud detection system tailored to the Philippine context. It integrates real-time anomaly detection and historical classification through an accessible and modular platform, bridging research-driven techniques with practical implementation needs.

**Synthesis**

The reviewed literature collectively emphasizes the growing importance of hybrid and ensemble models in financial fraud detection, particularly those combining supervised and unsupervised learning techniques. Many of the studies explored various machine learning algorithms, each with unique strengths and limitations, demonstrating that no single model alone can address the dynamic and evolving nature of fraudulent financial behavior.

Supervised learning models such as K-Nearest Neighbor (KNN), Random Forest, and XGBoost were frequently highlighted for their effectiveness in classifying transactions based on labeled historical data. For example, Saputra & Suharjito (2019) found that KNN achieved strong accuracy in identifying fraudulent transactions within an Indonesian dataset, while Hashemi et al. (2022) demonstrated that XGBoost significantly outperformed other supervised models in precision, recall, and ROC-AUC metrics. Similarly, Alfaiz & Fati (2022) and Ashtiani & Raahemi (2022) confirmed XGBoost’s robustness and efficiency in handling imbalanced datasets, which is a common challenge in fraud detection.

Unsupervised learning models were equally prominent in the literature. Isolation Forest and Local Outlier Factor (LOF) were validated across several studies for their ability to detect novel and subtle anomalies in real-time transaction streams. Thennakoon et al. (2019) proposed a framework using Isolation Forest to analyze streaming financial data, successfully flagging anomalous behavior with minimal latency. Likewise, Pourhabibi et al. (2020) and Simaiya et al. (2020) recommended unsupervised approaches like LOF in situations where labeled data is limited or continuously changing, noting their usefulness in identifying context-dependent fraud scenarios.

Several studies advocate for hybrid and ensemble methods, acknowledging the complementary nature of combining supervised and unsupervised models. Sailusha et al. (2020) introduced a hybrid model blending Isolation Forest with a supervised neural network, which demonstrated improved detection rates over standalone models. Similarly, Adusumilli et al. (2020) recommended pairing anomaly detectors with tree-based classifiers to optimize both accuracy and fraud discovery in real-time systems. These hybrid configurations are especially valuable when dealing with emerging fraud tactics that are not yet reflected in historical training data.

Additionally, practical systems like FraudLabs Pro, AMLYZE, Marble, and Tazama were reviewed for their implementation strategies. These platforms demonstrate real-world success using layered detection approaches, combining machine learning, rule-based scoring, and real-time monitoring. They also highlight the importance of user-centric features like customizable alerts, visual analytics, and low-latency processing, which the current study seeks to integrate into its system design.

In synthesizing these studies, it becomes clear that a hybrid model consisting of K-Nearest Neighbor, XGBoost, Isolation Forest, and Local Outlier Factor would effectively leverage both historical patterns and real-time anomaly detection. This configuration aligns with recent trends in the field and addresses key research gaps, such as the need for scalable, interpretable, and locally applicable fraud detection systems. By adopting a modular and hybrid approach, the proposed system contributes not only a technically sound detection framework but also a practical tool tailored for deployment in Philippine financial environments.

**METHODOLOGY**

**Research Design**

This study employs a Developmental-Experimental Research Design, which is appropriate for technology-oriented projects that aim to create, implement, and evaluate a functional system based on current innovations. The purpose of this research design is to guide the development of a web-based application for detecting fraudulent financial transactions using a hybrid machine learning approach composed of both supervised and unsupervised algorithms: K-Nearest Neighbor, XGBoost, Isolation Forest, and Local Outlier Factor.

The study involves two major phases. The first is the developmental phase, which includes system design, algorithm integration, user interface creation, and implementation using open-source tools and public datasets. The second phase is the experimental phase, wherein the performance of the hybrid model is tested and evaluated against a labeled dataset using metrics such as accuracy, precision, recall, and F1-score. This phase assesses how well the system can detect fraudulent transactions in various scenarios, including both known patterns and novel anomalies.

This approach allows the researchers to iteratively refine the system based on observed outputs, ensuring that both theoretical insights and practical results guide the final application. The developmental-experimental framework is aligned with related studies that similarly developed intelligent fraud detection tools and evaluated their efficacy in simulated or real-world environments.

**Sources of Data**

The primary source of data used in this study is a publicly available dataset titled Credit Card Fraud Detection Dataset 2023, published on Kaggle. This dataset contains over 550,000 anonymized credit card transaction records made by European cardholders in the year 2023 (Elgiriyewithana, N., 2023). Each transaction entry includes various numerical features derived from the original transaction data, along with a label indicating whether the transaction is fraudulent or legitimate. The dataset is particularly suited for machine learning applications as it provides a balanced structure for evaluating the performance of fraud detection models in both supervised and unsupervised settings.

**Participants of the Study**

The participants involved in the testing and evaluation phase of the study are college students from Cavite State University – Imus Campus. These participants were selected to simulate user interaction with the web-based fraud detection system, including tasks such as uploading transaction data, interpreting flagged results, and providing feedback on the interface and usability. Their responses and observations serve as a valuable basis for identifying possible system improvements and ensuring that the platform is user-friendly, informative, and appropriate for use in local financial contexts.

**Sampling Technique**

This study employs purposive sampling to select participants who will engage with and evaluate the developed web-based fraud detection system. The target population consists of Information Technology and Computer Science students from Cavite State University – Imus Campus. These participants were selected based on their foundational knowledge of digital systems, machine learning, or cybersecurity, which enables them to provide meaningful feedback on the system’s usability, interface design, and overall performance. Since the primary purpose of participant involvement is to test the application’s interface and report user experience, purposive sampling ensures that the data collected is relevant to the research goals.

**Data to be Gathered**

The study involves two types of data:

1. Machine Learning Dataset: The main technical dataset consists of over 550,000 credit card transaction records from the publicly available “Credit Card Fraud Detection Dataset 2023” on Kaggle. Each record includes anonymized transaction features and a binary classification indicating fraud or legitimate activity. This dataset will be used to train, validate, and test the hybrid machine learning models (KNN, XGBoost, Isolation Forest, and LOF).

Data Gathering Procedure:

* + The dataset will be downloaded from Kaggle.
  + Data preprocessing will be done using Python libraries (Pandas, NumPy), including normalization and handling of class imbalance using SMOTE.
  + The dataset will be split into training and testing sets (e.g., 70:30 ratio).
  + Performance metrics (accuracy, precision, recall, F1-score) will be computed after model evaluation.

1. User Feedback Data (Usability Testing): Qualitative and quantitative feedback will be gathered from selected participants using evaluation forms and guided usability tasks within the system.

Research Instruments:

* + A Likert-scale-based questionnaire to evaluate ease of use, design clarity, feature functionality, and responsiveness.
  + Task-based usability tests where participants will interact with the system (e.g., upload transaction data, interpret flagged results, test fraud alert features).
  + Short open-ended interview prompts or comment boxes to collect additional feedback.

**Statistical Treatment of Data**

For the machine learning evaluation, statistical analysis will include:

* Accuracy: Measures the proportion of correctly predicted transactions over the total.
* Precision: Measures the proportion of transactions flagged as fraud that are truly fraudulent.
* Recall (Sensitivity): Measures the system’s ability to correctly identify actual fraud cases.
* F1-Score: Harmonic mean of precision and recall, providing a balance between the two.
* Confusion Matrix: To visualize model performance on true positives, false positives, true negatives, and false negatives.
* ROC-AUC (Receiver Operating Characteristic – Area Under Curve): Used to evaluate the trade-off between true positive rate and false positive rate for all classification thresholds.

For the usability feedback:

* Descriptive statistics such as mean, median, and standard deviation will be used to summarize participant responses on Likert-scale items.
* Responses from open-ended questions will undergo thematic analysis to identify common usability insights or interface issues.

**REFERENCES**

Adusumilli, S. B.-V., Damancharla, H., Metta, A. R. (2020). Machine Learning Algorithms for Fraud Detection in Financial Transactions. Retrieved from: https://ijsdcs.com/index.php/ijsdcs/article/view/639

Alaghbari, K.A., Saad, M.H.M., Hussain, A. et al (2022). Complex event processing for physical and cyber security in datacentres - recent progress, challenges and recommendations. J Cloud Comp 11, 65 (2022). Retrieved from: https://doi.org/10.1186/s13677-022-00338-x

Alfaiz, N. S., & Fati, S. M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics*, *11*(4), 662. https://doi.org/10.3390/electronics11040662

Amlyze (2024). Amlyze. Empowering compliance through regulatory insiders. Retrieved from: https://amlyze.com/

Elgiriyewithana, N. (2023). Credit Card Fraud Detection Dataset 2023. Retrieved from: https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023/data

Embroker Team (2024). What is the fraud triangle? (Three components explained). Retrieved from: https://www.embroker.com/blog/fraud-triangle/

Equifax Inc. (2025). Welcome to Kount Developer. Digital identity and fraud prevention solutions. Retrieved from: https://developer.kount.com/hc/en-us

FraudLabs Pro (n.d.). FraudLabs Pro. Secure Your Business From Online Payment Fraud. Retrieved from: https://www.fraudlabspro.com/

Linux Foundation Projects (n. d.). Tazama. Resources for deploying Tazama. Retrieved from: https://tazama.org/community/technical/

M. N. Ashtiani and B. Raahemi (2022). "Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 72504-72525, 2022. Retrieved from: https://doi.org/10.1109/ACCESS.2021.3096799

Marble (2025). Marble. Next Generation Decision Engine for Fraud and Compliance. Retrieved from: https://github.com/checkmarble/marble

Obeng, S., Iyelolu, T. V., Akinsulire, A. A. and Idemudia, C. (2024). Utilizing machine learning algorithms to prevent financial fraud and ensure transaction security. Retrieved from: https://doi.org/10.30574/wjarr.2024.23.1.2185

Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). Credit Card Fraud Detection Using Machine Learning. 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). Retrieved from: https://sci-hub.se/https:/ieeexplore.ieee.org/abstract/document/9121114

Saputra, Adi; Suharjito (2019).  International Journal of Advanced Computer Science and Applications; West Yorkshire Vol. 10, Iss. 9 (2019). DOI:10.14569/IJACSA.2019.0100943. Retrieved from: https://www.proquest.com/openview/6f17cc61f7d88de70e3ade6c284b3fa3/1?cbl=5444811&pq-origsite=gscholar

Simaiya, Sarita & Kumar Lilhore, Dr & Sharma, Sanjeev. (2020). An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods. MATTER: International Journal of Science and Technology. Retrieved from: https://www.researchgate.net/publication/341932015

S. K. Hashemi, S. L. Mirtaheri and S. Greco (2022). "Fraud Detection in Banking Data by Machine Learning Techniques," in IEEE Access, vol. 11, pp. 3034-3043, 2023. Retrieved from: https://ieeexplore.ieee.org/abstract/document/9999220

Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019). Real-time Credit Card Fraud Detection Using Machine Learning. 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). Retrieved from: https://sci-hub.se/https:/ieeexplore.ieee.org/abstract/document/8776942