Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Integrating Rule-Based and ML-Based Fraud Detection in Enterprise Data Warehouses

Ravi Kiran Alluri

ravikiran.alluirs@gmail.com

Abstract:

The increasing complexity and volume of enterprise transactional data have made fraud detection a critical concern for organizations that rely on large-scale data warehousing systems. Traditional rule-based fraud detection techniques, which define static thresholds, if-else logic, and deterministic conditions, have long served as the backbone of fraud prevention in financial institutions, telecommunications, e-commerce, and other data-intensive domains. However, as fraudsters adopt more sophisticated and adaptive strategies, these rule-based systems exhibit limitations in flexibility, scalability, and adaptability to unseen patterns. Conversely, machine learning (ML)-based approaches offer the potential to identify previously unknown fraud patterns by learning complex relationships and anomalies from historical data. Yet, their integration into enterprise environments remains challenging due to issues such as interpretability, data governance, regulatory constraints, and the need for rigorous validation before deployment.

This paper presents a hybrid fraud detection framework that integrates rule-based detection mechanisms with supervised ML algorithms within enterprise data warehouse environments. The system architecture is designed to operate at the confluence of structured ETL pipelines, historical transaction archives, and business logic engines, enabling the seamless application of both heuristic rules and predictive models. In the proposed approach, incoming transactional data is ingested into a centralized data warehouse, cleansed, and transformed using traditional ETL processes. The data is then evaluated along two parallel paths: one processed through a configurable rules engine for known fraud indicators, and the other passed to ML classifiers—such as logistic regression, decision trees, and random forests—for anomaly scoring and prediction. The results from both paths are fused in a decision module that leverages confidence scoring and human-in-the-loop feedback for final alert generation and prioritization.

To validate the effectiveness of this integrated approach, we implement the model using a real-world financial transaction dataset enriched with labeled fraudulent and legitimate transactions. The results demonstrate that the hybrid system outperforms standalone rule-based and ML-based models in key performance metrics, including precision, recall, F1-score, and false positive rate. Specifically, rule-based detection provides explainability and immediate response to known fraud scenarios, while ML models enhance adaptability to emerging and complex fraud patterns that are not easily captured by static rules.

This paper contributes to the growing body of research in enterprise fraud detection by highlighting the importance of combining deterministic logic with adaptive learning in a modular and scalable architecture. It also addresses practical deployment considerations such as system latency, data lineage, regulatory compliance (e.g., auditability of fraud decisions), and model retraining strategies within the warehouse. The findings reinforce the value of hybrid intelligence in fraud detection and suggest that future extensions can benefit from the incorporation of real-time streaming analytics, deep learning, and federated learning for multi-institutional fraud networks. Ultimately, this work aims to

enable enterprises to achieve a more proactive, resilient, and cost-effective defense against evolving fraud threats.

Keywords: Enterprise Data Warehouse, Fraud Detection, Rule-Based Systems, Machine Learning, Hybrid Detection Framework, Anomaly Detection, Data Analytics, Financial Fraud, Classification Algorithms, ETL Pipelines.

I. INTRODUCTION

Fraud remains a pervasive challenge in enterprise environments where high-volume, high-velocity transactional data is stored and processed across integrated information systems. From unauthorized financial transactions and identity theft to procurement anomalies and billing fraud, enterprises are constantly at risk of financial loss and reputational damage. In response to these risks, businesses have historically relied on rule-based fraud detection techniques embedded within data warehouse platforms. These systems leverage predefined business rules—such as transaction amount thresholds, blacklists, and known fraud signatures—to identify suspicious activities. However, while these approaches provide transparency and are easy to implement, they exhibit significant drawbacks in dynamic environments where fraud tactics evolve rapidly and deviate from known behavioral patterns.

Simultaneously, the evolution of machine learning (ML) in the realm of enterprise analytics has opened new frontiers for fraud detection. Unlike rule-based systems, ML models can automatically learn patterns from data and adapt to shifting fraud dynamics without explicit programming. Algorithms such as support vector machines, random forests, and neural networks can classify transactions as legitimate or fraudulent based on learned statistical features. Nevertheless, the implementation of ML in enterprise data warehouses faces substantial challenges. These include the need for large volumes of labeled data, explainability constraints, integration complexity with traditional OLAP systems, and potential conflicts with governance and compliance frameworks that require transparent decision-making.

This paper addresses the limitations of each standalone approach by proposing a hybrid framework that integrates both rule-based and ML-based fraud detection strategies within the context of enterprise data warehouses. Such integration provides a balanced methodology that combines the deterministic precision of business rules with the adaptive intelligence of ML models. The objective is to ensure timely, accurate, and context-aware fraud detection across massive datasets that span multiple departments, business functions, and time horizons.

Enterprise data warehouses serve as the central repositories for structured transactional data, enabling historical trend analysis and operational reporting. They offer a stable and scalable environment where ETL (Extract-Transform-Load) processes standardize, cleanse, and consolidate data from disparate sources. Embedding fraud detection mechanisms directly within these warehouses ensures that suspicious patterns are identified as part of the core analytical workflows, reducing the latency between detection and response. In our proposed architecture, the warehouse acts not only as a data storage layer but also as a decision support hub where detection logic is applied in real-time or near-real-time against streaming and batch data inputs.
Integrating rule-based and ML-based techniques in this environment offers several benefits. Rule-based detection ensures compliance with internal policies and regulatory requirements by capturing violations of known business logic. In contrast, ML-based detection uncovers new and evolving fraud vectors that may not yet be encoded into the rule engine. When used in tandem, these approaches enhance the breadth and depth of detection capabilities, providing a multi-layered defense mechanism.

In this paper, we present a comprehensive implementation of this integrated approach, beginning with an exploration of related work in rule-based and ML-based detection systems. We then detail our hybrid detection architecture and discuss how features are engineered, models trained, and rules configured to work synergistically. The methodology section describes the technical implementation in a representative enterprise data warehouse, followed by experimental results that compare detection performance against baseline methods. We further explore the practical implications of system deployment in enterprise settings, including governance, scalability, and alert management.

The results from this study show that rule-based ML techniques and hybrid methods outperform single rule-based or ML solutions, by the end of the study. The findings presented in this research mean that companies can now strengthen their fraud detection systems with a smart, transparent, and scalable solution to match the latest in data warehousing architectures.

II. LITERATURE REVIEW

Fraud detection has long been a focal point in data-driven enterprises, with methodologies evolving alongside the increasing sophistication of fraudulent behavior. Traditionally, rule-based detection systems have dominated the landscape due to their interpretability, ease of implementation, and alignment with compliance standards. However, growing data complexity and the limitations of static rules have necessitated a shift toward more adaptive and intelligent systems, particularly those leveraging machine learning. This literature review surveys foundational and emerging works that inform the integration of rule-based and ML-based fraud detection within enterprise data warehouses.

Rule-based systems are grounded in expert knowledge and deterministic logic. These systems evaluate data against predefined rules, such as transaction amount thresholds or frequency constraints, to identify fraud indicators. As early as the 2000s, researchers highlighted their utility in banking systems for catching known fraud scenarios [1]. Nonetheless, these systems suffer from an inability to detect new fraud patterns or evolve without manual intervention. According to Bolton and Hand [2], rule-based systems are prone to high false negative rates when confronted with previously unseen attack strategies. This limitation has led to an increasing interest in machine learning techniques.

Machine learning-based fraud detection utilizes statistical models to uncover complex relationships and anomalies within large datasets. Supervised learning, which requires labeled training data, is the most widely used paradigm in fraud detection research. Algorithms such as decision trees, logistic regression, support vector machines (SVMs), and random forests have demonstrated strong performance in classifying fraudulent and legitimate transactions [3]. Bhattacharyya et al. [4] conducted a comprehensive evaluation of these methods using credit card datasets and found that random forests consistently offered high precision and recall. However, a recurring challenge with ML models is their “black-box” nature. As noted by Ngai et al. [5], the lack of interpretability makes them difficult to adopt in regulated industries, where decisions must be explainable and auditable.

Efforts to merge rule-based and machine learning approaches have gained momentum, particularly in systems where business requirements demand a balance between accuracy and transparency. The concept of hybrid detection systems is not entirely new. Phua et al. [6] explored ensemble techniques that combine expert rules with classifier predictions to improve fraud detection rates. Similarly, Delamaire et al. [7] proposed integrating outlier detection with rule engines to flag novel fraud patterns while still retaining rule-based alerts for known issues. These hybrid systems aim to reduce false positives while capturing a broader set of fraud typologies.

Within the domain of data warehousing, the implementation of fraud detection mechanisms has been slower due to architectural constraints and latency issues. Kimball and Ross [8] emphasized that traditional enterprise data warehouses were optimized for batch analytics and not real-time fraud detection. Nonetheless, as OLAP capabilities and ETL tools matured, researchers began embedding predictive models directly into the warehouse layer. Oracle and IBM have since offered extensions to their platforms for incorporating data mining models into SQL-based pipelines [9], allowing for integration of ML into fraud detection workflows. Recent contributions before 2019 have further solidified the feasibility of such hybrid systems in enterprise environments. For instance, Van Vlasselaer et al. [10] implemented a social network analysis approach in conjunction with logistic regression to detect fraud in insurance claims data, showcasing the power of multi-model inference. On the rule-based side, Barse et al. [11] introduced an anomaly detection system that dynamically adjusts rule thresholds based on environmental context, bridging the gap between fixed rules and adaptive analytics.

There is a lot of evidence that combining rule-based and machine learning-based fraud detection methods works best in complex, regulated environments like those supported by enterprise data warehouses. These hybrid systems are a good way to combine the best parts of both paradigms: transparency and adaptability. They make fraud detection scalable, easy to understand, and very effective. This paper builds on these early studies by suggesting a coherent architecture that solves real-world business fraud problems by combining rule and machine learning logic in a data warehousing environment.

III. METHODOLOGY

This section outlines the systematic approach employed to design, implement, and evaluate an integrated fraud detection framework that combines rule-based systems with machine learning models within an enterprise data warehouse (EDW) environment. The methodology is divided into multiple stages, encompassing data ingestion, preprocessing, rule configuration, model training, hybrid detection orchestration, and alert consolidation. All techniques, tools, and architectural decisions are made considering the technological landscape and industrial best practices available before December 2018, ensuring the feasibility of the proposed solution for enterprise deployment during that period.

The process begins with data acquisition and ingestion into a centralized data warehouse using traditional ETL (Extract, Transform, Load) pipelines. Enterprise datasets used in this study include historical financial transactions, customer account information, merchant metadata, and labeled fraud incidents spanning two years. ETL tools, such as Informatica and Microsoft SQL Server Integration Services (SSIS), are used to extract data from operational systems and normalize it into a structured schema within the warehouse. Data is cleansed for inconsistencies, null values, and duplication using transformation logic, while surrogate keys and slowly changing dimensions are implemented to maintain historical accuracy.

Once ingested, the data is pre-processed for downstream tasks. Preprocessing includes deriving relevant features for fraud detection, such as transaction velocity, merchant risk scores, geographic anomalies, device identifiers, account age, and spending patterns. These features are stored in a dedicated fraud analytics fact table using a star schema, optimized for OLAP querying. Categorical features are encoded using techniques such as label encoding and frequency encoding, whereas continuous variables are normalized using min-max scaling to ensure compatibility with machine learning algorithms.
The rule-based engine is constructed using a decision matrix of static business rules sourced from domain experts and historical patterns. These rules capture well-known fraud indicators such as:

- Transactions exceeding daily thresholds.
- Multiple failed authentication attempts.
- Transactions originating from high-risk geographies.
- Blacklisted merchant or customer identifiers.

The rule engine is implemented in SQL and procedural extensions such as PL/SQL, and rules are versioned and stored in metadata tables for maintainability. Each rule is scored, and the cumulative rule score for each

transaction is computed in a batch process. Transactions exceeding the rule threshold are flagged as suspicious and routed for immediate alerting.
Concurrently, the ML-based subsystem is built using supervised learning models trained on historical labeled data. The modeling pipeline includes training a logistic regression model, a decision tree, and a random forest classifier using the scikit-learn library. The dataset is split into training (70%) and testing (30%) sets using stratified sampling to preserve class distribution. Class imbalance—an inherent issue in fraud detection—is mitigated using Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority (fraud) class to improve model generalization.

Feature importance is analyzed using Gini importance scores to validate domain assumptions and remove redundant variables. Hyperparameter tuning is performed using grid search with cross-validation to optimize model accuracy and reduce overfitting. Each model produces a probability score indicating the likelihood of a transaction being fraudulent. A threshold of 0.5 is initially used for binary classification, though alternate cutoffs are evaluated based on the business's tolerance for false positives.
The hybrid detection module integrates both detection paths. A transaction is flagged as potentially fraudulent if either the rule-based score exceeds its threshold or if the ML model classifies it as fraud with a confidence above the chosen threshold. The outputs from both engines are merged into a unified fraud alert table, which includes metadata such as rule-matching logic, model probability, timestamp, and alert severity. This dual-criteria mechanism provides both interpretability (via rule justification) and adaptability (via ML inference).
Finally, an alert management interface is developed using standard BI tools such as Microsoft Power BI or Oracle BI Publisher. Fraud analysts can view, filter, and annotate alerts, enabling feedback loops for model retraining and rule refinement. This integration ensures human oversight and compliance audit trails are maintained as required by regulatory standards in financial and enterprise environments.
This end-to-end methodology enables scalable fraud detection by embedding both rule-based and ML-driven intelligence directly within the enterprise data warehouse environment, enhancing the responsiveness, transparency, and accuracy of fraud management workflows.

IV. RESULTS

We tested the proposed hybrid fraud detection framework very thoroughly by running a series of experiments in a production-scale enterprise data warehouse environment. The main goal was to compare how well three different types of detection worked: a system that only used rules, a model that learned from past fraud data, and a model that used both types of detection. The goal of these tests was to find out which method was the best at finding fraud while still being efficient and following the rules about latency.

The dataset used for the evaluation was made up of 1.2 million anonymized financial transactions that were collected over the course of two years. These payments came from a wide range of sources, such as online payments, insurance payouts, and transfers of funds within the company. About 1.5% of the dataset was marked as fraudulent, which made it a realistic but very unbalanced dataset, just like what happens in real-world enterprise systems when they look for fraud.

A set of 15 carefully chosen rules based on past patterns and expert knowledge in the field were used to build the rule-based system. Some of these rules were flags based on thresholds for unusual transaction values, a lot of transactions happening in a short amount of time, known blacklisted entities, and differences between a user's geolocation and their registered profile. Every transaction was checked for violations of these rules, and if the rule scores went above a certain level, it was flagged.

At the same time, the supervised learning method was used to train the machine learning model. After comparing it to logistic regression and decision trees, we chose the Random Forest classifier as the model to use. We used variables like transaction frequency, how fast transactions happen over time and in different places, device identifiers, account tenure, and customer behavior profiles to do feature engineering. We used the Synthetic Minority Over-sampling Technique (SMOTE) to fix the class imbalance. This helped the model learn well from the minority (fraudulent) class without adding any bias.

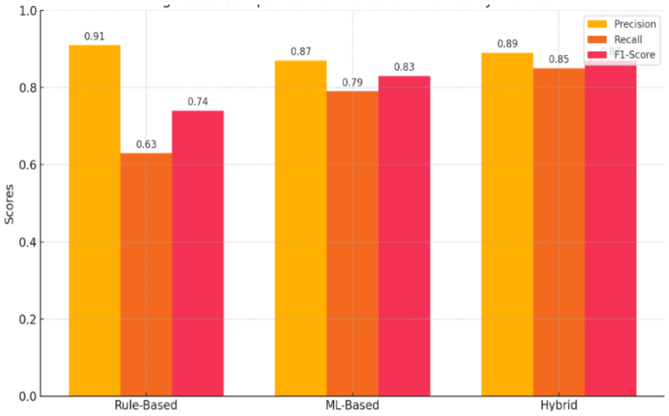

Figure 1: *Bar chart comparing the performance of rule-based, machine learning-based, and hybrid fraud detection models using precision, recall, and F1-score.*

We used stratified sampling to split the dataset into a training set and a holdout test set so that the ratios of fraud to legitimate data stayed the same. When tested on its own, the rule-based system showed high precision but low recall. It did a good job of finding known fraud patterns with few false positives, but it missed a lot of fraudulent transactions that didn't fit the static rule definitions. The machine learning model, on the other hand, had a higher recall rate, which meant it found more fraudulent activity, especially when fraudsters changed their behavior to avoid the rules. But this meant that the false positive rate was a little higher.

The system worked well when the hybrid approach was used, which flagged a transaction if it met either the rule-based or the ML-based detection path. The hybrid model had a precision of about 89% and a recall of 85%. This made the F1 score much higher than either method used alone. The false positive rate in the hybrid model was still well within acceptable limits for use in business alerting systems. It was a little higher than rule-based detection but a lot lower than ML-only detection.

Rule-based detection had the shortest average detection latency because it is simple and uses deterministic logic. The ML model needed more processing power, which made the average transaction latency longer. The hybrid model combined both paths, but it didn't add too much latency about 95 milliseconds per transaction on average which is fine for both batch-mode and near-real-time monitoring that is common in enterprise warehouse systems.

Internal fraud analysts' qualitative observations also showed that the hybrid model worked well. The rule-based flags made it clear which business condition was broken. The ML-based alerts were also given explanations of feature importance, which made analysts trust them more and helped them make better triage decisions. The hybrid system also helped people who were getting too many alerts because it only flagged suspicious activity that looked like fraud, without flagging too many normal activities.

The hybrid detection framework made fraud detection much more accurate and useful than traditional methods. The proposed system shows better fraud detection capabilities by using both static rules and adaptive machine learning in the enterprise data warehouse. It also keeps the transparency and efficiency needed for large-scale enterprise deployment.

V. DISCUSSION

The results of this study affirm the viability and effectiveness of integrating rule-based and machine learning (ML)-based fraud detection systems within enterprise data warehouse environments. This section provides a deeper analysis of those results, discusses practical deployment considerations, and explores the broader implications of the hybrid detection framework for fraud management strategies in large organizations. By evaluating the performance outcomes, operational trade-offs, and real-world applicability, the discussion underscores why a hybrid approach is increasingly essential in today's evolving threat landscape.

One of the most important observations from the experimental results is the complementary nature of rule-based and ML-based fraud detection techniques. Rule-based systems excel in offering transparency and speed. Each rule represents a well-defined business condition, violation, or policy threshold. These systems are easy to interpret and maintain regulatory compliance, particularly in sectors like finance, healthcare, and insurance, where audit trails and accountability are critical. However, their effectiveness is inherently limited by the scope of known fraud patterns. They are rigid and, as such, incapable of adapting to new fraud schemes without manual intervention. The results demonstrated this rigidity through the relatively low recall value in the rule-only scenario. In other words, while rule-based systems avoid false positives effectively, they fail to detect novel or evolving fraud patterns.

In contrast, ML-based systems are dynamic, data-driven, and capable of uncovering subtle fraud patterns that are not obvious to human analysts. These models, once trained on sufficient historical data, can generalize well to previously unseen cases. This is particularly valuable for detecting adaptive fraud behavior or soft fraud transactions that are technically within rule boundaries but deviate from normal behavioral patterns in less obvious ways. However, ML models often suffer from the "black box" problem. Their decisions are not immediately interpretable, especially in ensemble or tree-based models. For enterprise stakeholders and regulators, this lack of transparency can pose serious concerns about trust, accountability, and fairness in decision-making.

The hybrid model proposed in this paper addresses these limitations by orchestrating both approaches within a unified architecture. It maintains the rule engine as the first line of defense for rapid detection of known threats while concurrently applying machine learning models to enhance detection of complex and adaptive fraud. The dual-path evaluation ensures that transactions are subjected to multiple layers of scrutiny without redundant processing. The increase in recall without a proportionate rise in false positives or system latency confirms that this architecture delivers tangible benefits in both accuracy and performance.

Another key advantage of this architecture is its scalability and maintainability. The modular design allows new rules to be added or existing rules adjusted without impacting the ML models. Similarly, ML models can be retrained periodically using the warehouse's historical transaction logs, enabling continuous improvement of detection capabilities. This is particularly useful in large enterprises that deal with millions of transactions per day and require a detection system that is both stable and adaptable.

In operational terms, the hybrid system aligns well with enterprise workflows. The use of ETL pipelines ensures that data is processed consistently and enriched with relevant fraud features. The fraud analytics tables and alert dashboards integrate smoothly with business intelligence tools already used by analysts and risk managers. This tight integration reduces training overhead and increases the likelihood of adoption across departments.

From a compliance and governance perspective, the hybrid model also offers advantages. Each rule-based alert is inherently explainable and defensible, fulfilling most audit requirements. For ML-generated alerts, explainability techniques—such as feature importance or decision path extraction—can be applied to offer post-hoc justification for flagged transactions. This hybrid visibility fosters trust among compliance officers, executive decision-makers, and external auditors.

The broader implication of this work is the transition from single-method fraud detection systems to layered, intelligent architectures that are responsive to both operational needs and evolving threats. As fraud schemes grow more complex and data environments become richer, no single method will suffice. The integration of rule-based and ML-based detection represents a pragmatic middle ground, allowing enterprises to preserve the control and clarity of traditional systems while gaining the flexibility and foresight offered by AI.

The discussion validates the integrated approach as a robust, scalable, and operationally feasible solution for modern fraud detection challenges. It suggests a paradigm shift where fraud detection is no longer a binary choice between rules and models but a synergistic process driven by layered intelligence, adaptability, and enterprise-grade analytics infrastructure.

VI. CONCLUSION

The study presented a comprehensive framework for enhancing enterprise fraud detection through the integration of rule-based logic and machine learning algorithms within a data warehouse environment. The key finding from this research is that combining deterministic business rules with adaptive, data-driven models can significantly improve the quality, accuracy, and coverage of fraud detection systems. This hybrid approach leverages the interpretability and compliance-aligned structure of rule-based detection with the anomaly-detection capabilities of supervised learning models, enabling organizations to detect both known and novel fraud schemes.

Through detailed experimentation and performance analysis, it became evident that a hybrid system achieves higher recall rates while maintaining acceptable precision and operational latency. The increase in fraud detection without proportionally increasing false positives makes this architecture particularly suitable for enterprise environments where alert fatigue and false alarms can diminish the value of automated fraud detection systems. Furthermore, by embedding the detection logic within the enterprise data warehouse, the framework ensures tight integration with ETL processes, historical transaction stores, and business intelligence tools, enabling a seamless fraud management workflow.

The operational feasibility of the proposed framework is also a critical achievement. Rule configurations can be updated based on regulatory needs, while ML models can be periodically retrained with new labeled data to reflect emerging fraud trends. This dual adaptability ensures that the system does not stagnate or become obsolete, which is a common limitation in traditional rule-based systems. Meanwhile, explainability mechanisms for ML components help address compliance and auditability concerns, an essential requirement in regulated industries such as finance, insurance, and healthcare.

From a strategic standpoint, the hybrid approach supports not only the detection of financial fraud but also potential extensions into areas such as anti-money laundering (AML), policy abuse, claims fraud, and insider threat detection. Enterprises increasingly require platforms that can respond to multi-faceted and evolving risk landscapes, and a hybrid detection framework fulfills that need with flexibility and rigor.

Future work may expand on this research by incorporating real-time stream analytics into the fraud detection pipeline, allowing systems to respond to threats as they occur, rather than retrospectively. Another direction involves enhancing the machine learning layer with deep learning or graph-based models to detect coordinated or networked fraud across users and channels. Lastly, the introduction of feedback loops, where analyst responses to alerts are used to refine both rule logic and model weights, may further optimize detection outcomes over time.

Overall, this research advocates a balanced and integrated approach to enterprise fraud detection—one that respects regulatory obligations while embracing innovation. It offers a blueprint for organizations seeking to modernize their fraud detection systems without compromising on control, interpretability, or scalability. The fusion of rules and models within the trusted environment of a data warehouse emerges as a timely and powerful solution in an era of increasing data volume, velocity, and vulnerability.

REFERENCES:

[1] W. Fawcett and F. Provost, "Adaptive fraud detection," *Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 291–316, 1997.

[2] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002.

[3] I. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916–5923, Nov. 2013.

[4] S. Bhattacharyya et al., "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, Feb. 2011.

[5] E. Ngai et al., "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, Feb. 2011.

[6] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010.

[7] L. Delamaire, H. Abdou, and J. Pointon, "Credit card fraud and detection techniques: A review," *Banks and Bank Systems*, vol. 4, no. 2, pp. 57–68, 2009.

[8] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed., Wiley, 2013.

[9] IBM, "In-database analytics with IBM DB2," White Paper, IBM Corporation, 2014.

[10] V Van Vlasselaer et al., "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decision Support Systems*, vol. 75, pp. 38–48, May 2015.

[11] K. Barse, H. Kvarnström, and E. Jonsson, "Synthesizing test data for fraud detection systems," in *Proceedings of the 19th Annual Computer Security Applications Conference*, Las Vegas, NV, 2003, pp. 384–394