Research Paper
Impact Factor: 7.358
Peer Reviewed & Indexed Journal
www.ijmsrr.com
Special Issue
IJMSRR
E- ISSN - 2349-6746
ISSN -2349-6738

AUTONOMOUS BUSINESS DECISION GOVERNANCE ENGINE: A RISK-AWARE FRAMEWORK FOR RESPONSIBLE AI-DRIVEN ENTERPRISE AUTOMATION

Dr.M Durga Prasad* Dr. Balusu Nandini**
**Ph.D., Department of Computer Science &Engineering, Telangana University,Dichpally, Telangana.*
***Associate Professor, Department of Computer Science &Engineering,Telangana University, Dichpally, Telangana.*

Abstract
Artificial intelligence (AI) is widely applied in modern financial systems to detect fraudulent transactions. Machine learning models analyze large volumes of transaction data and identify patterns that may indicate fraud. However, many existing fraud detection systems primarily rely on prediction outputs and often lack governance mechanisms that supervise automated decisions. In financial systems, even a small number of missed fraud cases may lead to significant financial losses if transactions are automatically approved. This study presents an Autonomous Business Decision Governance Engine (ABDGE) that supervises AI-based fraud detection systems before automated actions are executed. The proposed framework integrates three components: A Random Forest classifier for fraud prediction, an Isolation Forest model for anomaly detection, and SHAP-based explainable artificial intelligence for interpreting model behavior. These components are combined to compute a governance risk score that evaluates the safety of automated decisions. Based on the computed risk, transactions are routed into three categories: fully automated processing, human review, or escalation for further investigation. Experiments were conducted on the PaySim financial transaction dataset containing more than 6.3 million transactions. The Random Forest model achieved high predictive performance but still missed a small number of fraudulent transactions. In the experimental evaluation, the ABDGE governance layer escalated all fraud cases missed by the machine learning model, thereby reducing the risk of unsafe automated approvals. The results show that integrating governance mechanisms with AI systems improves reliability, transparency, and trust in automated financial decision processes.

Keywords:Fraud detection, Explainable AI, Governance risk, Random Forest, Isolation Forest, SHAP, Trustworthy AI.

Introduction

Digital financial systems have experienced rapid growth due to the widespread adoption of online banking, mobile payments, and electronic commerce(Williams et al., 2021). While these technologies provide convenience and efficiency, they also increase the risk of fraudulent financial activities. Detecting fraudulent transactions has therefore become a critical task for financial institutions(Sadineni, 2020).Machine learning techniques are increasingly applied in fraud detection because they can process large volumes of transaction data and identify patterns that may indicate fraudulent activity(Ali et al., 2022). Algorithms such as decision trees, random forests, and neural networks have demonstrated strong predictive performance in detecting fraudulent transactions.

Despite their effectiveness, machine learning models are not perfect. Even highly accurate models can still produce occasionally misclassifications(Höfler, 2005). In fraud detection, a false negative occurs when a fraudulent transaction is incorrectly classified as legitimate. When such transactions are automatically approved by financial systems, they may result in financial loss and reduced trust in automated decision processes.Another important challenge is the lack of transparency in many machine learning models. Many complex machine learning models behave as black-box

systems(Evans et al., 2019), making it difficult to understand the reasoning behind their predictions. In financial applications, transparency and accountability are essential for maintaining trust and regulatory compliance. Recent research emphasizes the importance of trustworthy and responsible AI systems(Lu et al., 2023), which incorporate explainability, transparency, and human oversight. Instead of relying solely on prediction outputs, AI systems should include governance mechanisms that evaluate the risk associated with automated decisions.

To address these challenges, this paper proposes an Autonomous Business Decision Governance Engine (ABDGE) that supervises AI-driven fraud detection systems. The proposed framework integrates machine learning predictions, anomaly detection, and explainable AI analysis to evaluate the risk of automated decisions.
The main contributions of this work are summarized as follows:

1. A governance framework that supervises AI-based fraud detection systems before automated decisions are executed.
2. A governance risk model that combines explain ability signals, anomaly detection scores, and machine learning prediction probabilities.
3. A decision routing mechanism that categorizes transactions into automated processing, human review, and escalation levels.

Experimental validation demonstrating that the governance layer successfully escalates fraud cases missed by the machine learning model.

Literature Review

AI has become a critical technology for improving fraud detection and financial security in digital transaction systems(Zainal, 2023). Fraud detection has been extensively studied in the fields of data mining and machine learning. Early fraud detection systems relied primarily on rule-based methods, where expert-defined rules were used to identify suspicious transactions. Although rule-based systems are interpretable, they often struggle to detect new evolving fraud patterns(Mohite & Ouarbya, 2024).Machine learning approaches enable financial institutions to analyze large transaction datasets and identify patterns that may indicate fraudulent behavior. Ensemble methods such as Random Forest are widely used because they combine multiple decision trees to produce robust predictions. Random Forest models are particularly effective in handling high-dimensional data and capturing nonlinear relationships among features(Quaye, 2024).

In addition to supervised learning methods, anomaly detection techniques are commonly used to identify unusual patterns in financial transactions. Isolation Forest is a well-known anomaly detection algorithm that isolates observations that differ significantly from normal data points(Fadul, 2023). Such techniques can help detect previously unseen fraud patterns.Explainable artificial intelligence (XAI) has also gained attention in recent years(Das & Rad, 2020). Methods such as SHAP (SHapley Additive Explanations) allow researchers to interpret machine learning predictions by measuring the contribution of each feature to the model's output. In financial systems, model transparency is essential to ensure that automated decisions can be understood and audited. Although these approaches improve predictive performance and interpretability, many existing fraud detection systems still focus primarily on classification accuracy. Relatively fewer studies have explored the integration of prediction models, anomaly detection, and explainability that supervise automated decisions(Ayeola, 2025).

Although machine learning models have significantly improved fraud detection performance, most

existing systems focus mainly on prediction accuracy. Limited research has addressed about AI decisions should be supervised before being executed in real financial systems. This gap highlights the need for governance frameworks that combine prediction models, anomaly detection, and explainability to ensure safe and trustworthy automated decision. The approach presented in this study addresses this limitation by introducing a governance layer that evaluates the safety of AI-based decisions before they are executed.

Methodology

This study introduces an Autonomous Business Decision Governance Engine designed to improve the reliability of AI-based fraud detection systems. The framework combines supervised machine learning, anomaly detection, explainable AI techniques, and a governance layer for decision routing.The methodology includes data preprocessing, model training, explainability analysis, governance risk calculation, and decision routing.

A. Dataset Description

The PaySim financial transaction dataset was used to train and evaluate the proposed ABDGE framework.

B. Data Preprocessing and Feature Engineering

To improve fraud detection capability, additional features were derived from the original attributes. Two additional features were created to capture balance inconsistencies in transactions.

The sender balance error was computed as

$$errorBalanceOrig = oldbalanceOrg - newbalanceOrig - amount$$

Similarly, the receiver balance error was calculated as

$$errorBalanceDest = oldbalanceDest + amount - newbalanceDest$$

These features highlight differences between expected and actual account balances, which may indicate fraudulent activity. Categorical transaction types were converted into numerical variables using one-hot encoding. After preprocessing, the dataset consisted of 12 predictive features used for model training.Non-informative attributes such as account identifiers (nameOrig and nameDest) were removed since they do not contribute to predictive modeling.

C. Dataset Partitioning

The dataset was divided into training and testing subsets using stratified sampling to maintain the original class distribution. The training set contained 4,453,834 transactions, while the testing set consisted of 1,908,786 transactions. A test ratio of 30% was used to evaluate the performance of the model on unseen data.

D. Fraud Detection Model

A Random Forest classifier was employed as the primary fraud detection model. Random Forest is an ensemble learning algorithm that builds multiple decision trees and combines their outputs to improve prediction accuracy.The model was configured with several parameters to ensure stable training performance. A total of 150 decision trees were used in the model, with the maximum depth of each tree limited to 10. To address the severe class imbalance commonly present in fraud detection datasets, the class weight parameter was set to balanced so that fraudulent transactions received appropriate importance during the learning process. A random seed value of 42 was also specified to maintain reproducibility of the experimental results. After the training phase, the model generated two outputs for every transaction in the test dataset. The first output was a fraud prediction

label indicating whether the transaction was classified as fraudulent or legitimate, while the second output provided a fraud probability score that represents the likelihood of the transaction being fraudulent. These outputs were subsequently used as key inputs for the governance layer in the proposed system.

E. Anomaly Detection

To detect abnormal transaction patterns that may not be captured by the supervised model, an Isolation Forest algorithm was applied. Isolation Forest detects anomalies by isolating unusual observations using randomly generated trees.Each transaction was assigned an anomaly score, where higher scores indicate more suspicious behavior. The anomaly scores were normalized to a 0–1 scale using Min-Max normalization to enable integration with other governance metrics.

F. Explainable AI Analysis

Explainability was incorporated using SHAP (SHapley Additive Explanations). SHAP values measure how each feature influences the model's prediction. A Tree Explainer was used to compute SHAP values for the Random Forest classifier. Due to computational constraints, SHAP values were calculated for a randomly selected subset of 10,000 test transactions. Absolute SHAP values were used to measure the strength of each feature's influence. These values were then normalized to generate standardized feature importance scores. A matrix was constructed to store the normalized SHAP contributions of each feature for the sampled transactions.

G. Explainability Risk Computation

Explainability risk reflects how strongly different features influence the model's predictions. It is computed as the average magnitude of normalized SHAP contributions across all features and sampled transactions. Higher values indicate a more complex decision process.

H. Governance Risk Model

To govern AI-driven decisions, a composite governance risk score was introduced. The governance risk combines three components:

- Explainability risk derived from SHAP analysis
- Anomaly score obtained from Isolation Forest
- Fraud probability predicted by the Random Forest model

The governance risk was computed using a weighted formulation, where the weights were selected empirically to emphasize explainability uncertainty while incorporating anomaly behavior and model confidence:

Governance Risk = 0.5 × Explainability Risk + 0.3 × Anomaly Score + 0.2 × Fraud Probability

All components were normalized before computing the final governance risk score.To classify transactions according to their risk level, two governance thresholds were determined from the distribution of governance risk values. The first threshold ($\tau 1$) represents the 50th percentile, while the second threshold ($\tau 2$) represents the 80th percentile. These thresholds were used to classify transactions into low, moderate, and high governance risk levels.

I. ABDGE Decision Engine

The Autonomous Business Decision Governance Engine (ABDGE) routes transactions into three decision categories based on the computed governance risk and anomaly scores:

- Fully Automated: low-risk transactions that can be processed automatically

- Human Review: moderate-risk transactions requiring manual verification
- Escalation: high-risk transactions requiring further investigation

Results and Discussion

A. Dataset Overview

The proposed ABDGE framework was evaluated using the PaySim financial transaction dataset, which simulates mobile money transactions based on real financial behavior. The dataset consists of 6,362,620 transactions, with features including transaction type, amount, sender and receiver balances, and fraud indicators.For evaluation, the dataset was split using stratified sampling to preserve the distribution of fraudulent transactions. The dataset was divided into training and testing subsets as described in the methodology section.Fraudulent transactions constitute a very small fraction of the dataset, making this a highly imbalanced classification problem.

B. Evaluation Methodology

The evaluation was conducted in two stages:
Machine Learning Fraud Detection:
The trained Random Forest model was evaluated on the test dataset to generate fraud predictions and probability scores.
Governance Evaluation with ABDGE:
The ABDGE governance layer evaluates each transaction's risk by integrating three signals:

1. Random Forest fraud probability
2. Isolation Forest anomaly score
3. SHAP-based explain ability risk

The combined governance risk score determines whether a transaction can be processed automatically or requires human review or escalation.

C. Machine Learning Model Performance

The Random Forest classifier demonstrated strong predictive performance on the test dataset:

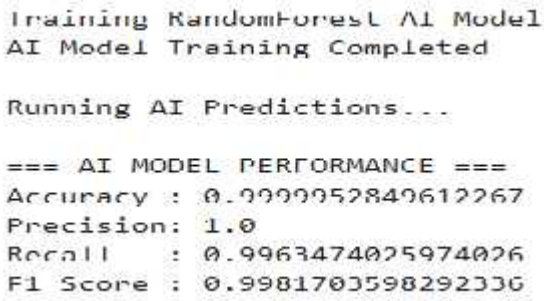

Figure 1: Performance Metrics of the Random Forest Fraud Detection Model

The Random Forest model achieved very high predictive performance with high precision and F1 score, indicating it reliably identifies fraudulent transactions. However, even the few missed fraud cases are extremely costly in a commercial financial setting, emphasizing the importance of a governance layer to catch every high-risk transaction.

The confusion matrix for the test dataset is shown below:

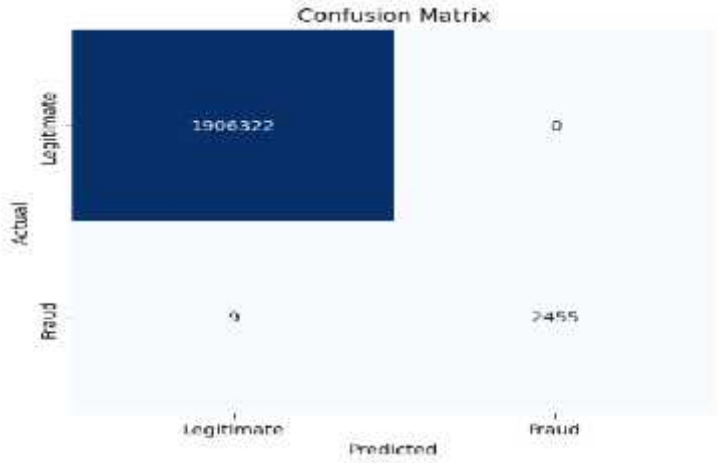

Figure 2: Confusion Matrix of the Fraud Detection Model Before Applying the Governance Model.

Out of 2,464 fraudulent transactions, the model successfully detected 2,455 cases, while 9 transactions were missed. Although the model shows excellent performance, these missed fraud cases highlight the need for an additional governance mechanism to ensure safe automation.

D. Governance Risk Thresholds

The ABDGE framework calculates governance risk thresholds based on the distribution of the risk scores:

$\tau 1 = 0.03$

$\tau 2 = 0.078$

These thresholds categorize transactions into different levels of oversight:

- Low risk → Fully Automated
- Medium risk → Human Review
- High risk → Escalation

E. ABDGE Decision Distribution

Applying the ABDGE governance engine to the test dataset resulted in the following distribution of transaction decisions:

Table 1: Showing distribution of transaction decisions with ABDGE Governance

Decision Category	Number of Transactions
Fully Automated	954,178
Human Review	572,891
Escalation	381,717

Approximately 50% of transactions were fully automated, while the remaining transactions required either human review or escalation. This ensures that potentially risky transactions receive appropriate oversight.

F. Handling Missed Fraud Cases

Although the Random Forest classifier missed 9 fraudulent transactions, the ABDGE governance engine successfully routed all nine cases to the escalation category due to elevated governance risk and anomaly scores as shown in Figure 3.

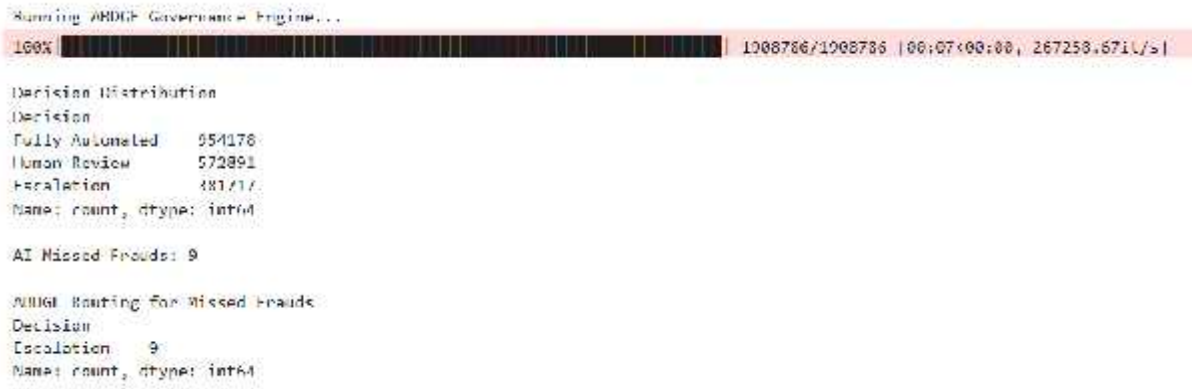

Figure 3: ABDGE execution interface illustrating detection of fraudulent transactions missed by the Random Forest classifier.

This demonstrates that the governance layer provides an additional safety mechanism, preventing fraudulent transactions from being automatically approved even when the machine learning model fails.

G. Comparison with Existing Fraud Detection Methods

To highlight the advantages of ABDGE, a comparison with existing fraud detection approaches was is shown in Table 2.

Table 2: Comparison of ABDGE with Existing Fraud Detection Approaches

Method	Machine Learning	Explainability	Anomaly Detection	Governance Decision Layer
Traditional Rule-Based Systems	No	No	No	No
Basic Machine Learning Models	Yes	No	No	No
Explainable AI Models	Yes	Yes	No	No
Hybrid ML + Anomaly Detection	**Yes**	**No**	**Yes**	**No**
Proposed ABDGE Framework	Yes	Yes	Yes	Yes

The comparison illustrates that most existing systems primarily focus on prediction accuracy and lack mechanisms for transparency, anomaly supervision, or governance. Unlike traditional fraud detection approaches, ABDGE combines machine learning prediction with explainability and anomaly analysis under a governance layer. The Random Forest model identifies potential fraud, while SHAP helps understand which features influence the prediction. Isolation Forest is used to capture unusual transaction patterns that the supervised model may miss. Based on these signals, the governance engine decides whether a transaction can be processed automatically or should be reviewed by humans, improving the reliability of automated financial decisions.

The experimental results demonstrate the benefits of integrating governance mechanisms into AI-based fraud detection systems: Although the Random Forest model achieved high accuracy, the missed fraud cases highlight the risks of relying solely on automated predictions.

The ABDGE governance layer adds a safety net by incorporating explain ability and anomaly detection into a risk assessment framework. High-risk transactions are successfully routed to human review or escalation, preventing unsafe automated approvals.By combining multiple signals,

ABDGE aligns with principles of trustworthy AI, emphasizing transparency, accountability, and human oversight. Overall, the framework improves both the accuracy and reliability of financial fraud detection systems. It ensures that AI predictions are supervised, and high-risk transactions are evaluated before automation, which is critical for real-world financial operations.

Conclusion

This study introduced the Autonomous Business Decision Governance Engine (ABDGE) for supervising AI-based fraud detection systems. The framework integrates machine learning predictions, anomaly detection, and explainable AI analysis to evaluate the safety of automated financial decisions. Experimental evaluation on the PaySim dataset demonstrated that the ABDGE framework successfully escalated all fraud cases missed by the machine learning model, thereby reducing the likelihood of unsafe automated approvals.

Future research may extend the ABDGE framework to other domains such as credit risk assessment, cyber security, and healthcare decision support.

References

1. Ali, A., AbdRazak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., &Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. Applied Sciences, 12(19), 9637.
2. Ayeola, F. (2025). AI-Driven Anomaly Detection and Data Governance: Securing Financial Systems, Cybersecurity Networks, and Regulatory Compliance in Multicultural Digital Ecosystems. Cybersecurity Networks, and Regulatory Compliance in Multicultural Digital Ecosystems (September 04, 2025)
3. Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. arXiv Preprint arXiv:2006.11371.
4. Evans, B. P., Xue, B., & Zhang, M. (2019). What's inside the black-box? A genetic programming method for interpreting complex machine learning models. 1012–1020.
5. Fadul, A. M. A. (2023). Anomaly detection based on isolation Forest and local outlier factor. *Africa University*.
6. Höfler, M. (2005). The effect of misclassification on the estimation of association: A review. International Journal of Methods in Psychiatric Research, 14(2), 92–101.
7. Lu, Q., Zhu, L., Whittle, J., & Xu, X. (2023). Responsible AI: Best practices for creating trustworthy AI systems. Addison-Wesley Professional.
8. Mohite, R., &Ouarbya, L. (2024). Interpretable anomaly detection: A hybrid approach using rule-based and machine learning techniques. 1–10.
9. Quaye, G. E. (2024). Random forest for high-dimensional data.
10. Sadineni, P. K. (2020). Detection of fraudulent transactions in credit card using machine learning algorithms. 659–660.
11. Williams, M., Yussuf, M. F., &Olukoya, A. O. (2021). Machine learning for proactive cybersecurity risk analysis and fraud prevention in digital finance ecosystems. Ecosystems, 20, 21.
12. Zainal, A. (2023). Role of artificial intelligence and big data technologies in enhancing anomaly detection and fraud prevention in digital banking systems. International Journal of Advanced Cybersecurity Systems, Technologies, and Applications, 7(12), 1–10.