Frontline Marketing, Management and Economics Journal
ISSN: 2752-700X

Explainable AI in Banking Compliance: Leveraging Large Language Models for AML and KYC Decision Support

Sakib_Salam_Jamee[1], I K M SAAMEEN YASSAR[2], Md Arif Hossain[3], Mohammad Musa Mia[4], Molay Kumar Roy[5]

[1]Department of Management Information Systems, University of Pittsburgh, PA, USA
[2]Masters of Science and Information Technology, Washington University of Science and Technology, USA
[3]Master of Science in Management Information System, College of Business, Lamar University, Beaumont, TX, US
[4]Master of Business Administration, International American University, Los Angeles, California
[5]Ms in Digital Marketing & Information Technology Management, St. Francis College, USA

A R T I C L E I N f O

Article history:
Submission Date: 29 Nov. 2025
Accepted Date: 22 Dec. 2025
Published Date: 07 Jan. 2025
VOLUME: Vol.06 Issue01
Page No. 06-12
DOI:
https://doi.org/10.37547/marketing-fmmej-06-01-02

A B S T R A C T

In this study, we propose an explainable artificial intelligence framework for banking compliance that integrates traditional machine learning models with large language models to support Anti-Money Laundering and Know Your Customer decision-making. The proposed approach emphasizes regulatory transparency, auditability, and human-centered interpretability while maintaining strong predictive performance. Using an open-source financial dataset from the UCI Machine Learning Repository, we demonstrate how explainable modeling and natural language reasoning can enhance compliance decision support systems. This research aligns with U.S. national interests by improving financial system integrity, reducing compliance costs, and strengthening risk monitoring capabilities in regulated institutions.

Keywords: Explainable Artificial Intelligence, Banking Compliance, Anti-Money Laundering (AML), Know Your Customer (KYC), Large Language Models, SHAP, Real-Time Risk Monitoring.

INTRODUCTION

The rapid digitalization of banking operations and the exponential growth of financial transactions have significantly increased the complexity of regulatory compliance in the global financial system. Anti-Money Laundering and Know Your Customer regulations require financial institutions to continuously monitor customer behavior, transaction patterns, and risk indicators in near real time. Traditional rule-based compliance systems are no longer sufficient to address the scale, velocity, and sophistication of modern financial crimes, leading institutions to adopt data-driven and artificial intelligence-based solutions.

Despite their operational advantages, many AI-driven compliance systems suffer from a lack of transparency, which poses substantial challenges for regulatory approval, auditability, and legal accountability. Regulators increasingly demand not only accurate risk predictions but also clear explanations that justify why a customer or transaction is flagged as suspicious. This requirement has elevated explainable artificial intelligence from a desirable feature to a regulatory necessity in banking compliance.

In parallel, real-time data streams generated from transactional systems, customer interaction platforms, and digital payment networks have introduced new opportunities for proactive compliance monitoring. Real-time analytics enable earlier detection of anomalous behavior, faster intervention, and reduced financial and reputational risk. However, integrating real-time data with explainable decision-making remains an open research challenge.

In this study, we propose an expanded explainable AI framework for banking compliance that incorporates both historical open-source data and simulated real-time transaction streams. By integrating machine learning models with large language models, the framework supports continuous AML and KYC decision-making while generating regulator-aligned explanations in near real time. Prior research has explored the use of AI for automating compliance documentation and interpreting regulatory guidelines; however, limited work has investigated the integration of structured risk models with large language models to generate regulator-aligned explanations for AML and KYC decisions. This study addresses this gap by presenting an integrated framework that aligns predictive accuracy with regulatory transparency.

Literature Review

The increasing complexity of financial crimes and regulatory obligations has driven extensive research into the application of artificial intelligence for Anti-Money Laundering and Know Your Customer compliance. Early compliance systems were predominantly rule-based, relying on predefined thresholds and expert-crafted heuristics. While effective for simple scenarios, such systems lack scalability and adaptability, leading to high false-positive rates and significant operational costs. As financial transactions became more voluminous and heterogeneous, researchers began exploring machine learning techniques to enhance risk detection and automate compliance decision-making.

Supervised learning models such as logistic regression, decision trees, and support vector machines have been widely adopted for financial risk classification tasks. Logistic regression remains a common baseline due to its interpretability and regulatory familiarity; however, its linear assumptions limit its ability to capture complex behavioral patterns. Ensemble models, including random forests and gradient boosting machines, have demonstrated superior performance in detecting fraudulent or anomalous financial behavior by modeling nonlinear relationships and feature interactions. Empirical studies consistently report higher recall and robustness for ensemble-based methods in AML-related tasks, particularly in imbalanced datasets where high-risk cases are rare but critical.

Despite their predictive strength, complex machine learning models introduce transparency challenges that conflict with regulatory requirements. This limitation has motivated significant research into explainable artificial intelligence. Post-hoc explanation techniques, such as Local Interpretable Model-Agnostic Explanations and SHapley Additive exPlanations, have emerged as dominant approaches for interpreting black-box models. SHAP, grounded in cooperative game theory, provides both global and local explanations and has been widely adopted in financial risk modeling due to its theoretical consistency and alignment with regulatory audit processes. Studies in credit scoring and fraud detection have shown that SHAP-based explanations can effectively bridge the gap between model complexity and regulatory transparency.

More recently, researchers have begun exploring the integration of natural language processing and large language models within financial compliance workflows. Large language models have demonstrated strong capabilities in text summarization, reasoning, and contextual explanation, making them suitable for interpreting structured model outputs and generating human-readable narratives. Initial studies have examined the use of language models for regulatory document analysis, policy interpretation, and compliance reporting automation. However, these applications primarily focus on unstructured text rather than decision-level explainability.

Only a limited body of research has investigated the combined use of structured predictive models and large language models for explainable compliance decision support. Existing studies often treat explainability and natural language generation as separate processes, lacking a unified framework that aligns quantitative feature attributions with regulator-oriented explanations. Furthermore, real-time compliance monitoring remains underexplored, with most prior work relying on static historical datasets that fail to capture behavioral drift and evolving risk patterns.

This study extends the existing literature by presenting an integrated explainable AI framework that combines ensemble machine learning models, SHAP-based feature attribution, and large language model-driven narrative explanations. Unlike prior approaches, the proposed framework supports both historical and simulated real-time data streams, enabling continuous AML and KYC decision support. By aligning predictive accuracy with interpretability and temporal responsiveness, this research contributes a scalable and regulator-ready solution to the growing demands of modern banking compliance.

Methodology

Data Collection and Dataset Description

In this research, we utilized the Bank Marketing Dataset obtained from the UCI Machine Learning Repository as an open-source proxy for compliance-related customer decision data. The dataset consists of 45,211 customer records with 17 attributes capturing demographic information, financial characteristics, and interaction histories. Although originally designed for marketing outcome prediction, the dataset closely resembles real-world AML and KYC profiling scenarios, where customer behavior and background information inform approval or rejection decisions.

The target variable indicates whether a client subscribed to a term deposit, which we treated as a proxy for a binary compliance decision. This abstraction enables the modeling of risk-based customer classification while preserving ethical data usage and reproducibility.

Data Preprocessing

We conducted exploratory data analysis to identify inconsistencies, missing values, and outliers prior to model training. To provide transparency regarding the data preparation process, we summarize the dataset characteristics used during preprocessing in Table 1.

Table 1: Dataset Description and Preprocessing Summary

Feature Name	Feature Type	Description	Preprocessing Applied
age	Numerical	Client age	Standardization
job	Categorical	Type of occupation	Label normalization and encoding
marital	Categorical	Marital status	Mode imputation and encoding
education	Categorical	Education level	Mode imputation and encoding
default	Categorical	Credit default history	Binary encoding
balance	Numerical	Average yearly balance	Outlier treatment and scaling
housing	Categorical	Housing loan status	Binary encoding
loan	Categorical	Personal loan status	Binary encoding
contact	Categorical	Contact communication type	Encoding
day	Numerical	Last contact day	Scaling
month	Categorical	Last contact month	Encoding
duration	Numerical	Contact duration	Log transformation
campaign	Numerical	Number of contacts	Scaling
pdays	Numerical	Days since last contact	Missing value handling
previous	Numerical	Number of previous contacts	Scaling
poutcome	Categorical	Outcome of previous	Encoding

		campaign	

Numerical attributes were standardized using z-score normalization, while categorical attributes were transformed into machine-readable formats using encoding strategies that preserved semantic consistency. Missing values were imputed using median values for numerical features and modal values for categorical features. To address class imbalance, we applied resampling techniques that preserved minority-class signals relevant to high-risk compliance scenarios.

Feature Extraction and Engineering

Feature extraction in this study was designed to capture compliance-relevant behavioral, demographic, and interaction-based signals that reflect real-world AML and KYC assessment criteria. We focused on identifying features that represent customer stability, financial behavior consistency, and engagement patterns commonly evaluated by compliance analysts during risk assessment. Behavioral attributes related to interaction frequency, contact duration, and response history were aggregated to reflect both habitual behavior and deviations from established norms.

To address the dynamic nature of financial risk, we incorporated temporal feature extraction using sliding time windows. Within each window, statistical measures such as rolling means, variances, growth rates, and change-point indicators were computed to capture short-term anomalies and long-term behavioral trends. These temporal features enabled the detection of sudden increases in transaction intensity or irregular interaction patterns that may signal suspicious activity.

Feature engineering further enhanced the representational power of the dataset by constructing composite indicators aligned with compliance reasoning. Engagement intensity scores were derived by combining interaction frequency and duration metrics, while financial stability proxies were generated using balance-related attributes and historical behavior consistency. Interaction terms between demographic variables and behavioral indicators were introduced to capture nonlinear relationships frequently observed in AML investigations. This enriched feature space improved both predictive performance and interpretability by aligning model inputs with human decision-making logic.

Model Development

The predictive component of the framework was developed using supervised learning models trained to estimate compliance-related outcomes. Let the feature vector for a customer be denoted as $x \in \mathbb{R}^n$. The probability of a positive compliance outcome was modeled as:

$$P(y = 1 \mid x) = \sigma(w^T x + b)$$

where $\sigma(\cdot)$is the logistic function, $w$represents learned weights, and $b$is the bias term? Gradient boosting models were additionally employed to capture nonlinear relationships.

To enhance explainability, we computed SHAP values $\phi_i$for each feature, representing its marginal contribution to the prediction. These values served as structured inputs to a large language model, which generated natural language explanations by conditioning on feature importance scores, customer profiles, and decision thresholds.

Explainability and LLM Integration

The large language model functioned as an interpretive layer, translating quantitative model outputs into regulator-aligned narratives. Given a set of SHAP values and predicted risk scores, the LLM produced explanations that articulated why a customer was classified as low or high risk, referencing specific attributes and behaviors. This approach enabled transparent, auditable decision support consistent with regulatory expectations.

Model Evaluation

Model evaluation was conducted using a rigorous and multi-dimensional framework to ensure predictive reliability, robustness, and regulatory suitability. Standard classification metrics, including accuracy, precision, recall, and F1-score, were employed to assess baseline performance. Particular emphasis was placed on recall, as false negatives represent critical failures in AML and KYC systems by allowing high-risk customers to remain undetected.

To evaluate robustness under dynamic conditions, we assessed model performance across both historical datasets and simulated real-time streaming environments. Rolling-window evaluation was applied to measure stability over time and sensitivity to concept drift caused by evolving customer behavior. This approach allowed for continuous monitoring of model degradation and ensured that performance remained consistent as new behavioral data was introduced.

Explainability was evaluated through both global and local interpretability analyses. Global SHAP value aggregation was used to identify dominant risk drivers across the customer population, providing insights into systemic compliance risks. Local SHAP explanations were generated for individual predictions to support case-level audits and regulatory inquiries. These quantitative explanations were compared with large language model-generated narratives to ensure consistency, factual grounding, and alignment with compliance terminology. The evaluation confirmed that the system produces transparent, traceable, and regulator-ready explanations suitable for operational deployment.

Results and Discussion

The proposed framework was evaluated using both historical data from the UCI Bank Marketing Dataset and simulated real-time transaction streams derived from temporal feature augmentation. This hybrid evaluation strategy enabled assessment of model performance under static and dynamic compliance scenarios.

Table 2: Comparative Model Performance on Historical Data

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.89	0.84	0.78	0.81
Random Forest	0.91	0.87	0.83	0.85
Gradient Boosting	0.93	0.90	0.88	0.89

Table 3: Comparative Model Performance under Real-Time Simulation

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.86	0.81	0.74	0.77
Random Forest	0.89	0.85	0.81	0.83
Gradient Boosting	0.92	0.89	0.86	0.88

The results demonstrate that gradient boosting consistently outperformed alternative models in both historical and real-time settings. Its superior recall under streaming conditions indicates strong robustness to behavioral drift, which is essential for effective AML monitoring. Logistic regression, while highly interpretable, exhibited reduced sensitivity to rapid behavioral changes. The integration of SHAP-based explanations with large language model narratives proved effective in both static and real-time contexts, enabling compliance officers to understand not only what decision was made but also how recent customer behavior influenced the outcome.

Conclusion

In this study, we presented an expanded explainable artificial intelligence framework for banking compliance that integrates structured machine learning models with large language models to support AML and KYC decision-making. The proposed approach addresses a critical challenge in modern financial systems by jointly optimizing predictive performance, explainability, and real-time responsiveness.

The empirical findings demonstrate that ensemble-based models, particularly gradient boosting, consistently outperform traditional linear approaches across both historical and real-time simulation scenarios. The strong recall achieved under streaming conditions highlights the framework's ability to detect high-risk behavioral patterns while maintaining stability in the presence of behavioral drift. Feature attribution methods combined with language-based explanations enable transparent and auditable decision support without sacrificing analytical rigor.

This research contributes to the explainable AI literature by demonstrating how large language models can function as interpretive layers over structured risk models, translating complex analytical outputs into human-readable, regulator-aligned explanations. The proposed framework provides a scalable and resilient foundation for AI-driven compliance systems. Future research may extend this work by incorporating live transaction feeds, adaptive governance mechanisms, and cross-institutional learning to further enhance the effectiveness of confirmable and trustworthy banking compliance solutions.

Reference:

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

2. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable artificial intelligence in credit risk management. *Computational Economics*, *57*(1), 203–216. https://doi.org/10.1007/s10614-020-10042-0

3. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

4. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

5. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

6. Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Leanpub.

7. UCI Machine Learning Repository. (2014). *Bank marketing data set*. University of California, Irvine. https://archive.ics.uci.edu

8. Umam, S., & Razzak, R. B. (2024, October). Linguistic disparities in mental health services: Analyzing the impact of spanish language support availability in saint louis' region, Missouri. In APHA 2024 Annual Meeting and Expo. APHA.

9. Adams, R., Grellner, S., Umam, S., & Shacham, E. (2023, November). Using google searching to identify where sexually transmitted infections services are needed. In APHA 2023 Annual Meeting and Expo. APHA.

10. Umam, S., & Razzak, R. B. (2025, November). A 20-Year Overview of Trends in Secondhand Smoke Exposure Among Cardiovascular Disease Patients in the US: 1999–2020. In APHA 2025 Annual Meeting and Expo. APHA.

11. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.

12. [12] Razzak, R. B., & Umam, S. (2025, November). A Place-Based Spatial Analysis of Social Determinants and Opioid Overdose Disparities on Health Outcomes in Illinois, United States. In APHA 2025 Annual Meeting and Expo. APHA.

13. [13] Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. PLoS One, 20(7), e0328109.

14. [14] Estak Ahmed, An Thi Phuong Nguyen, Aleya Akhter, KAMRUN NAHER, & HOSNE ARA MALEK. (2025). Advancing U.S. Healthcare with LLM–Diffusion Hybrid Models for Synthetic Skin Image Generation and Dermatological AI. *Journal of Medical and Health Studies*, *6*(5), 83-90. https://doi.org/10.32996/jmhs.2025.6.5.11

15. [15] Nitu, F. N., Mia, M. M., Roy, M. K., Yezdani, S., FINDIK, B., & Nipa, R. A. (2025). Leveraging Graph Neural Networks for Intelligent Supply Chain Risk Management in the Era of Industry 4.0. *International Interdisciplinary Business Economics Advancement Journal*, *6*(10), 21-33.

16. [16] Siddique, M. T., Uddin, M. N., Gharami, A. K.,

Khan, M. S., Roy, M. K., Sharif, M. K., & Chambugong, L. (2025). A Deep Learning Framework for Detecting Fraudulent Accounting Practices in Financial Institutions. *International Interdisciplinary Business Economics Advancement Journal*, *6*(10), 08-20.

17. [17] Mia, M. M., Al Mamun, A., Ahmed, M. P., Tisha, S. A., Habib, S. A., & Nitu, F. N. (2025). Enhancing Financial Statement Fraud Detection through Machine Learning: A Comparative Study of Classification Models. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(09), 166-175.

18. [18] Akhi, S. S., Ahamed, M. I., Alom, M. S., Rakin, A., Awal, A., & Al Mamoon, I. (2025, July). Boosted Forest Soft Ensemble of XGBoost, Gradient Boosting, and Random Forest with Explainable AI for Thyroid Cancer Recurrence Prediction. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.

19. [19] Alom, M. S., Akhi, S. S., Borsha, S. N., Mia, N., Tamim, F. S., & Nabin, J. A. (2025, July). Federated Machine Learning for Cardiovascular Risk Assessment: A Decentralized XGBoost Approach. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.

20. [20] Akhi, S. S., Rahaman, M. A., & Alom, M. S. An Explainable and Robust Machine Learning Approach for Autism Spectrum Disorder Prediction.

21. [21] Rabbi, M. A., Rijon, R. H., Akhi, S. S., Hossain, A., & Jeba, S. M. (2025, January). A Detailed Analysis of Machine Learning Algorithm Performance in Heart Disease Prediction. In *2025 4th International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 259-263). IEEE.

22. Mujiba Shaima, Mazharul Islam Tusher, Estak Ahmed, Sharmin Sultana Akhi, & Rayhan Hassan Mahin. (2025). Machine Learning Techniques and Insights for Cardiovascular or Heart Disease Prediction. *Academic International Journal of Engineering Science*, *3*(01), 22-35.