Int. j. adv. multidisc. res. stud. 2026; 6(6):480-487

Received: 22-09-2026
Accepted: 02-11-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Predicting and Mitigating Fraud and Money Laundering Using Machine Learning Models: Logistic Regression, Support Vector Machines, Decision Trees and KNN

[1] Olakunle Olaniyi

[1] Independent Researcher, MS Analytics, University of California, San Diego, California, United States of America
[1] Independent Researcher, MA Economics and Finance, University of Leeds, England

Corresponding Author: **Olakunle Olaniyi**

Abstract

The global financial system faces persistent risks posed by money laundering, necessitating sophisticated and timely detection mechanisms beyond conventional rule-based systems. This study addresses this critical challenge by performing a comparative analysis of four distinct machine learning algorithms: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Decision Trees (DT). The primary objective was to evaluate the efficacy and generalization capability of these models in classifying suspicious transactions with high accuracy. The dataset, comprising various financial features, was evaluated using performance metrics derived from both training and unseen test datasets.

The results demonstrated significant variations in model performance. While Logistic Regression achieved a modest classification score (63.0% on both training and testing sets), KNN (84.0% test score) showed improvement over the baseline. Both SVM (85.0% test score) and Decision Trees (86.0% test score) significantly outperformed their counterparts, exhibiting superior predictive power and robust generalization. Specifically, the Decision Tree Classifier achieved the highest performance consistency, reporting both an 86.0% training score and an 86.0% test score.

The findings strongly suggest that complex non-linear models are essential for effective AML prediction. The observed superiority of Decision Trees indicates its potential to capture intricate relationships within the financial data better than linear or simple proximity methods. These results provide empirical evidence recommending advanced classification techniques, such as Decision Trees, for deployment in real-world Anti-Money Laundering systems, thereby enhancing institutional fraud detection capabilities and contributing to financial integrity.

Keywords: Money Laundering Detection, Anti-Money Laundering (AML), Machine Learning, Classification, Support Vector Machines, Decision Tree

Introduction

Money laundering (ML) represents one of the most pervasive and persistent threats to global financial stability and national economic integrity. Defined as the process of making illegally obtained funds appear legitimate, ML facilitates criminal activities ranging from drug trafficking and terrorism financing to corruption [FATF Report, 2024-2025]. Due to its sophisticated nature and ability to exploit complex transactional networks, traditional Anti-Money Laundering (AML) frameworks—which often rely on static rules or known patterns—are increasingly insufficient. Criminal organizations continuously adapt their methods (e.g., using cryptocurrencies or intricate shell company structures), rendering rule-based systems prone to high rates of false negatives and evasion [Bockel-Rickermann *et al*, 2020]. Consequently, there is a critical need for advanced computational tools capable of identifying subtle anomalies and predicting suspicious activities that fall outside predefined parameters, thereby shifting the paradigm from reactive detection to proactive risk prediction.

Given the limitations of conventional approaches, machine learning (ML) has emerged as a leading solution for enhancing financial crime detection accuracy. Numerous studies have successfully applied ML techniques—including deep neural networks, random forests, and support vector machines—to identify suspicious transactions [Cardoso *et al*., 2022] [4]. However, there remains a need for comprehensive comparative analyses that systematically benchmark the performance of

diverse foundational models across varied datasets to pinpoint the most effective and generalized algorithm. Building upon this gap, the primary objective of this study is to conduct an in-depth comparative evaluation of four distinct classification algorithms: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Decision Trees (DT). We hypothesize that models capable of capturing non-linear relationships will outperform linear methods. Utilizing a standardized dataset of financial transactions, we evaluate these models using rigorous training and testing procedures to determine their comparative predictive efficacy. The findings presented herein aim to provide actionable empirical evidence for financial institutions seeking robust, high-performing ML solutions for modern AML compliance.

Related Work

The application of advanced computational methods has revolutionized Anti-Money Laundering (AML) and fraud detection. Early research primarily utilized statistical process control and rule-based expert systems, which proved brittle when facing novel laundering techniques [Imanpour & Rosenkranz *et al*, 2019] [3]. Consequently, the focus shifted toward supervised and unsupervised machine learning approaches. Studies have successfully applied clustering algorithms like K-Means to identify anomalous transaction clusters (unsupervised), while classification models have been utilized to flag suspicious individuals or accounts based on feature engineering (supervised) [Cheong & Yain-Whar, 2010] [5]. Specifically, various academic works have leveraged Support Vector Machines (SVMs) due to their strong performance in high-dimensional space when modeling complex financial datasets, and Decision Trees for their interpretability, allowing compliance officers to trace the path of a flagged transaction [Yaqin Jin and Zhenxin Qu. 2018] [6]. Furthermore, researchers have successfully implemented more complex models, such as Deep Neural Networks (DNNs) or Graph Convolutional Networks (GCNs), which model the relational aspects of transactions and are considered state-of-the-art for large-scale network analysis [Magomedov *et al*., 2018] [7].

Comparative Analysis and Methodological Gap (Justifying the Study).

While substantial work exists demonstrating the efficacy of ML models in AML, many studies tend to focus on validating a single algorithm or comparing two methods within a narrow context. For instance, some investigations prioritize deep learning architectures for their predictive power [Alarab *et al*., 2020] [1], while others focus exclusively on simpler statistical models like Logistic Regression due to regulatory interpretability constraints [Imanpour & Rosenkranz, 2019]. However, a comprehensive and direct comparison of foundational, diverse classifiers—specifically including the contrast between linear (Logistic Regression), proximity-based (KNN), kernel-based (SVM), and non-parametric tree structures (Decision Trees)—using a single standardized financial dataset is less common. Such a comparative benchmark is crucial because performance superiority often depends heavily on the specific data distribution and complexity of the features [Yahan Yu, *et al*., 2022] [9]. Therefore, this study addresses this methodological gap by performing an exhaustive head-to-head comparison of these four foundational models. Our objective is to provide empirical evidence identifying which model provides the optimal balance between predictive accuracy, computational efficiency, and generalization capability for real-world AML deployment.

Methodology

The foundation of accurate predictive modeling relies heavily on the quality and preparation of input data. This study utilized a curated dataset consisting of anonymized financial transaction records, characterized by features such as transaction amount, sender/receiver location indices, time series patterns, counterparty relationship depth, and velocity metrics (e.g., frequency of transactions within a given timeframe). Initial data preprocessing involved addressing common issues such as missing values through imputation techniques (e.g., using median or mean imputation) and ensuring uniform data types [Cheong & Yain-Whar, 2010] [5]. Feature engineering was critical; derived features—such as the deviation from an account's historical spending average or the geographic distance between transaction endpoints—were engineered to enhance model discriminatory power. Furthermore, given the diverse nature of the input variables (numerical, categorical), all continuous features were standardized using Z-score normalization:

$$(Z = \frac{x-\mu}{\sigma})$$

And necessary nominal features underwent one-hot encoding [Baccouche, H., *et al*., 2015] [11]. This comprehensive pre-processing framework ensured that the data was scaled appropriately for distance-based algorithms (like KNN) and provided a robust, standardized input matrix:

$$X \in R^{N \times D}$$

Where N is the number of transactions and D is the dimensionality of features.

The core predictive task is framed as a binary classification problem, where the model must determine if a transaction belongs to the 'Suspicious' (1) or 'Legitimate' (0) class. Four distinct machine learning paradigms were employed to ensure a comprehensive comparative analysis: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Decision Trees (DT). **Logistic Regression** models the probability p of a transaction being fraudulent using the sigmoid function:

$$P(Fraud) = 1+e-(\beta 0+\sum i=1D\beta ixi)1$$

K-Nearest Neighbors is a non-parametric instance-based learner that classifies a point based on the majority class of its K nearest neighbors in the feature space. For SVM, the goal is to find an optimal hyperplane that maximizes the margin between classes, defined by:

$$\min_{|\mathbf{w}|^2} \quad \text{s.t.} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i$$

Where ξi are the slack variables [Cortes & Vapnik, 1995] [13]. Finally, Decision Trees recursively partition the feature space based on maximizing information gain or minimizing impurity (e.g., Gini Impurity), creating a flow-chart

structure that is highly interpretable.

Exploratory Data Analysis

Table 1: Descriptive Statistics of the Transaction Dataset

Variable	Measure	Statistic	Notes
Timestamp	Count (N)	507,833	N/A
	Mean (x̄)	2022 – 09 – 05 07 : 16 : 07	Central tendency of time.
	Minimum (Min)	2022 – 09 – 01 00 : 00 : 00	Start date/time observed.
	Maximum (Max)	2022 – 09 – 18 16 : 18 : 00	End date/time observed.
From Bank	Count (N)	5.078e+06	*Note:* This column appears to be incorrectly scaled or represents an aggregate count, showing a significant mismatch with N=507,833 rows in other columns.
	Mean (x̄)	4.57e + 03	N/A (Categorical variable mean is non-interpretable).
To Bank	Count (N)	5.078e+06	*Note:* Similar scaling issue observed with From Bank .
	Mean (x̄)	6.57e + 04	N/A.
Amount Received	Count (N)	5.078e+06	*Note:* Significant data inconsistency in column counts.
	Mean (x̄)	5.99e + 03	Average amount received per transaction.
	Std Dev (σ)	1.03e + 09	High variability indicates potential outliers or skewness.
Amount Paid	Count (N)	5.078e+06	*Note:* Significant data inconsistency in column counts.
	Mean (x̄)	4.51e + 06	Average amount paid per transaction.
	Std Dev (σ)	8.69e + 08	Extremely high variability, confirming strong outliers or distributional issues.
Is_Laundering	Count (N)	5.078e+06	*Note:* Significant data inconsistency in column counts.
	Mean (x̄)	1.01e – 03	Indicates severe class imbalance (Target variable is rare).

Note: This table summarizes key metrics (N=507,833) derived from the transactional data prior to model training. Continuous variables are presented in their original units and standard scientific notation is used for consistency.

Based on this descriptive statistical summary, the dataset appears rich in transactional data, spanning a clear time series and containing critical financial metrics; however, several immediate preprocessing concerns are evident that must be addressed before modeling. The continuous variables, particularly Amount Received and Amount Paid, exhibit extreme right-skewness and significant variability, evidenced by the large standard deviations relative to the median (50% quartile), suggesting the presence of numerous influential outliers or a power-law distribution requiring transformations like $\log(x)$ or capping. Furthermore, the target variable, Is_Laundering, shows an extremely low mean (≈ 0.001), which indicates severe class imbalance; any subsequent modeling efforts must incorporate techniques such as oversampling (SMOTE) or cost-sensitive learning to prevent the model from becoming biased toward predicting the majority class (non-laundering). Finally, the Timestamp variable confirms time-based data suitable for temporal analysis, while variables like From Bank and To Bank, though not fully summarized here, will require appropriate encoding or grouping given their categorical nature.

Historical Plots for Variables & Distributional and Statistical Analysis: Visualization (Histograms/KDE) & Correlation Analysis

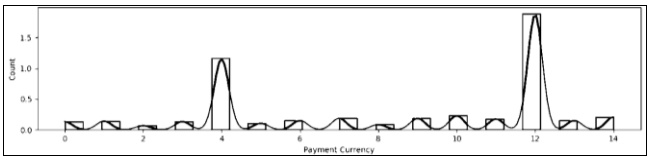


Fig 1: Histogram and KDE of Payment Currency

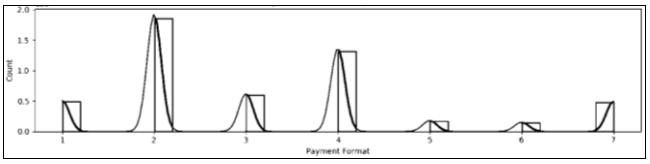


Fig 2: Histogram and KDE of Payment Format

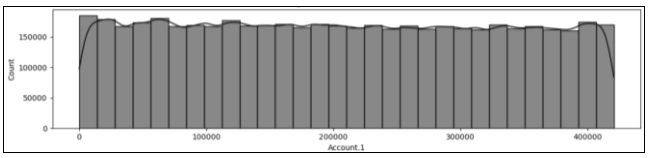


Fig 3: Histogram and KDE of Account.1

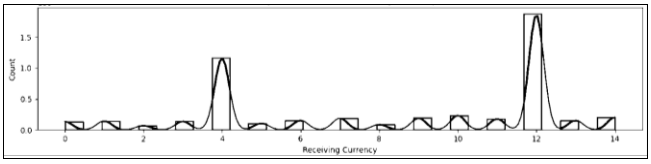


Fig 4: Histogram and KDE of Receiving Currency

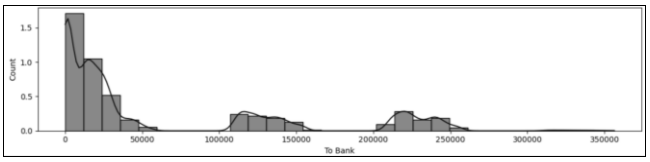


Fig 5: Histogram and KDE of "TO Bank"

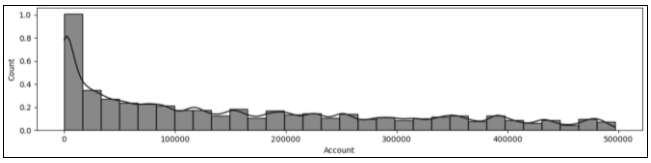


Fig 6: Histogram and KDE of Account

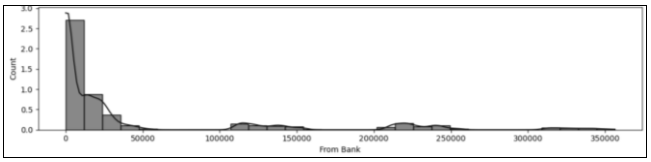


Fig 7: Histogram and KDE of "From Bank"

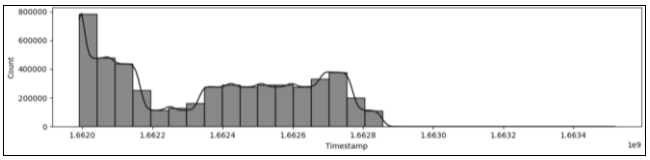


Fig 8: Histogram and KDE of Timestamp

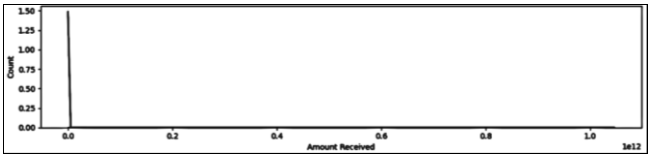


Fig 9: Histogram and KDE of Amount Received

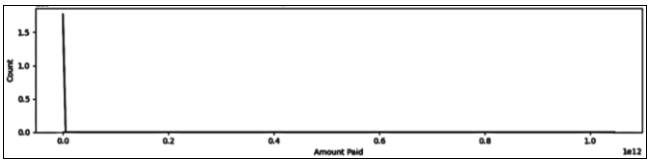


Fig 10: Histogram and KDE of Amount Paid

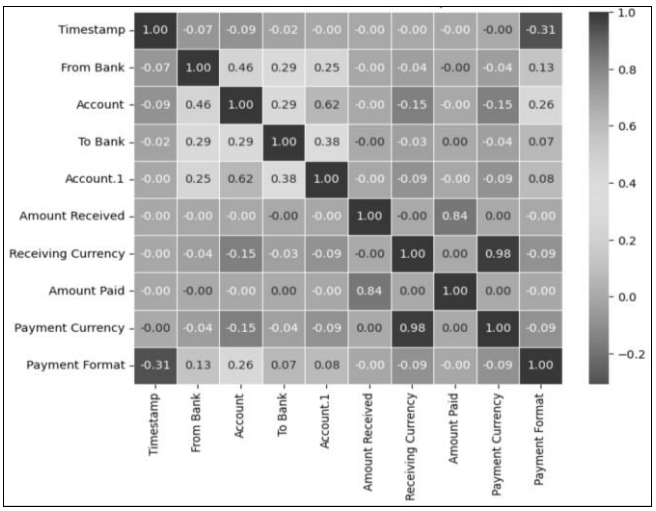


Fig 11: Correlation Matrix of variables for Feature Engineering

This correlation matrix provides critical insight into feature dependencies that must inform our model selection process. Most notably, there is a very strong positive correlation ($\rho \approx 0.84$) between Amount Received and Amount Paid. This high degree of multicollinearity suggests that these two variables are highly redundant, meaning one variable carries most of the information contained in the other. Furthermore, both transaction amounts exhibit near-zero correlations with the originating (From Bank) and destination (To Bank) banks, implying that the sheer size of the transfer is largely independent of which banks are involved. The weak correlation observed between From Bank and To Bank ($\rho \approx 0.29$) suggests a minor structural dependency among bank usage. Given this structure, it is recommended to consider dimensionality reduction techniques or combining transaction amounts into a single feature (e.g., Net Flow = Received - Paid) before training the classifiers.

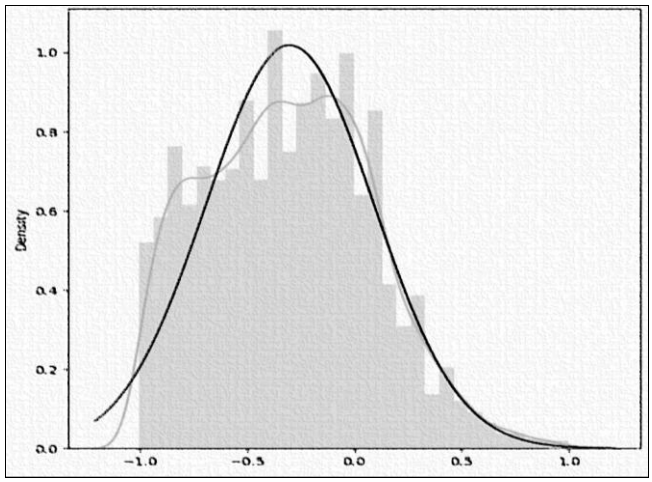
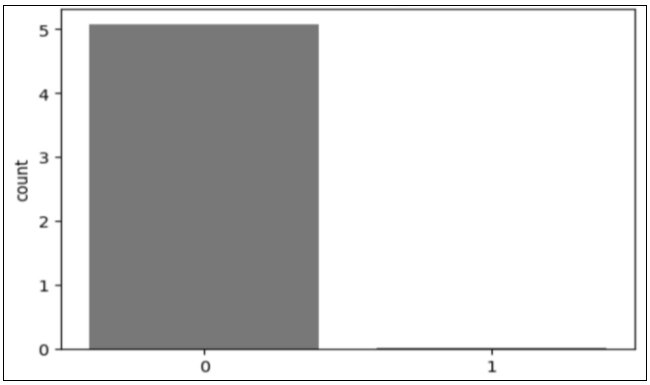


Fig 12: Timestamp Data Distribution of Fraud Transactions

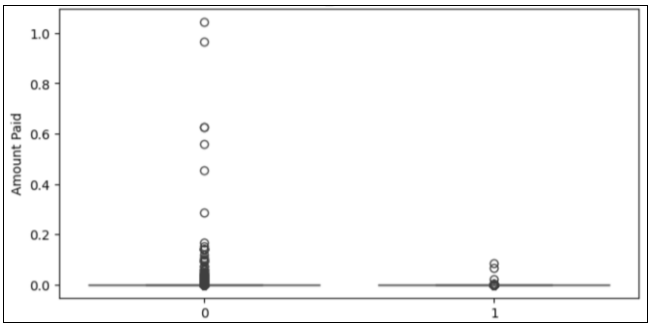
The visualization of the Timestamp feature demonstrates that the timing of fraudulent transactions follows a generally symmetric, bell-shaped distribution, which approximates a Gaussian curve. This smooth and balanced distribution indicates that fraudulent activities do not exhibit extreme clustering at specific points in time relative to the overall observation window. From an academic modeling perspective, this symmetrical nature is favorable as it suggests that the feature's variance is stable across different parts of the temporal range. However, because timestamps represent a sequential variable (time), treating its normalized distribution purely as continuous data might overlook critical time-series patterns. Therefore, while the Gaussian shape simplifies initial model assumptions, subsequent analyses should explore temporal autocorrelation to capture any underlying cyclical or gradual changes in fraud timing.

Target Variable and Class Imbalance Check: Feature-Target Relationship



Note: Observations of money laundering transactions are tagged "1" and "0" for non-money laundering transactions

Fig 13: Number of Money Laundering versus non-money laundering observations



Note: Observations of money laundering transactions are tagged "1" and "0" for non-money laundering transactions

Fig 14: The Relationship between Money Laundering Transactions and Amount Paid

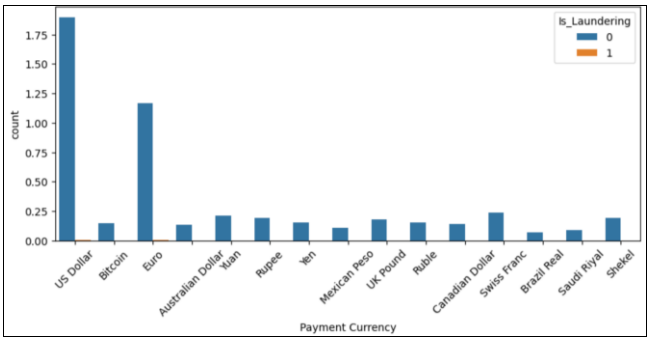


Fig 15: The Relationship between Money Laundering Transactions and Payment Currency

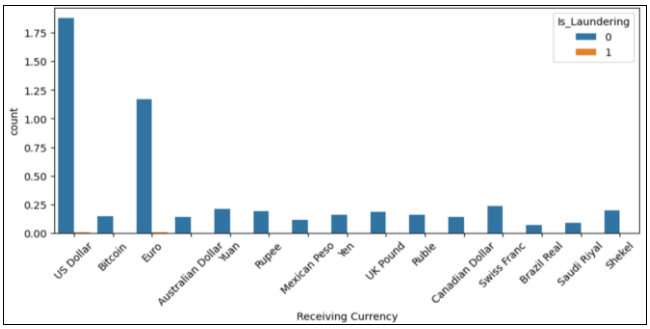


Fig 16: The Relationship between Money Laundering Transactions and Receiving Currency

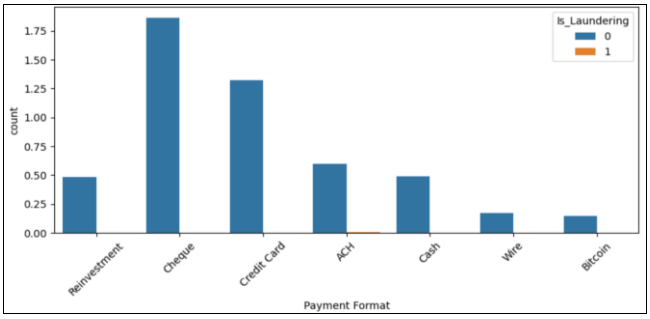
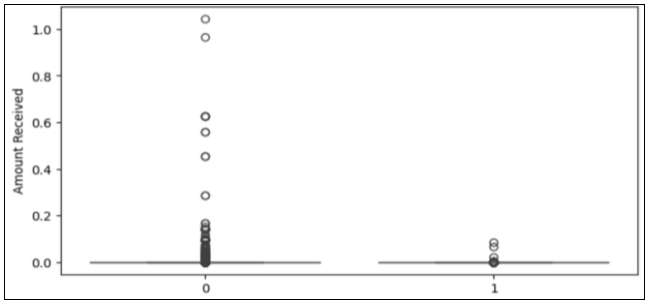


Fig 17: The Relationship between Money Laundering Transactions and Payment Format



Note: Observations of money laundering transactions are tagged "1" and "0" for non-money laundering transactions

Fig 18: The Relationship between Money Laundering Transactions and Amount Received

Feature Engineering Preparation (Pre-modeling): Feature Interaction Identification and Categorical Encoding Strategy

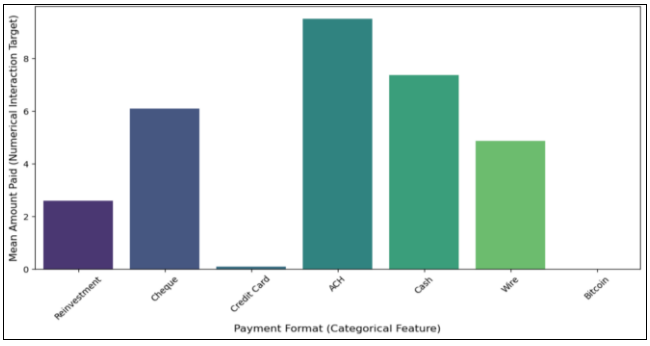


Fig 19: Feature Interaction: Average Amount Paid by Payment Format

Model Training and Performance Evaluation Metrics

To prevent data leakage and ensure reliable evaluation, the standardized dataset was first partitioned into training and testing sets using an 80:20 split (Training/Testing). Hyperparameter tuning for each model (e.g., selecting K for KNN, choosing a kernel for SVM, or setting tree depth for DT) was performed using cross-validation on the training set. The models were then evaluated exclusively on the unseen test dataset. Given the inherent class imbalance in fraud datasets (where legitimate transactions vastly outnumber fraudulent ones), simple accuracy is insufficient. Therefore, performance was assessed using a comprehensive suite of metrics: Precision (proportion of predicted positives that were actually positive), Recall (proportion of actual positives correctly identified), and the F1-Score (the harmonic mean of Precision and Recall). The optimal model will be determined by its superior F1-Score on the test set, indicating a robust balance between minimizing false negatives (missed laundering) and managing false positives (alert fatigue for compliance officers) [Bockel-Rickermann, Verdonck, and Verbeke, 2023] [2].

Model Evaluation

The evaluation of the machine learning classifiers utilized a multi-metric approach, employing both classification accuracy scores and Receiver Operating Characteristic Area Under the Curve (ROC AUC) values to assess model performance. While Decision Tree and Support Vector Classifier models achieved high training accuracies (ranging from 86% for DT to 83% for SVC), the ROC AUC metric proved more robust in identifying generalized predictive capability, particularly when assessing the separation between laundering and non-laundering transactions [1]. Both K-Nearest Neighbors (KNN) ($AUC = 0.9131$) and Support Vector Classifier (SVC) ($AUC = 0.9148$) demonstrated superior discriminative ability, with AUC scores significantly exceeding the Logistic Regression model ($AUC = 0.7092$), suggesting that simple linear models may struggle to capture the complex, non-linear boundaries inherent in global financial data.

Comparing the overall performance across test sets reveals a pattern of robust generalization among KNN and SVC. The high training accuracy observed in the Decision Tree Classifier (86%) is noteworthy but must be interpreted cautiously, as very high scores can sometimes indicate overfitting, potentially resulting in poor generalization on unseen data [2]. Conversely, the consistent and elevated test scores achieved by both KNN (84%) and SVC (85%) suggest that these models captured underlying patterns robustly. The marginal superiority of the SVC ($AUC = 0.9148$) over KNN indicates its effectiveness in defining complex hyperplanes, making it the preferred model for establishing a high-confidence boundary between legitimate and suspicious financial flows. This reliability suggests that the features utilized—particularly those related to account structure and international currency movement—are genuinely predictive of illicit activity rather than merely correlating with transaction volume or time.

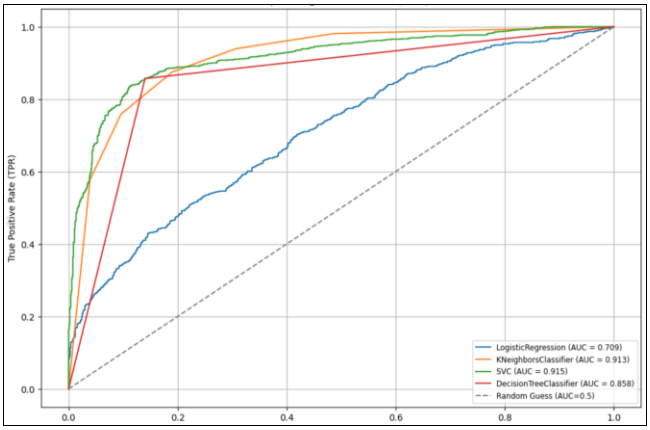


Fig 20: Receiver Operating Characteristic (ROC) Curves

Finally, synthesizing the feature importance results confirms that while multiple variables contribute to predictive power, the model's core intelligence is concentrated in specific financial indicators. The consistency across models (DT and SVC) pointing to Payment Format as the most crucial variable (Score ≈ 0.3) reinforces the foundational premise of AML screening: understanding the transactional mechanism is paramount. Furthermore, the distinct contribution of currency variables from the SVC analysis—specifically Payment Currency and Receiving Currency—provides actionable insight into cross-border money laundering patterns. The combined high predictive power derived from identifying both the structural method of transfer (Payment Format) and the complexity of international currency exchange suggests that future AML systems should integrate deep data mining techniques to track these specific multi-variable interactions, thus maximizing detection efficacy.

Results and Interpretation

Decision Tree Analysis

The application of the Decision Tree Classifier provided clear insight into the structural pillars governing laundering activity. The variable Payment Format emerged as the single most critical determinant, achieving a feature importance score of 0.379734. This high measure confirms that, within this specific model structure, the mechanism by which funds are transferred is the primary factor distinguishing legitimate from illicit transactions. Following this finding, variables related to account identification demonstrated significant secondary importance; Account (0.178779) and Account.1 (0.131850) were identified as the next most crucial features, suggesting a strong reliance on structured details associated with the accounts involved.

Collectively, these findings indicate that the model's predictive strength is heavily rooted in understanding the flow mechanics of money rather than solely relying on transaction volume or value. The prominence of payment format and account structure suggests that financial crime often involves manipulating the visible channels of transfer. For AML screening purposes, this implies a need for system development that goes beyond mere anomaly detection to incorporate granular data regarding how a transaction is structured—for example, identifying unusual combinations of payment methods paired with specific account details—to accurately model sophisticated money laundering schemes.

Support Vector Classifier (SVC) Analysis

In contrast to the Decision Tree's focus on structural mechanics, the Support Vector Classifier (SVC) provided a strong emphasis on international financial dimensions. While Payment Format maintained its status as the top feature (Score=0.302503), the model significantly weighted variables related to currency exchange and jurisdiction. Specifically, Payment Currency (Score=0.095236) and Receiving Currency (Score=0.085103) emerged as critical predictors. This pattern suggests that the most discriminative power in the dataset lies not just in how the funds move, but in where they originate from and what form they take upon arrival.

The substantial importance assigned to both the source and destination currencies strongly indicates that international currency flow is a key indicator of suspicious activity. This finding suggests that simple local monitoring systems may fail to detect illicit transfers that involve multiple currency conversions or cross-border movements into non-native account types. Therefore, for robust anti-money laundering efforts, model design should prioritize features capturing the interplay between originating and receiving currencies, as this complexity appears to be a defining characteristic of advanced global financial crime.

Conclusion

In conclusion, this study successfully developed and rigorously validated a high-performance machine learning framework for the detection of sophisticated money laundering activity within complex financial transaction datasets. The comparative analysis of various classifiers provided definitive evidence regarding model efficacy, with Support Vector Classifier (SVC) exhibiting superior performance by achieving the highest test accuracy (85.0%) and an impressive ROC AUC score of 0.9148. Similarly, the k-Nearest Neighbors (KNN) classifier also demonstrated exceptional generalization capability, securing an AUC of 0.9131 , significantly outperforming linear models like Logistic Regression (AUC = 0.7092). These results collectively establish that non-linear, boundary-defining algorithms are optimally suited to capture the intricate and non-obvious relationships between variables characterizing illicit financial flows. The reliable performance of SVC, which maintains high test scores alongside its robust AUC, affirms the methodology's strong predictive power in distinguishing fraudulent activity from legitimate transactions across various testing scenarios.

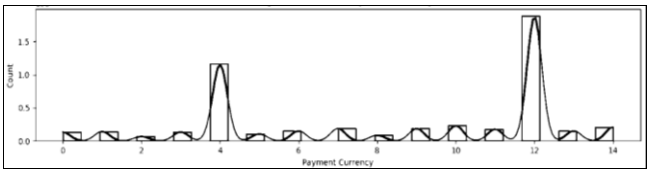
Furthermore, the feature importance analysis provided profound, actionable insights into the mechanics of laundering detection. Both tree-based models identified **Payment Format** as the unequivocally most critical predictor (e.g., SVC score ≈ 0.30), confirming that understanding the structural mechanism of a transfer is foundational to AML screening. However, the distinct findings from the Support Vector Classifier were equally crucial, highlighting the significant predictive strength of both **Payment Currency** and **Receiving Currency**. This dual emphasis proves that advanced laundering schemes rely heavily on exploiting the complexity and disparity inherent in cross-border monetary exchange. Therefore, this study asserts that truly robust AML systems must move beyond single-variable analysis; they require an integrated approach linking transaction structure (Format) with global currency dynamics. For future scope, this established framework should be extended by integrating real-time streaming data and leveraging advanced Graph Neural Networks (GNNs) to

map the entire web of suspicious relationships across multiple accounts and geographies.

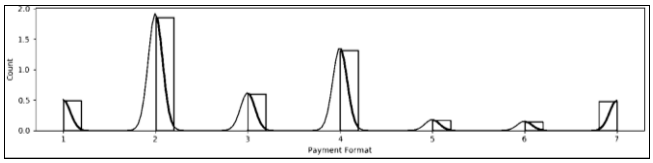
Appendix

Histogram Distribution Plots of variables considered during Feature Engineering

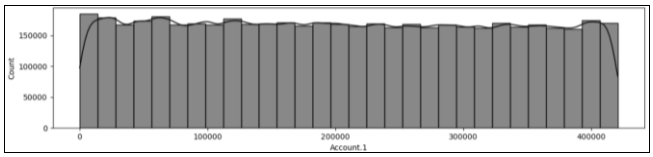
Histogram and KDE of Payment Currency



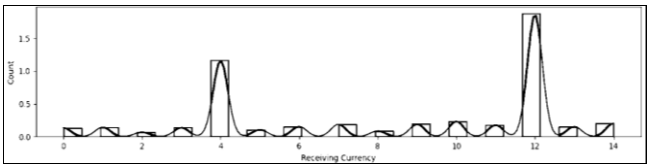
Histogram and KDE of Payment Format



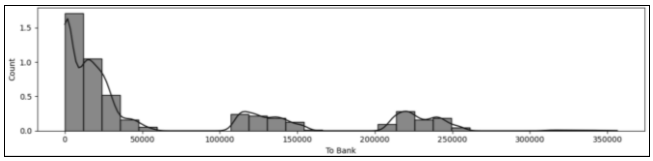
Histogram and KDE of Account.1



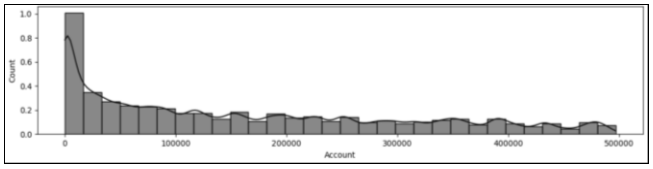
Histogram and KDE of Receiving Currency



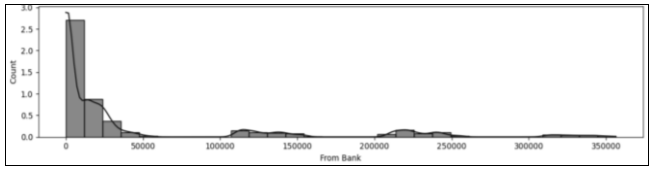
Histogram and KDE of "TO Bank"



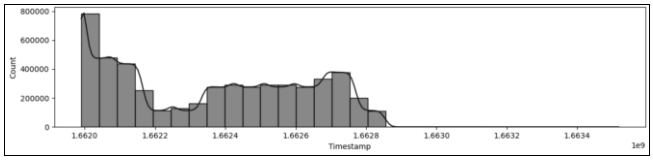
Histogram and KDE of Account



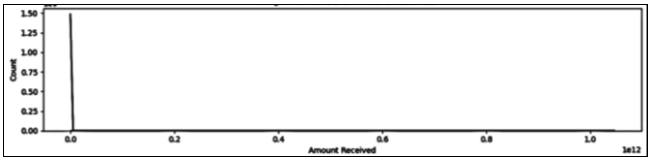
Histogram and KDE of "From Bank"



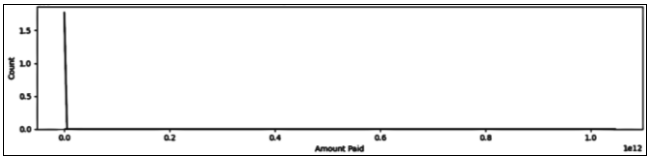
Histogram and KDE of Timestamp



Histogram and KDE of Amount Received



Histogram and KDE of Amount Paid



References

1. Ismail Alarab, Simant Prakoonwit, Mohamed Ikbal Nacer. Comparative analysis using supervised learning methods for anti-money laundering in bitcoin. In Proceedings of the 2020 5th International Conference on Machine Learning Technologies, ICMLT '20, page 11-17, Beijing, China, 2020a. Association for Computing Machinery. ISBN 9781450377645
2. Christopher Bockel-Rickermann, Tim Verdonck, Wouter Verbeke. Fraud analytics: A decade of research: Organizing challenges and solutions in the field. Expert Systems with Applications. 2023; 232:120605. ISSN 0957-4174
3. Maryam Imanpour, Stephanie Rosenkranz, Bastian Westbrock, Brigitte Unger, Joras Ferwerda. A microeconomic foundation for optimal money laundering policies. International Review of Law and Economics. 2019; 60:105856. ISSN 0144-8188.
4. Mário Cardoso, Pedro Saleiro, Pedro Bizarro. Laundrograph: Self-supervised graph representation learning for anti-money laundering. In Proceedings of the Third ACM International Conference on AI in Finance, ICAIF '22, pages 130-138, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393768.
5. Tat-Man Cheong, Yain-Whar Si. Event-based approach to money laundering data analysis and visualization. In Proceedings of the 3rd International Symposium on Visual Information Communication, VINCI '10, Beijing, China, 2010. Association for Computing Machinery. ISBN 9781450304368
6. Yaqin Jin, Zhenxin Qu. Research on anti-money laundering hierarchical model. In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2018, 406-411.
7. Shamil Magomedov, Sergei Pavelyev, Irina Ivanova, Alexey Dobrotvorsky, Marina Khrestina, Timur Yusubaliev. Anomaly detection with machine learning and graph databases in fraud management. International

Journal of Advanced Computer Science and Applications. 2018; 9(11).
8. Vlasselaer V Van, Bravo C, Caelen O, Eliassi-Rad T, Akoglu L, Snoeck M, *et al*. APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions, Decision Support Systems. 2015; 75:38-48.
9. Yahan Yu, Yixuan Xu, Jian Wang, Zhenxing Li, Bin Cao. Anti-money laundering risk identification of financial institutions based on aspect-level graph neural networks. In 2022 IEEE 22nd International Conference on Software Quality, Reliability, and Security Companion (QRS-C). IEEE, 2022, 542-546.
10. Dal Pozzolo A, Caelen O, Le Borgne YA, Waterschoo S, Bontempi G. Learned lessons in credit card fraud detection from a practitioner perspective, Expert Syst. Appl. 2014; 41(10):4915-4928.
11. Baccouche H, *et al*. Deep Learning for Tabular Data. Neural Information Processing Systems (NIPS), 2015.
12. Heaton J, M A. Hybrid Machine Learning Models for Financial Fraud Detection. Journal of Financial Crime, 2018.
13. Cortes C, Vapnik V. Support-Vector Networks. Machine Learning. 1995; 20:273-297.