JOURNAL OF TECHNOLOGY MANAGEMENT AND BUSINESS
ISSN: 2289-7224 e-ISSN: 2600-7967

JTMB

Vol. 12 No. 2 (2025) 98-106
https://publisher.uthm.edu.my/ojs/index.php/jtmb

Logistic Regression Modelling for Anti Money Laundering Detection for Reporting Institutions

Nor Adlin Sofea Nor Hairudin[1], Noryanti Muhammad[1]*, Md Ibnu Hisyam Mohamad[2], Wan Zarazillah Wan Abu Bakar[2]

[1] *Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300, Kuantan, Pahang, MALAYSIA*

[2] *Suruhanjaya Syarikat Malaysia (SSM), No 7, Jalan Stesen Sentral 5, Kuala Lumpur Sentral, 50623 Kuala Lumpur, MALAYSIA*

*Corresponding Author: noryanti@umpsa.edu.my
DOI: https://doi.org/10.30880/jtmb.2025.12.02.007

Article Info

Received: 22 August 2025
Accepted: 1 December 2025
Available online: 20 December 2025

Keywords

Anti money laundering, logistic regression, reporting institutions

Abstract

Money laundering is a significant threat to global financial systems and national security, and Malaysia's Anti-Money Laundering, Anti-Terrorism Financing and Proceeds of Unlawful Activities Act 2001 (AMLA 2001) mandates reporting institutions, such as banks and casinos, to identify and report suspicious transactions. Traditional rule-based approaches to Anti-Money Laundering (AML) detection face limitations, including high false-positive rates and a lack of adaptability to evolving criminal tactics. This research aims to address these challenges by identifying key parameters within the AMLA 2001 framework, developing a detection model that incorporates these parameters and leverages logistics regression techniques, and evaluating its accuracy, precision, recall, and capacity to reduce false positives. Employing a multi-pronged methodology, including literature review, data collection, model development, and evaluation, the study uncovered critical AMLA 2001 parameters predictive of illicit activities and created a model that enhances detection efficiency and accuracy. The findings have practical implications for any reporting institutions and regulatory bodies, contributing significantly to efforts to combat money laundering and safeguard the integrity of country's financial system.

1. Introduction

Money laundering refers to the process of concealing the origins of illicitly obtained funds, making them appear legitimate. In contrast, Anti-Money Laundering (AML) encompasses the legal and regulatory measures aimed at preventing such activities (Berkan Oztasa, 2024). The Anti-Money Laundering Act (AMLA) plays a crucial role in identifying suspicious financial behaviors, such as large, unexplained cash withdrawals or frequent small transactions designed to avoid detection. Other red flags include fund transfers to or from financially unstable countries, which may signal cross-border laundering schemes. Additionally, associations with individuals who have criminal backgrounds can also raise serious concerns about potential illicit involvement (Hamin et al., 2015).

Detecting money laundering and terrorism financing remains a persistent challenge due to the increasingly sophisticated techniques used to obscure financial trails (Plaksiy et al., 2018). Traditionally, rule-based systems have been employed to flag suspicious transactions. However, these systems often produce high false-positive rates, resulting in increased compliance costs and potentially overlooking high risk cases (Tertychnyi et al., 2020; Tiwari et al., 2020). As data science and machine learning technologies advance, they offer promising solutions

for improving the accuracy and efficiency of AML detection systems. Machine learning models including logistic regression, bagging, random forest, boosting, and stacking can be utilized to detect patterns in large-scale transaction data and more effectively identify high risk entities (Plaksiy et al., 2018; Tiwari et al., 2020).

In Malaysia, efforts to combat financial crime have intensified through the enforcement of the Anti-Money Laundering, Anti-Terrorism Financing and Proceeds of Unlawful Activities Act 2001 (AMLA). In general, this research aims to enhance the enforcement of Malaysia's Anti-Money Laundering, Anti-Terrorism Financing and Proceeds of Unlawful Activities Act 2001 (AMLA) by developing a more precise, data-driven inspection targeting system. Recognizing the limitations of traditional rule-based methods, which generate many false alerts, the study leverages company level compliance data specifically records on BO, CDD, and RKP. By applying statistical modelling, particularly logistic regression techniques, to this structured dataset, the research seeks to build a decision support system that can more accurately identify noncompliant or high risk companies for inspection, thereby improving the efficiency and effectiveness of the country's efforts to combat money laundering and terrorism financing

2. Literature Review

Money laundering is known as an activity where it hides the sources of illegal money while Anti Money Laundering (AML) is an action to combat such corruption (Berkan Oztasa, 2024). A cornerstone of AML efforts is the Anti-Money Laundering Act (AMLA), which focuses on identifying red flag behaviours. These red flags are known as essentially heuristic rules derived from historical criminal typologies. For instance, "structuring" (frequent small transfers to avoid reporting thresholds) and transactions with high-risk jurisdictions are well documented indicators in AML literature (e.g., Financial Action task Force (FATF) is recommendations). Anti money Laundering Act (AMLA) focus on identifying unusual activities such as large sum of money being withdrawn without clear reference or justifiable reason (Berkan Oztasa, 2024). Another example is the frequent transfer of small amount of money to avoid detection by the authorities, sending money to financially weak countries or receiving money from weak financial country also is considered suspicious activities, as it may be linked to money laundering. Additionally, individuals who personally engage or dealing with those having criminal background raise further concern of potential involvement in illegal activities (Hamin et al., 2015).

Detecting money laundering and terrorism financing has known to been challenging due to sophisticated techniques used by criminals to conceal their activities (Plaksiy et al., 2018). There are two primary approaches been used to identify suspicious organization that involve in money laundering and terrorism financing which is rule based approach and mathematical and machine learning-based approach (Tiwari et al., 2020). The rule-based approach represents a knowledge driven paradigm, where expert defined "if-then" rules are used for detection. While transparent, its main limitation, as noted by Plaksiy et al. (2018), is its static nature and high false positive rate, leading to alert fatigue and high operational costs. Traditional detection systems often rely on rule based models that flag suspicious transactions, yet these systems may suffer from high false positive rates, burdening financial institutions with significant compliance costs and missed opportunities to identify high risk transactions (Plaksiy et al., 2018).

As the machine learning and data science continue to rise with rapid evolution, there is coming new opportunities to develop a seamless Anti-Money Laundering detection model with higher accuracy and efficiency to analyse these illicit activities in large transaction datasets (Plaksiy et al., 2018). To detect suspicious activities with automate decision-making, mathematical and machine learning techniques with advanced algorithms being implemented to detect patterns in data. Machine learning models operate on a predictive paradigm(Nasdaq Verafin, 2019). By learning patterns from historical data, they can identify complex, nonlinear relationships that are difficult to encode with static rules. This shift is from explicit programming to implicit pattern recognition (Jullum et al., 2020). Through the use of systematic machine learning techniques, machine learning models like logistic regression (Md Sum et. al, 2020), bagging, random forest, boosting and stacking will improve the accuracy of identifying high-risk and suspicious companies or organization through evaluating variables and utilizing models (Tiwari et al., 2020). Each model has a specific theoretical strength, for example, it is known that logistic regression provides probabilistic and interpretable outcomes, while ensemble methods like random forest (bagging) and boosting improve predictive power by combining multiple models to reduce variance or bias, respectively.

Malaysia has intensified its effort to combat financial crimes, particularly money laundering and terrorism financing in recent years through the regulatory of the Anti-Money Laundering, Anti-Terrorism Financing and Proceeds of Unlawful Activities Act 2001 (AMLA). This enforcement framework has been significant important in addressing financial issues as evidenced by several high-profile cases. This establishes the practical context for the research, situating the application of the aforementioned theories and models within a specific national regulatory environment. The effectiveness of this framework can be evaluated through its application in high profile cases, which serve as real world validation of the system.

J. of Technology Management and Business Vol. 12 No. 2 (2025) p. 98-106

3. Methods

This section describes the research techniques used to examine the detection of Anti-Money Laundering activities within Reporting Institution based on the frameworks outlined in the AMLA 2001. This study uses AMLA parameters to create and improve detection model that is suited to the operational and regulatory requirements of reporting institutions. The study adopts hybrid design combining descriptive, diagnostic and predictive analytics to explore how AMLA parameters can be used to detect potentially suspicious behaviour within the companies. The focus is on unsupervised learning using Random Forest model to build an efficient AML detection model requires a thorough understanding of the study design, data source, data processing, feature selection, and model development stages, all of which are covered in this section.

Through a systematic examination of AMLA characteristic, this methodology seeks to improve compliance efforts within reporting institution by facilitating the early detection of possible money laundering activities. This section lays the groundwork for putting into practice a scalable system going over each stage which is shown in Figure 1.

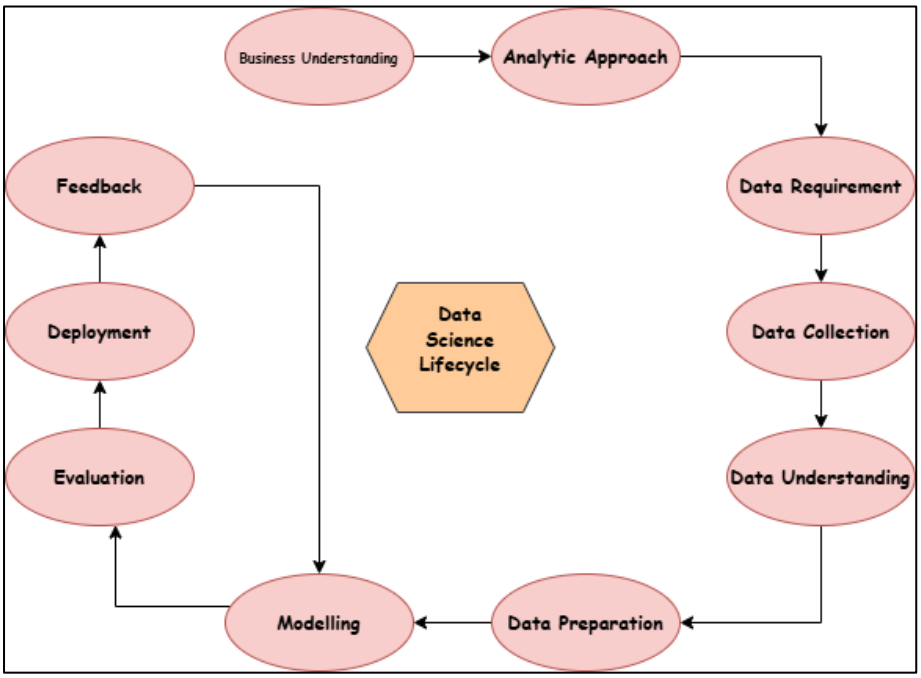

Fig. 1 *Data science lifecycle (Yadav, 2020)*

This study aims to evaluate the effectiveness of a machine learning based detection model designed to help reporting institutions comply with Malaysia's Anti-Money Laundering, Anti- Terrorism Financing and Proceeds of Unlawful Activities Act 2001 (AMLA). Institutions such as company secretaries, lawyers, and accountants are required to identify and report suspicious financial activities, but they often face challenges due to the growing complexity of money laundering schemes, high transaction volumes, and increasing regulatory pressure. To address these issues, this research proposes a data driven model that utilizes AMLA compliance parameters to identify irregularities and potential non-compliance. The model leverages unsupervised learning techniques to detect abnormal patterns in transactional and organizational data, which are then categorized into risk levels. The analytic approach is layered, beginning with descriptive analytics to summarize overall compliance trends, followed by diagnostic analytics to uncover key patterns linked to high-risk behaviour, and concluding with predictive analytics to forecast future noncompliant cases. By combining statistical methods with domain specific regulatory knowledge, the proposed model aims to improve detection accuracy, reduce false positives, and support more effective monitoring and enforcement within the reporting ecosystem.

3.1 Data Collection

The data collection process is to build and effective Anti-Money Laundering detection model specifically for reporting institutions. The methods and process of collecting data may vary across institutions, depending on internal compliance framework. However, this collected data follows the standard and requirements outlined in the Malaysia's Anti-Money Laundering, Anti- Terrorism Financing Act 2001 (AMLA). In this study, the data has

been simulated based on requirement needed for this research and has been guided by Companies Commission of Malaysia used from its regulatory review of certain companies. The institutions are responsible for ensuring that the sources data are reliable and data integrity is guaranteed throughout the procedures. The key features are list as in Table 1.

Table 1 *Dataset description*

Features	X	Data Type	AMLA Context
No. Syarikat		Numeric	Companies' registration number
Nama_Syarikat		String	Companies' registration name
Bayaran_Umpukan_TS	x1	Binary	Lump-Sum Payment Transaction Screening
Dokumen_sokongan	x2	Binary	Cosec keep all copies of client
Dokumen_sokongan_pemegang_ luar	x3	Binary	Supporting documents for foreign shareholders
Dokumen_sokongan_legalperson	x4	Binary	Supporting documents verifying the legal person's structure and ownership
Due_Payable_TS	x5	Binary	Payments or financial obligations are screened
EDD_risiko_tinggi_ CRP	x6	Binary	Whether enhanced due diligence was applied for high-risk clients.
Hasil_pemprofilan_ risiko_CRP	x7	Binary	Outcome/Results of the client's risk profile (1=low risk / 0=high risk)
Kelulusan_penguru san_diperoleh_CRP	x8	Binary	Whether management approval is obtained for customer risk profiling
Medium_CRP	x9	Binary	Medium/Channel through which customer risk profile is obtained
Medium_maklumat pengarah	x10	Binary	Medium/Channel through which director information is obtained
Menyimpan_maklumat_legal person	x11	Binary	Whether the institutions keep records on legal person
Pemegangsaham_tidak_berdaftar	x12	Binary	Whether there are unregistered (informal) shareholders.

3.2 Data Preparation

The data preparation phase involves cleaning, transforming and standardization data for optimal performance in the detection model. Initially, the raw dataset is derived from interview and documentary reviews during Anti-Money Laundering and Counter Terrorism Financing evaluation of companies' secretaries and was compiled into a structured tabular format consist of binary indicators across AML compliances variables. Each row represents a company, and each binary column indicates whether the company complies as 1 or does not comply as 0 with a specific AMLA requirement. Therefore, no encoding was necessary due to uniform binary structure and outliers were not present due to categorical nature of data. Data cleaning includes handling missing values, correcting inconsistencies, and removing duplicates. Missing values in this dataset are considered as conditionally missing values, which do not indicate as entry errors. Instead, these missing values exist due to certain companies not being applicable on specific compliance variables because it fulfilled prior requirements. Hence the imputation 'df.fillna(1)' was made to mark the missing values as compliant as 1. The dataset was first inspected with 'df.info()' to understand the structure, data types and presence of missing values. All compliance columns are converted into integer '.astype(int)' to ensure columns are in the proper integer format which help prevent potential issues during model training and evaluation stages.

3.3 Risk Score Development for Anti-Money Laundering

This section outlines the systematic development of composite risk scoring features intended to quantify the AMLCFT compliance level of companies. Initially a subset of binary compliance indicators was selected form the dataset, representing key Anti-Money laundering (AML) control areas such as Know Your Customer (KYC), Record keeping (RKP) Beneficial Ownership Transparency (BO) and Enhanced Due Diligence (EDD). Non-compliance fields like company identifiers suchas 'No.Syarikat' and 'Nama_syarikat' were excluded to focus purely on binary columns.

Hence, all binary columns were converted into integer format '0' for non-compliant, '1' for compliant to ensure compatibility with correlation analysis and weighting procedures, which adapted from Lim et al (2024).

Next, Spearman's correlation was computed to assess the relative importance of each binary column in determining AML risk. Each feature was assessed for its correlation strength against selected proxy indicators representing overall AML performance. The absolute correlation values of all other features were extracted and normalized such that their sum equals 1. This process was repeated for multiple key indicators (19 compliance variables) in total, reflecting various AML dimensions. Therefore, mean correlation weight across all targets was then computed for each feature which served as the base weight, indicating how strongly each compliance parameter aligns with overall AML observance.

A penalty weight was applied based on each feature's compliance frequency across institutions to prevent risk dilution from frequently compliant indicators. Each binary variable, the frequency of compliance and non-compliance was calculated and normalized. A penalty score was computed as in Equation (1) to penalize common indicators that may not strongly differentiate low-risk and high-risk entities. Finally, the final weight for each binary variable was the product of its average correlation weight and frequency-based penalty.

$$\boldsymbol{WWpenalty} = \mathbf{1} - \left(\frac{\boldsymbol{Frequency\ of\ compliance}}{\boldsymbol{Total}}\right) \quad (1)$$

According to Lim et. al, (2024) a composite risk score was computed using the finalized feature weights as in Equation (2) for each entities using a weighted sum of its binary compliance indicators. Additionally, raw risk scores were normalized to a scale of 0 to 10 utilizing 'minmaxscaler' to ensure interpretability and comparability.

$$Risk\ Score = \sum_{j=1}^{n} x_{ij} \cdot w_j \quad (2)$$

where $x_{iii} \in \{0,1\}$ = The compliance status of institution *i* on feature *jj* and w_{iii} = Final weight *jj*.

3.4 Risk Score Classification

The classification of the risk score developed from previous section is calculated by considering a rule- based classification approach, which was implemented to assign each entity into distinct AML risk categories. This stratification facilitates clearer interpretation of the predicted risk and supports supervisory prioritization of regulatory bodies. Thus, each entity's normalized risk score was categorized into risk levels as shown in Table 2. The thresholding approach was determined based on expert judgment and aligned with risk-based supervision guidelines, where entities are grouped according to the severity of their compliance gaps. This facilitates a more efficient allocation of regulatory resources and allows for proportionate supervisory action based on institutional risk exposure.

Table 2 *Risk level*

Group Code	Average	Level
1	0.00 – 3.00	Low
2	3.01 – 6.00	Medium
0	6.01 – 10.00	High

In addition, to guaranteed risk level into actionable supervisory outcomes, and inspections recommendation layer, was introduced. Institutions were mapped to inspection priorities as shown in Table 3. This classification enables the regulator to make proactive, risk-sensitive decisions in line with AML/CFT supervisory mandates. For instance, high-risk entities may undergo detailed audits or EDD, while low-risk entities may be exempted from immediate oversight.

Table 3 *Inspection logic*

Group Code	Level	Inspect
1	Low	No Immediate Action Required
2	Medium	Review Recommended
0	High	Inspection Required

3.5 Logistic Regression

Logistic regression is a statistical modeling technique used for predicting the probability of a categorical outcome based on one or more predictor variables. Unlike linear regression, which predicts continuous values, logistic regression is designed to handle binary or multinomial classification problems by modeling the log-odds of the outcome. In cases where the target variable has more than two unordered classes Low, Medium, High-risk levels, the logistic regression model is extended to multinomial logistic regression. It estimates a separate set of coefficients for each class except a reference class and uses the softmax function to compute class probabilities as shown in Equation 3. The class with the highest predicted probability is assigned as the output label. Multinomial logistic regression is particularly suitable when the outcome categories have no inherent order, as in the AML risk levels used in this study.

$$\boldsymbol{P(Y=k|X)} = \frac{e^{\beta_k^T X}}{\sum_{j=1}^{k} e^{\beta_k^T X}} \boldsymbol{for\ k=3} \tag{3}$$

Where :

Y = The categorical outcome (risk level)

X = The feature vector

β_k = The coefficient vector for class k

K = Total number of classes

4. Results and Discussion

In this section, the related results are shown and discussed. The final risk scores model for this study is shown in Equation (4).

$$\begin{aligned} Risk\ Score_{\text{High}} = & -0.24982078 + 0.082064x_1 - 0.258940x_2 - 0.416285x_3 + 0.049533x_4 \\ & + 0.082064x_5 + 0.494028x_6 + 0.313216x_7 + 0.494028x_8 + 0.494028x_9 \\ & + 0.142524x_{10} - 0.072155x_{11} - 0.329916x_{12} - 0.122193x_{13} + 0.154573x_{14} \\ & + 0.494028x_{15} + 0.512354x_{16} - 0.503811x_{17} + 0.114575x_{18} + 0.215567x_{19} \end{aligned}$$

Where:

x_1 = Bayaran_Umpukan_TS
x_2 = Dokumen_sokongan
x_3 = Dokumen_sokongan_legalperson
x_4 = Dokumen_sokongan_pemegang_luar
x_5 = Due_Payable_TS
x_6 = EDD_risiko_tinggi
x_7 = Hasil_pemprofilan_risiko_CRP
x_8 = Kelulusan_pengurusan_diperoleh_CRP
x_9 = Medium_CRP
x_{10} = Medium_maklumat_pengarah
x_{11} = Menyimpan_maklumat_legalperson
x_{12} = Pemegangsaham_tidak_berdaftar
x_{13} = Pemeriksaan_keatas_pelanggan
x_{14} = Pemeriksaan_pangkalan_data
x_{15} = Pemprofilan_risiko_CRP
x_{16} = Penilaian_RedFlag_STR_CRP
x_{17} = Rekod_KYC
x_{18} = Saham_dibayar
x_{19} = Simpan_dokumen_bayaran

The logistic regression model demonstrated satisfactory predictive capability. It achieved an overall classification accuracy of 100% as shown in Figure 2, indicating that most institutions were correctly classified. The precision and recall metrics were generally balanced across all three risk categories, with particularly strong precision observed for the high-risk class. The ROC AUC for high-risk detection was measured at 1.00 signifying a high level of discriminative power in identifying institutions that pose significant AML risks.

J. of Technology Management and Business Vol. 12 No. 2 (2025) p. 98-106

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         2
           1       1.00      1.00      1.00         3

    accuracy                           1.00         5
   macro avg       1.00      1.00      1.00         5
weighted avg       1.00      1.00      1.00         5

Confusion Matrix:
 [[2 0]
 [0 3]]
Accuracy Score: 1.0
ROC AUC Score: 1.0
```

Fig. 2 *Logistic regression classification report*

The confusion matrix as in Figure 2 confirms the perfect classification, with zero misclassifications observed. These results demonstrate the robustness of the model despite the relatively small test set. However, it is important to recognize the possibility of overfitting particularly given the dataset limited size and binary structure. Nevertheless, the model's precision and recall across both classes make it a reliable candidate for real-world AML compliance screening. The ROC curve Figure 3 illustrated a clear separation between high-risk and other categories, reinforcing the practical applicability of the model as a regulatory decision-support tool, which support the results in Figure 2.

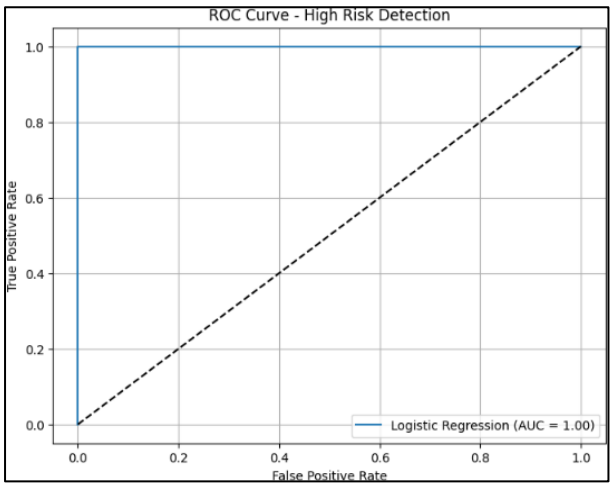

Fig. 3 *ROC curve*

The ROC curve showing a perfect AUC of 1.00 suggests that the logistic regression model achieved flawless classification between high-risk and low-risk entities. While this may appear ideal, such a result often indicates potential issues in the modeling process. One likely reason is overfitting, particularly the dataset is small, causing the model to memorize patterns that may not generalize to real-world data. Additionally, the features used may be too clean or perfectly separable, which is common in synthetic datasets lacking the noise and variability found in actual financial transactions. Another common cause is data leakage, where the model inadvertently gains access to information that directly or indirectly reveals the target label, such as using post-event variables or improperly engineered features. In some cases, biased or imbalanced sampling, where only extreme cases of compliance or non-compliance are included, can also lead to unrealistically high performance.

To improve the model's robustness and ensure its reliability, several steps should be taken in the future. Implementing cross-validation can help verify that the model's performance is consistent across different subsets of data. Using more realistic or noisy datasets, possibly incorporating real-world financial records, can make the model more adaptable to varied patterns. It is also crucial to audit the feature set for potential data leakage and to validate the model against a hold-out or external test set to assess generalization. Finally, relying solely on AUC is insufficient; other evaluation metrics such as precision, recall, the confusion matrix, and especially the false positive rate should be examined, as these are critical in the context of Anti- Money Laundering (AML) where false alerts can lead to significant operational costs.

The integration of a quantitative risk scoring framework with a multinomial logistic regression model provided a structured and interpretable approach to AML risk classification. The model successfully identified entities at varying levels of regulatory compliance and enabled risk-based prioritization in accordance with established supervisory practices. Despite the simplicity of the feature encoding and the relatively limited sample size, the model yielded reliable results and demonstrated scalability for broader applications.

The use of interpretable logistic regression algorithms is especially valuable in regulatory environments, where transparency and accountability are paramount. Moreover, the combination of data-driven scoring techniques with rule-based thresholds reflects a hybrid analytic framework that leverages both empirical evidence and expert judgment to inform risk assessments.

5. Conclusion

In conclusion, this paper effectively highlighted the early stage of Anti-Money laundering detection within the datasets by identifying key parameters of AMLA 2001, developing an Anti-Money Laundering algorithm and evaluating the efficiency of the model. Through this paper, it includes the identification of significant parameters outline in the Reporting obligation under AMLA 2001 to maximize the detection of Anti-Money Laundering action while minimizing false positives (FP). This model incorporating logistic regression model to detect money laundering risk for reporting institution using risk score model that has been created using the dataset to provide guidance. The outcomes include identification of significant AMLA parameters that guide institutions in prioritizing resources and focusing monitoring efforts on the most critical areas. This model is based on AMLA 2001 parameters, incorporating rule-based approaches to ensure interpretability and compliance with legal standards. This research has contributed to strengthening AML efforts within reporting institutions by providing a more effective and efficient means of detecting which companies should be inspected or not.

Acknowledgement

The authors would like to thank the Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing financial support and Suruhanjaya Syarikat Malaysia (SSM) for the business expert given in this study. The authors also would like to thank the reviewers for the valuable comments for improvements for this paper.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper. The authors have reviewed and approved the final version of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Muhammad, N., Mohamad, M. I. H., Abu Bakar, W. Z. W.;* ***data collection:*** *Muhammad, N., Nor Hairudin, N. A. S, Mohamad, M. I. H., Abu Bakar, W. Z. W;* ***analysis and interpretation of results****: Muhammad, N., Nor Hairudin, N. A. S;* ***draft manuscript preparation****: Muhammad, N., Nor Hairudin, N. A. S. All authors reviewed the results and approved the final version of the manuscript.*

References

Doppalapudi, P. K., Kumar, P., Murphy, A., Zhang, S., Rougeaux, C., & Stearns, R. (2022, October 7). The fight against money laundering: Machine learning is a game changer. McKinsey & Company

Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., & Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. Journal of Money Laundering Control, 23(1), 173-186.

Lim, J., & Muhammad, N. (2024). Developing a New Feature for Vulnerability Risk Scoring Model for Enhanced Cybersecurity. Journal of Statistical Modelling and Analytics, 6(2), 1–18. https://doi.org/10.22452/josma.vol6no2.5

Md Sum, R., & Abdul Khalik, Z. (2020). The Influence of Corporate Governance on Enterprise Risk Management Implementation: A Study on Non-Financial Public Listed Companies in Malaysia: ERM Implementation and Corporate Governance. Journal of Technology Management and Business, 7(1), 50-64.

Nasdaq Verafin. (2019). ML Higher Performance Analytics for Lower False Positives. Nasdaq. https://verafin.com/2019/08/machine-learning-higher-performance-analytics-for-lower-false-positives

Oztas, B., Cetinkaya, D., Adedoyin, F., Budka, M., Aksu, G., & Dogan, H. (2024). Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry. Future Generation Computer Systems, 159, 161–171. https://doi.org/10.1016/j.future.2024.05.027

Oztas, B., Cetinkaya, D., Adedoyin, F., Budka, M., Dogan, H., & Aksu, G. (2023). Enhancing Anti-Money Laundering: Development of a Synthetic Transaction Monitoring Dataset. Proceedings - 2023 IEEE International Conference on e-Business Engineering, ICEBE 2023, 47–54. https://doi.org/10.1109/ICEBE59045.2023.00028

Plaksiy, K., Nikiforov, A., & Miloslavskaya, N. (2018, August). Applying big data technologies to detect cases of money laundering and counter financing of terrorism. In 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW) (pp. 70-77). IEEE.

Tertychnyi, P., Slobozhan, I., Ollikainen, M., & Dumas, M. (2020, August). Scalable and imbalance-resistant machine learning models for anti-money laundering: A two-layered approach. In International Workshop on Enterprise Applications, Markets and Services in the Finance Industry (pp. 43-58). Cham: Springer International Publishing.

Tiwari, M., Gepp, A., & Kumar, K. (2020). A review of money laundering literature: the state of research in key areas. Pacific Accounting Review, 32(2), 271-303.

Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. arXiv:1908.02591.

Yadav, A. (2020, March 21). Steps to be followed in data science [Image]. Medium. https://medium.com/@arvindy/steps-to-be-followed-in-data-science-1a7866cf7463