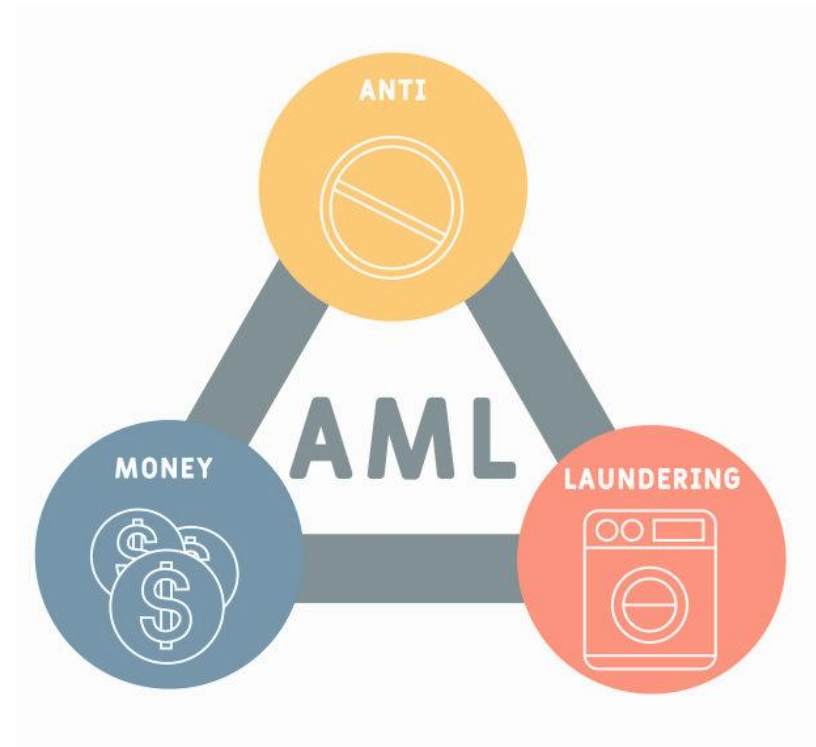


Explainable AI for Anti-money laundering in banking.



Student ID: 253001288

Student name: Aishah Abdulla Saeed Ali Aldhanhani

Contents

Abstract..... 3

Introduction..... 4

Background 5

 What is the Role of Explainable Artificial Intelligence (XAI) in Anti-Money Laundering (AML):..... 5

 Challenges and Limitations: 5

Problem Statement 6

Research Objectives 7

 Primary Goal: 7

 Sub-Objectives: 7

Research Scope 8

Literature Review: 9

Methodology..... 11

 1. Literature Review and Gap Analysis 11

 2. Data Collection and Preparation 11

 3. Model Design and Training 11

 4. Evaluation and Comparison 11

Expected Outcomes..... 12

Gantt Chart..... 13

References 14

Abstract

One of the challenges recently facing the banking sector is identifying whether transactions are suspicious or ordinary, and if they are suspicious, what is the reason that made the system select them as such. This is especially true when relying on non-traditional methods such as the Black Box. Despite the accuracy of these models, and although the answers are usually one hundred percent correct, they lack transparency, which leads to a deficit in trust between the customer and the banking sector. It also poses a risk to regulatory compliance.

The study that I have chosen proposes the development of an integrated model that detects suspicious transactions, but with an explanation based on Explainable Artificial Intelligence (XAI). Complete transparency and compliance with regulations and systems are the primary goal of this model. There are techniques considered to be the building blocks of this project, such as SHAP and LIME, to bridge the existing interpretability gap. The outcome of this study will be a transparent and explicit XAI system that provides results with reasons

Introduction

As I mentioned, the detection of suspicious transactions has become a profound problem in the banking sector. To move away from these methods, which some researchers have called "primitive methods," Artificial Intelligence (AI) has been placed at the forefront of the project as a fundamental and conditional solution to this problem. However, there are still some models in use that rely on the Black Box—meaning a result without a reason! which has begun to constitute a failure in the interpretive models

The essential reasons and drawbacks that financial analysts and auditors were facing in the banking sectors, or the financial sectors more broadly, were understanding the system's decision and the reasons that led the system to identify these transactions as suspicious, especially for transactions with very large monetary value or amounts.

Explainable Artificial Intelligence (XAI) offers the possibility of providing correct decisions with high accuracy, along with precise reasons and interpretations for these alerts on specific transactions

Background

Explainable Artificial Intelligence (XAI) is a fundamental axis for understanding the position of AI in risk management within the banking sector comprehensively, especially in the banking sector, and how its introduction can be a reason for limiting Anti-Money Laundering (AML) and whether it will affect credit reports.

The points where AI can be essential are:

- I. Fraud detection
- II. Anti-Money Laundering (AML)
- III. Risk management

What is the Role of Explainable Artificial Intelligence (XAI) in Anti-Money Laundering (AML):

In the key areas where AI focuses, especially Anti-Money Laundering:

- I. **Enhancing Transparency and Trust:** By providing the customer with the reason for classifying the transaction as suspicious.
- II. **Supporting Regulatory Compliance:** By keeping pace with consumer protection requirements.
- III. **Facilitating Investigation and Auditing:** By using techniques such as **SHAP and LIME** and **Decision Trees**, the logic followed in classifying transactions is provided.

Challenges and Limitations:

The implementation of Explainable Artificial Intelligence (XAI) has proven to be limited due to several factors, including:

- I. **Internal Bank Regulations.**
- II. **Potential Costs.**
- III. **Scarcity of Real-World Data** needed to train the system due to customer privacy concerns.

Problem Statement

These days in the banking and financial sector, many accounts are involved in money laundering, and identifying these suspicious transactions has become a major challenge, especially with traditional methods. The solution was to introduce Explainable Artificial Intelligence to become one of the foundations of banking transactions, making it easier to detect abnormal transactions that can be classified as money laundering. Although there are some methods used by certain banks that fall under the term “black box,” and the results are often satisfactory and accurate, they lack transparency and provide no reasons or explanations for the outcome (Kumar et al., 2023). The lack of transparency shakes the customer’s trust in the bank and creates fear when carrying out or initiating high-value transactions, fearing that they might be considered suspicious or related to money laundering.

The current models assist only in providing decisions, but they are not supported by explanations; in other words, they lack transparency. This represents a major weakness in the existing models, as the limited interpretability and the absence of any method to trace the reasons or outcomes produced by the system have become a significant challenge between customers and the banking sector. On the other hand, this also poses a risk to regulatory compliance and consumer protection.

In summary, the need to develop a model that provides decisions along with explanations has become a fundamental requirement in the banking sector. focusing on model interpretation and regulatory compliance. It emphasizes the importance of employing certain techniques (Zhou et al., 2024) Employing Explainable Artificial

Intelligence is now a necessity not only in name, but also as a means to enhance regulatory compliance and strengthen trust with customers. Implementing such a system will bridge a significant gap within the banking industry.

Research Objectives

Primary Goal:

To build a model based on Explainable Artificial Intelligence, enabling the detection of suspicious transactions with full transparency for customers.

Sub-Objectives:

- a. To analyze whether the currently used systems lack transparency, clarity, or the ability to provide explanations for the results they generate.
- b. To review whether modern approaches—such as Explainable Artificial Intelligence can address this problem, and whether identifying the strengths and weaknesses of the existing model will enable the new system to overcome these limitations.
- c. To design an ideal model that provides both results and their underlying reasons, using advanced AI tools, and to determine whether this will improve decisions related to detecting suspicious transactions or risks associated with credit reports.
- d. To evaluate whether the proposed advanced system (Explainable AI) produces accurate predictions, whether its explanations are clear, and whether those explanations align with and comply with regulatory requirements.
- e. To compare the results of the newly proposed system with those of the currently used systems, assessing the level of improvement and determining whether there is a significant difference in performance.

Research Scope

The research focused on the following:

- **Focus:** The primary focus and objective were in the area of Anti-Money Laundering (AML) using Explainable Artificial Intelligence (XAI).
- **Methodology:** Developing and building an interpretable machine learning model—for example, **Random Forest** or **XGBoost**—integrated with XAI interpretation tools such as **SHAP** and **LIME**.
- **Data:** The research relies on financial transactions for detecting and limiting money laundering operations by training the model using **synthetic data** or real but **limited datasets** due to privacy concerns.
- **Evaluation: Accuracy and prediction** are the goals of this model, as well as the **quality of the explanation** to ensure compliance with regulatory requirements.

Literature Review:

Several previous studies have addressed the use of artificial intelligence in the banking sector for fraud detection and risk management, as reviewed in the research paper by Kumar et al. (2023). The study emphasized the importance of interpretability in strengthening customer trust, ensuring regulatory compliance, and enhancing consumer protection. It also highlighted that most models currently used in the banking sector rely heavily on black-box algorithms. The difficulties and challenges faced were the inaccuracy and imbalance of information and the results of false positives. (Han et al., 2020) which creates significant difficulty in understanding the reasons behind classifying certain transactions as suspicious.

Many studies have previously been presented that addressed the use of artificial intelligence in the banking sector for fraud detection and risk management, and this was reviewed in the research paper by Kumar et al. (2023). It confirmed the importance of interpretability in enhancing customer trust, regulatory compliance, and consumer protection. It also confirmed that most of the models used in the banking sector rely entirely on black-box algorithms, which created difficulty and an obstacle in understanding the reasons behind classifying suspicious transactions. One of the challenges highlighted by the paper is the limited number of practical trials conducted and the complete reliance on synthetic data simulations. (Kumar et al., 2023)

In summary of what was mentioned in previous studies, integrating these technologies contributes to reducing bias and enhancing transparency, especially when the customers are distinguished and the bank does not want to lose this trust. Bracke et al. (2019) point out that in extremely complex Deep Neural Networks, achieving full transparency may slightly degrade the model's predictive power. Furthermore, the scarcity of real-world datasets remains a recurring limitation across almost all studies (e.g., Nguyen et al., 2025; Han et al., 2023), often due to strict privacy laws. These technologies will also contribute to strengthening the reasons and will support the alerts issued by the system in case suspicious transactions are detected.

Supporting this approach Han et al. (2023) demonstrated that combining Decision Trees with SHAP (SHapley Additive exPlanations) significantly increases the ease of result analysis. Nguyen et al. (2025) further proved through comparative analysis that algorithms like Random Forest. Belle and Papantonis (2021) provided the theoretical grounding for these tools, explaining how LIME and SHAP allow complex models to be understood locally and globally, even when handling massive banking datasets.

It was also mentioned in the context of previous studies that introducing these models after development that removes the obstacles and at the same time meets all regulatory requirements and enhancing trust in the banking sectors, not only between customers. There was a focus on future plans and directions to allow the XAI system time to expand across many expects. (Li et al., 2024)

The literature highlights that the benefits of XAI extend beyond technical accuracy to socio-technical factors like trust. Li et al. (2024) found that XAI directly encourages customer trust and institutional transparency. From a regulatory standpoint, Bracke et al. (2019).

Methodology

To achieve the project objectives, several methodologies could be employed; however, the following is the methodology I have adopted and will continue to implement throughout the project:

1. Literature Review and Gap Analysis

A systematic review of existing literature has been conducted to identify obstacles, constraints, and knowledge gaps. Regarding constraints, the focus is on the weakness of available databases, most of which consist of synthetic or "fake" data rather than real-world financial records.

2. Data Collection and Preparation

It is essential to acquire databases that are either real-world or high-fidelity synthetic data. A major focus will be placed on Data Preprocessing, specifically cleaning and balancing the datasets. This is crucial to reduce the rate of False Positives, a recurring issue identified in previous studies.

3. Model Design and Training

- **Algorithm Selection: Random Forest** and **XGBoost** have been selected as the primary algorithms due to their high performance in complex classification tasks.
- **XAI Integration:** These models will be integrated with interpretability tools (such as **SHAP** and **LIME**).

4. Evaluation and Comparison

- **Performance Evaluation:** The model will be assessed based on its **Accuracy**, **Precision**, **Recall**, and **F1-score** in detecting money laundering activities.
- **Interpretability Assessment:** The generated explanations will be evaluated for their clarity, usefulness, and alignment with **Regulatory Compliance** and consumer protection requirements.

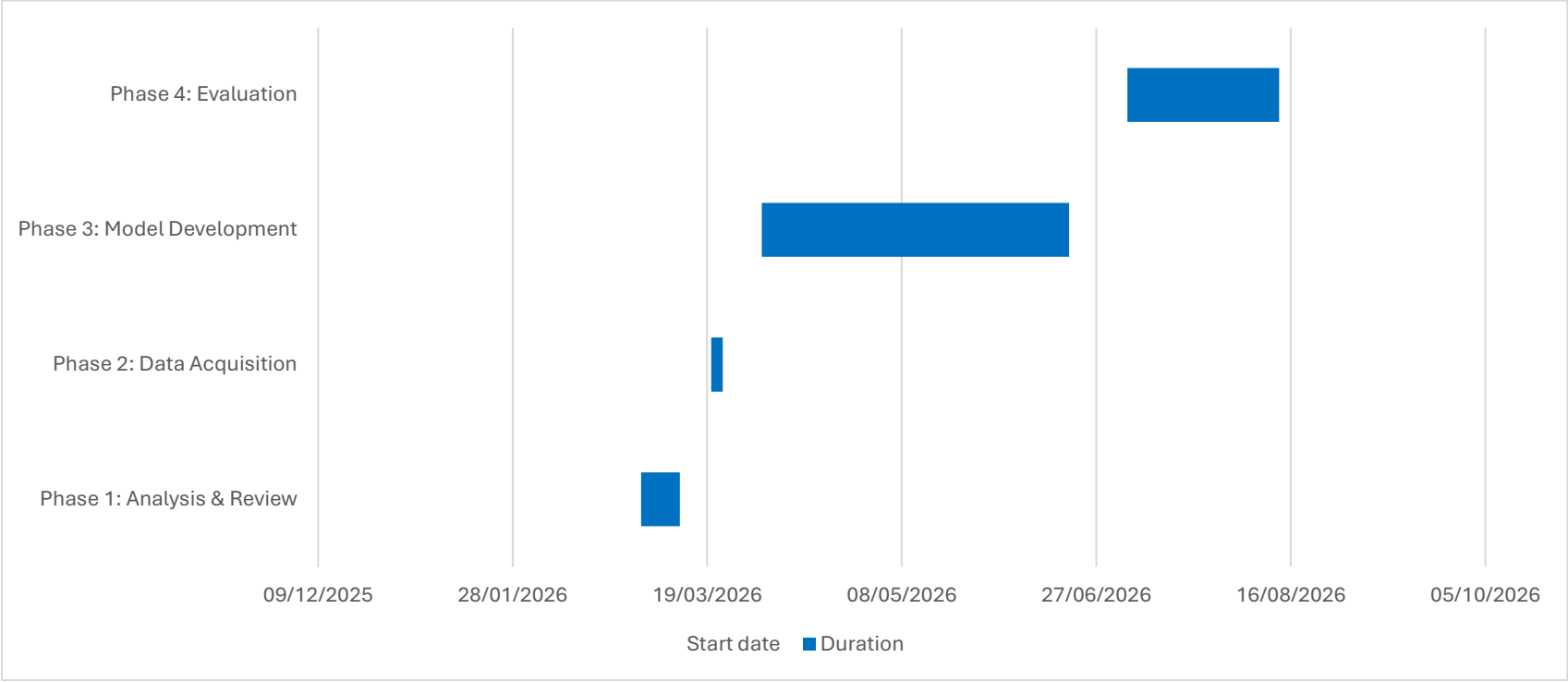
- **Comparison:** A comparative analysis will be conducted between the current "**Black Box**" systems (such as standard Deep Learning) and the proposed **XAI model** to quantify the resulting improvements in both performance and transparency.

Expected Outcomes

The expected outcomes of this research are as follows:

1. **An Explainable AI (XAI) Model:** Training an effective, interpretable system capable of detecting suspicious transactions while minimizing the rate of false positives.
2. **Enhancing Transparency and Trust:** To strengthen customer confidence, the system provides clear explanations for every decision it makes.
3. **A Framework for the Banking Sector:** Implementing Explainable AI within banking environments and seeking to overcome the constraints that hinder the real-world deployment of these models.

Gantt Chart



References

- Belle, V., & Papantonis, I. (2021).** Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data*.
- Bracke, P., Datta, A., Jung, C., & Sen, S. (2019).** Machine learning explained: Case studies in the financial sector. *Bank of England Working Paper*.
- Han, Y., Liu, Z., & Chen, X. (2020).** Artificial Intelligence for Anti-Money Laundering: A Review and Extension. *SSRN*.
- Han, Z., Wang, L., & Li, J. (2023).** Interpretable Machine Learning Models for Financial Fraud Detection. *AIP Publishing*.
- Kumar, R., Sharma, A., & Verma, P. (2023).** Applications of Explainable Artificial Intelligence in Finance - A Review. *Springer*.
- Li, M., Zhang, Q., & Huang, Y. (2024).** Explainable AI in Financial Services: Transparency and Trust in Risk Management. *Springer*.
- Nguyen, T., Tran, H., & Pham, K. (2025).** Comparative Analysis of Machine Learning Algorithms for Money Laundering Detection. *Springer*.
- Saraswat, M., Agarwal, S., & Tripathi, A. (2022).** Explainable AI in Anti-Money Laundering: An approach toward Transparent Financial Intelligence. *Journal of Banking and Financial Technology*.
- Sudjianto, A., Knauth, W., Singh, R., Yang, Z., & Zhang, A. (2021).** Explainable AI for Risk Management: Interpretable Machine Learning for Credit Risk and Anti-Money Laundering. *Journal of Financial Data Science*.
- Zhou, X., Chen, Y., & Li, D. (2024).** Explainable Artificial Intelligence (XAI) in Finance: A Systematic Literature Review. *Springer*.