COMP-1831 M01-2022-23

Technologies for Anti-Money Laundering and Financial Crime

Coursework for

MSc. Data Science

Sachin Suresh Dongare
001283296

School of computing and mathematics at

UNIVERSITY OF GREENWICH

Table of Contents

Table of Acronym’s

AML	Anti Money Laundering
AWS	Amazon web service
CA	Charted Accountant
FATF	Financial Action Task Force
FCA.	Financial Conduct Authority
GCP	Google cloud Platform
GDPR	General Data Protection Regulation
KYC	Know your customer
ML	Money Laundering
NCA	National Crime Agency
OCG	Organised Crime Groups
PML	Professional Money Laundering
PMLO	Professional Money Laundering organisation
SAR	Suspicious Activities Report
STR	Suspicious Transaction Report
TCSP	Trust and Company Service Provider
UK	United Kingdom
USD	United State Doller

Executive Summary

The AML as service would consist of three main components: a data collection component, a data analysis component, and a reporting component. The data collection component would be responsible for collecting data from a variety of sources, including financial institutions, law enforcement agencies, and government databases. The data analysis component would be responsible for analysing the collected data to identify suspicious activity. The reporting component would be responsible for generating reports on suspicious activity to regulatory agencies, financial institution and bank.

The platform would be designed to support the following FATF principles:

- Full client coverage: The platform would be designed to cover all customers of financial institutions, regardless of their size or risk profile.
- Accurate and complete data: The platform would be designed to collect accurate and complete data on all customers.
- Timely detection: The Platform would be designed to detect suspicious activity in a timely manner.
- Continuous monitoring: The Platform would be designed to continuously monitor customers for suspicious activity.

The Platform would also be designed to support the following regulatory requirements:

- Know your customer (KYC): The framework would be designed to collect and verify customer information, including name, address, date of birth, and source of funds.
- Customer due diligence (CDD): The framework would be designed to conduct ongoing risk assessments of customers and to take appropriate steps to mitigate risks.
- Suspicious activity reporting (SAR): The framework would be designed to generate SARs on suspicious activity.

The Platform would be designed to have the following architectural features:

- Scalability: The platform would be designed to be scalable to handle the large volume of data and transactions that are generated by financial institutions.
- Flexibility: The platform would be designed to be flexible enough to accommodate the different needs of different financial institutions like data regulation.
- Security: The platform would be designed to be secure to protect customer data and to prevent unauthorized access.
- Reliability: The framework would be designed to be reliable to ensure that it is available when needed.

1. Introduction

"*Money laundering is generally understood as the concealment of an illegitimate source of assets, providing an apparent legal origin*" (Chau,2020)

Money laundering is a global problem that undermines the integrity of financial systems and facilitates criminal activity, including drug trafficking, corruption, and terrorism. According to the United Nations Office on Drugs and Crime, it is estimated that between 2% and 5% of global GDP is laundered every year, which amounts to $800 billion to $2 trillion USD.

In the UK, money laundering is a significant problem that affects the financial industry, law enforcement, and society at large. The UK's National Crime Agency (NCA) estimates that hundreds of billions of pounds of dirty money is laundered through the UK each year, with a significant portion coming from overseas. The NCA also identifies various sectors that are particularly vulnerable to money laundering, including the property market, financial services, and professional services such as legal and accounting firms.

However, existing AML techniques have several drawbacks due to the ingenuity of Professional Money Launderers (PML) and Professional Money Laundering Organisations (PMLO). As They know how to outperform laws and regulations and find loopholes. As a result, reporting suspicious conduct to regulatory entities only on the basis of transaction analysis is difficult. but there is evidence to suggest that they are becoming increasingly ineffective at detecting and preventing money laundering. Criminals are using increasingly sophisticated techniques to launder money, and traditional AML systems may not be able to keep up with the evolving threat landscape. In addition, false positives generated by existing AML solutions can be time-consuming and costly to investigate, diverting resources away from more high-risk areas.

Therefore, there is a need for innovative, technology-driven solutions that can improve the effectiveness and efficiency of AML compliance efforts. An AML as a service platform that leverages advanced data analytics, machine learning, and artificial intelligence can help financial institutions detect and prevent money laundering more effectively and efficiently. This proposed solution will address the transaction analysis and validation using credit and asset management data. By automating routine tasks, such as transaction monitoring and risk assessment, and applying advanced analytics to detect suspicious patterns and behaviours, an AML as a service platform can help financial institutions reduce the risk of money laundering and comply with regulatory requirements.

2. Proposed Solution: AML as Service

Regulatory agencies, banks, and payment service providers can use AML as a service as part of a holistic solution to help them identify and detect money laundering activities. The service gathers, processes, and analyses vast volumes of financial data from many sources using cutting-edge technologies like data pipeline and storing data into data lake, including machine learning approach for transaction analysis, graph databases, visual analytics, and Pseudonymization privacy techniques.

The platform's main purpose is to spot trends and abnormalities that can point to instances of money laundering. And after spotting suspicious transaction to validate the role of user we are implementing holistic analysis based on established criteria like monthly transactions ,annual turnover, income declaration, credit score and historical transaction data. To avoid AML detection, money launderers frequently employ a variety of strategies and split big sums into smaller transactions. To contribute back to the originating source, these little transactions are finally collected at the opposite end. Aggregating transactions and doing network analyses can assist discover when minor transactions are gathered over time and probable links between senders and recipients can be made in order to detect this "divide and conquer" phenomena.

Our system accomplishes the necessary data prevention and protection procedures by applying necessary technology at data pre-processing section. The system gathers source data and uses KYC data to check user authenticity. Further data is sent to the data lake where machine learning techniques are used to detect suspicious transactions. If any transactions are found, a risk assessment is initiated for those transactions. Based on suspicious behaviour and the graph database, various networks, accounts, and organisations are examined to identify detailed patterns and link the accounts engaged in transactions. The same report is then submitted for validation in order to be confirmed, and by combining historical credit score data and transaction information, we can label the case and provide a final report.

Regulatory agencies, banks, and payment service providers who are in charge of adhering to AML requirements are the target audience for the solution. The platform may be used to spot high-risk transactions, look into transactions that have been identified, and validate capability of user to perform large transaction based on credit history and provide reports and dashboards that offer insights into the trends and patterns of money laundering operations.

3. Data access needs: GDPR

An AML platform as a service can make use of a variety of data sources, including but not limited to:

Customer Data: Customer data consists of details pertaining to a customer's identity, such as name, birthdate, address, and any official identification papers, such as a passport or driver's licence.

Transaction Data: Information concerning the transaction, such as the total amount moved, the sources and destinations of the funds, the nature of the transaction, and the people involved, are included in the term "transaction data."

Market Data: Market Data: This contains data on the market and overall state of the economy, such as stock prices, currency exchange rates, and commodities prices.

KYC: In order to confirm clients' identities and determine their risk of money laundering, two essential procedures are used: KYC (Know Your Customer) and CDD (Customer Due Diligence). Whereas CDD requires the evaluation of customer risk, KYC entails the gathering of consumer data. Data including the customer's name, address, date of birth, and identity papers are collected as part of KYC. The customer's identity is then confirmed using the provided information. By ensuring that financial institutions have correct client information and are able to spot questionable activity, KYC aids in the prevention of money laundering.

Financial Data: In order to confirm and validate the role of user in money laundering we are requesting Credit score data, Income declaration and National Insurance details.

All things considered, the information used in an AML platform as a service would depend on the precise specifications of the regulatory body and the participating financial institutions. For the platform to be effective in identifying instances of money laundering, it would need to be built to gather, process, and analyse data from several sources.

We are requesting regulator to let us use the above mentioned data, as its required and critical aspect in terms of validation of role in money laundering

4. Approach to data privacy

The General Data Protection Regulation (GDPR), an EU legislation that establishes guidelines for the processing and protection of personal data, would need to be followed by AML as a service for technologically identifying anti-money laundering transactions.

The solution will make sure that any personal data handled complies with GDPR guidelines in order to comply with GDPR. In order to do this, we employ the pseudonymization approach to hide the critical attribute like name, address, NI number. While the actual identity is being processed, validated, and presented to regulatory agencies.

Pseudonymization: This approach entails changing identifying information with a pseudonym or code for each user and allocating a special code that may be used to connect various pieces of data without disclosing the identity.

Additionally, since our system does not collect data directly from users, if a false positive result scenario arises in which it is determined that the incorrect user has been identified or evaluated, the platform will have a direct approach to delete the necessary data based on instructions received from banking institutions or regulatory bodies.

In order to adhere to the necessary technological and organisational safeguards in place to protect personal data against unauthorised or illegal processing as well as against accidental loss, destruction, or damage, We are also using data masking and pseudonymization. Moreover, it should have a data protection impact assessment (DPIA) in place to evaluate the risks related to the processing of personal data and to determine the most effective countermeasures.

We are storing this pseudonymous external data in our data lake for transaction analysis while masking other personal data, and the length of time for external data storage depends on the payment service provider. Along with platform's scalability. However, our system is looking for an annual approach to understand user behaviour because PML and PMLO use money mules who don't have criminal records and who are unable to carry out large transactions, so for analysis purposes we will store annual transaction data.

5. Architecture plan

5.1. Flow Chart

A high level flow chart for an Anti-Money Laundering (AML) system provides a visual representation of the various components and their interactions. It outlines how data is collected, processed, analysed, and reported to detect and prevent money laundering activities. The architecture diagram also helps to identify potential bottlenecks, dependencies, and risks that may impact the overall performance and effectiveness of the AML system. It provides a high-level overview of the system, including its infrastructure, technologies, and data sources

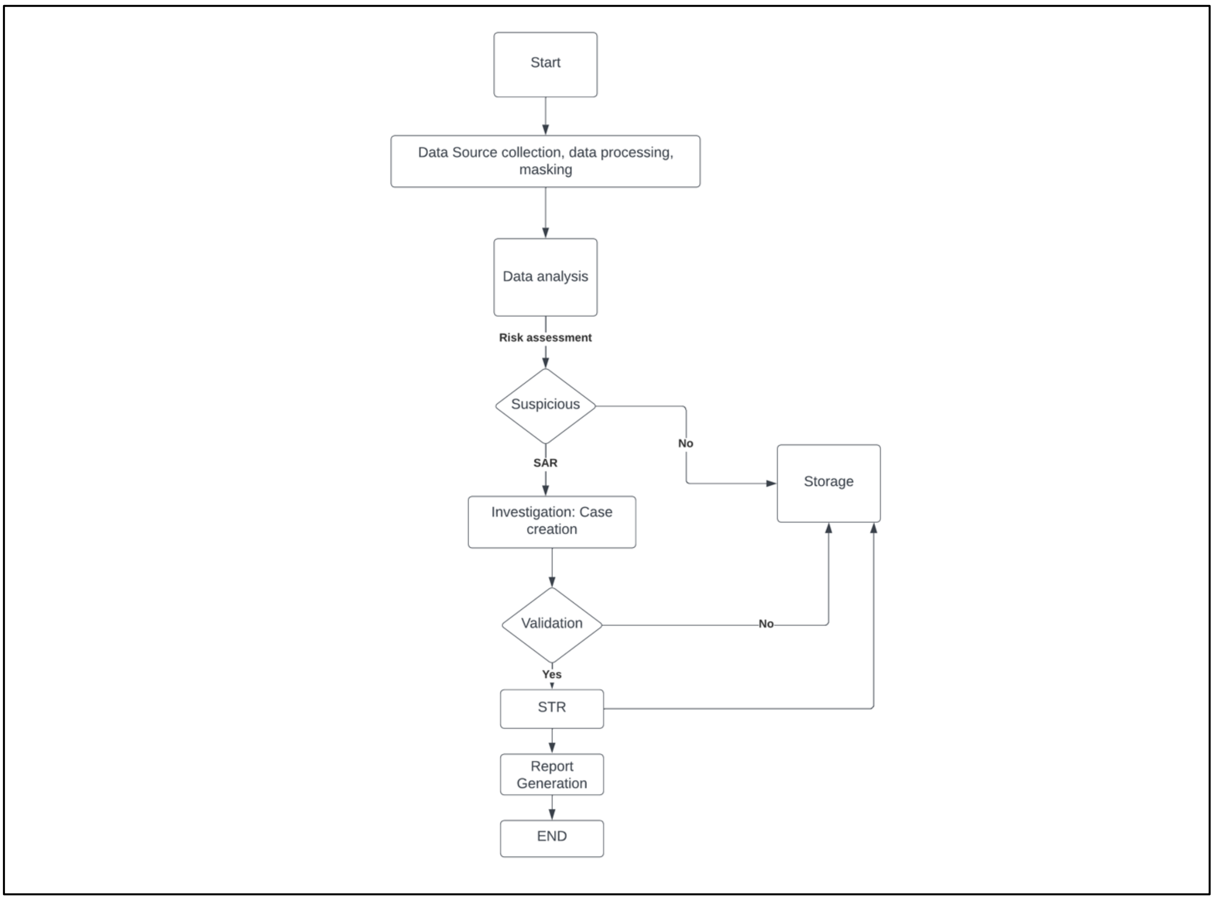

Figure 1 : AML Flow chart

Here's a high-level process flow diagram component for the proposed AML as a service solution with explained block.

Data source Collection: The platform collects transaction data from multiple sources, such as banks, financial institutions, payment processors, and other sources of financial data. The data is collected in real-time and stored in a secure data lake.

Data Pre-processing: The collected data is pre-processed using various masking techniques to ensure data privacy and confidentiality while maintaining the integrity of the data. The pre-processed data is then cleaned, normalized and transformed for analysis.

Data Analysis: The pre-processed data is analysed using machine learning models and rule-based algorithms to identify patterns and anomalies that may indicate money laundering activities. The analysis is performed using graph databases to enable network analysis and visualization of complex relationships and transactions.

Risk Assessment: Each transaction is assigned a risk score based on a risk-based approach to determine the likelihood of money laundering activities. The risk assessment takes into account various factors, such as the type of transaction, the location of the transaction, the parties involved, and the amount of money involved.

Suspicious Activity Detection: Transactions that are flagged as high-risk are further investigated to determine whether they are suspicious. This is done using a combination of graph analysis, machine learning models, and rule-based algorithms.

Investigation Management: When a transaction is flagged as suspicious, an investigation is initiated by assigning the case to an investigator or analyst. The investigation is managed using a workflow-based system that allows the investigator to review the transaction data, gather additional information, and communicate with other stakeholders involved in the investigation.

Validation: In addition to the above steps, the solution includes validation of the persons involved in the transaction by verifying their credit score data or asset valuation data.

Reporting: The platform generates reports and dashboards that provide regulatory bodies with insights into the trends and patterns of money laundering activities across multiple financial institutions and industries. The reports also include information on the effectiveness of the AML measures implemented by financial institutions and the level of compliance with regulatory requirements.

5.2. Data Architecture diagram

AML as service solution is based on a cloud-based architecture. This architecture allows us to scale our solution to meet the needs of our customers. It also allows us to provide our solution to businesses of all sizes.

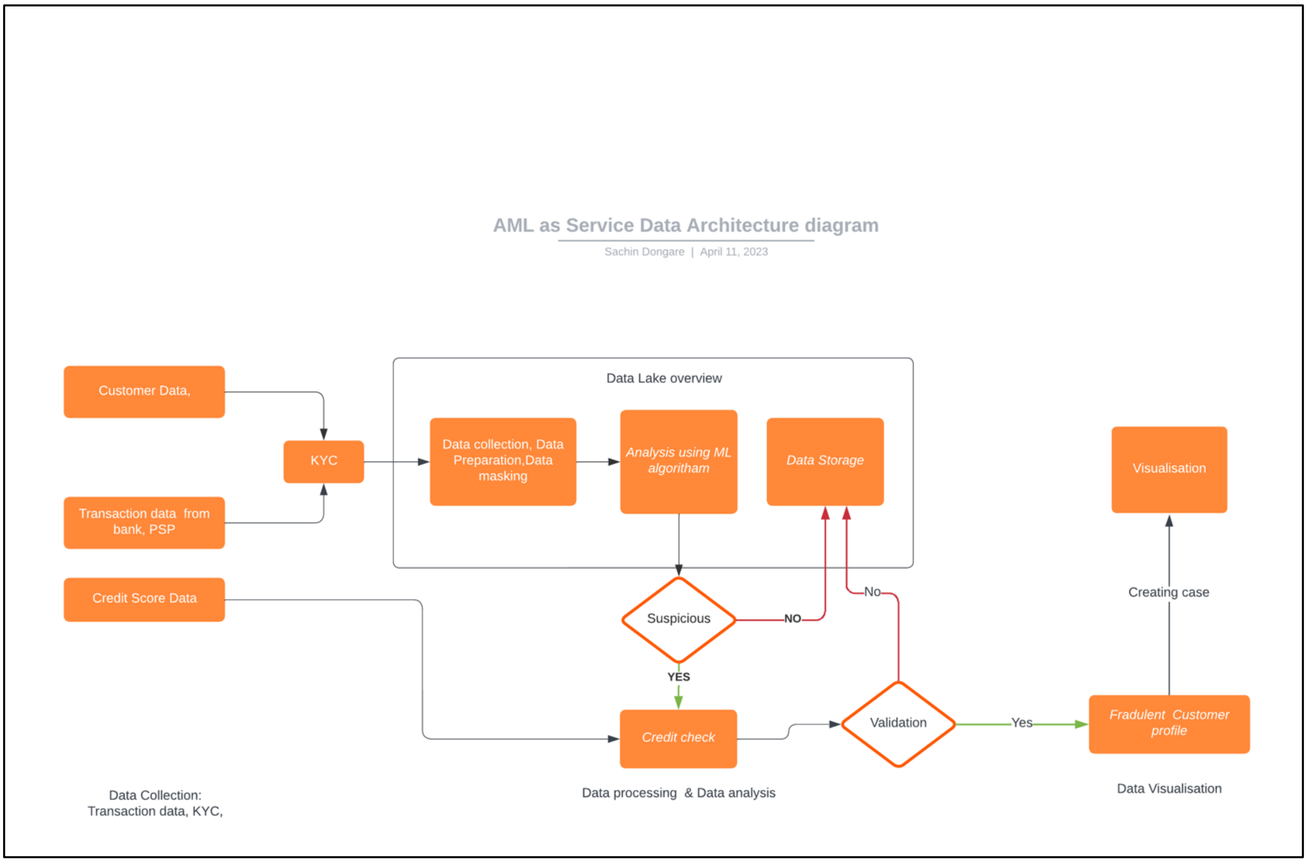
AML as Service Data Architecture diagram
Sachin Dongare | April 11, 2023
Customer Data,
Transaction data from bank, PSP
Credit Score Data
KYC
Data Lake overview
Data collection, Data Preparation,Data masking
Analysis using ML algoritham
Data Storage
Suspicious
NO
No
YES
Credit check
Validation
Yes
Fradulent Customer profile
Creating case
Visualisation
Data Collection: Transaction data, KYC,
Data processing & Data analysis
Data Visualisation

Figure 2 AML as service Data Architecture diagram

Here's a Data architectural diagram component for the proposed AML as a service solution with explained block.

Data collection: The platform needs customer data, transaction data from multiple sources, such as banks, financial institutions, payment processors, and other sources of financial data like credit score. The data is collected in real-time and stored in a secure data lake.

Data Processing: Collected data is pre-processed ,cleaned for analysis.

Data analysis: The pre-processed data is analysed using machine learning models and rule-based algorithms to identify patterns and anomalies that may indicate money laundering activities. The analysis is performed using graph databases to enable network analysis and visualization of complex relationships and transactions.

Data visualisation: Based on risk and investigation activities report is some transaction is found to be suspicious then required report is get visualised and presented to authorities.

6. Methodologies and tools

We are using a variety of tools and methodologies to provide our solution. These tools and methodologies include:

- Cloud service provider
- Data lake
- Data science:
- Risk-based approach

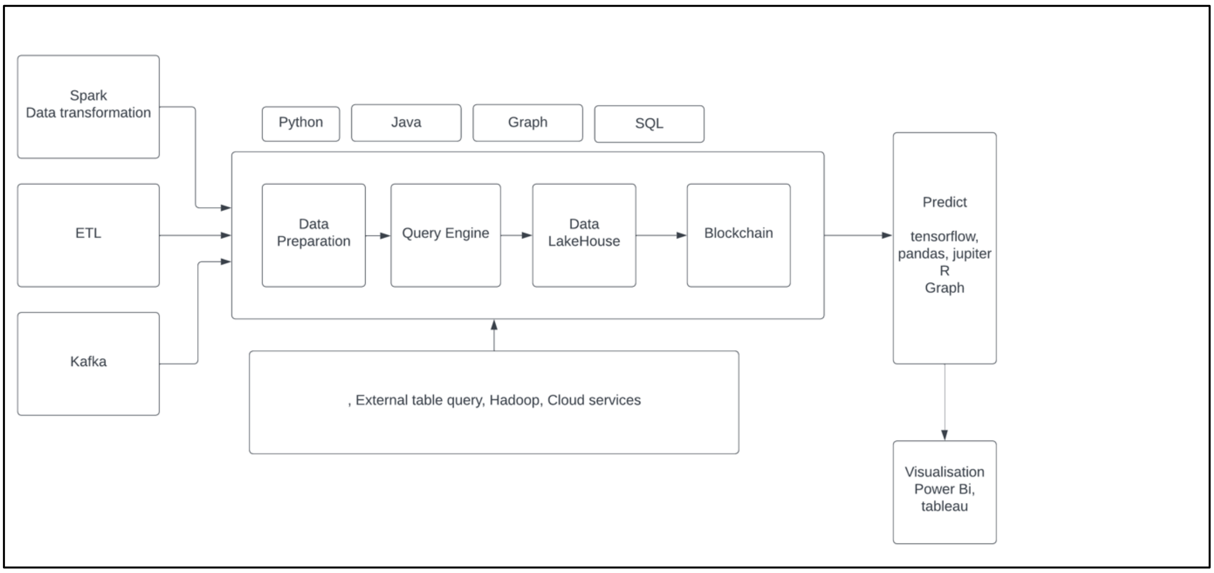

Figure 3: Tools used in AML service

6.1. Cloud Service provider.

Existing AML services are structured and SQL based to performing data extraction and loading is time consuming activity so to overcome that drawback we are using cloud based approach where we can utilise Big data technique to handle with big data using Java based approach or python based. But we mainly focusing on Python based approach for using complex analysis. In market we can find many cloud service provider offering that offer services for AML using Spark, Hadoop, and data lake technologies. Like. Microsoft azure, AWS, GCP. So we are Going with GCP as it pay as you go model and offers reliability. (Fig3)

6.2. Data Lake

Data lake services can be used to store and analyze large amounts of transaction data. This data can be used to train machine learning models and to identify suspicious activity. For example, a data lake could be used to store transaction data from multiple financial institutions. This data could then be used to train a machine learning model to identify patterns that are associated with money laundering.

Data Preparation: We are performing data extraction from financial institute and making it adhere to use in our solution as per our requirement and to achieve this we are using ETL tool i.e. Extract transfer and load mechanism to move data from one source to another. We will be applying additional data masking and Pseudonymization in data pre-processing stage only.

Query engine: We are using query engine that allows platform to interact with and retrieve data from the data lake. It allows platform to run queries on the data using SQL or other query languages, without having to specify the location or structure of the data. There are several query engines that will be used for AML in a data lake, such as Apache Hive, Presto. These engines enable our platform to perform advanced analytics, machine learning, and other data processing tasks on large volumes of data in the data lake. They also provide security and governance features, such as access control, encryption, and auditing, to ensure that the data is protected and compliant with regulatory requirements.

6.3. Data Science

Machine learning will be used to identify patterns in transaction data that are associated with money laundering, such as large, unusual transactions or transactions that involve multiple accounts. For example, a machine learning algorithm could be trained to identify transactions that are larger than a certain threshold or that involve multiple accounts that were opened in a short period of time. To improve accuracy and do complex analysis we are using neural network on transaction data to find complex pattern.

Graph methods will be used to analyze transaction data to identify relationships between different entities, such as customers, accounts, and transactions. These relationships can be used to identify suspicious activity, such as transactions that involve shell companies or smurfing. For example, a graph method could be used to identify transactions that involve customers who have opened multiple accounts or that involve accounts that have been used to transfer money to shell companies.

Data visualisation: tool is used for creating dashboard to show analytics and generate report to present to regulatory bodies.

6.4. Risk-based Approach

By defining the required set of rules and using the historical data of transaction along with credit score and income declaration we are going on next step to validate the identified risk involving transactions. To achieve this we are mapping the new purchase of assets information with the aggregated amount we found through analysis. Credit score and asset net worth data can be used to validate the same transaction user. For validating transactions using credit score data, algorithms like Linear Regression, Decision Trees, and Random Forest can be used to predict credit scores based on input features such as income, assets, and debt-to-income ratio. These algorithms can then be used in combination with AML algorithms to identify transactions that do not match the predicted credit score, which may indicate suspicious activity.

7. Limitations and future scoop

There are several limitations to consider for an AML as a service statup, including:

1. Data quality: The efficacy of AML detection can be significantly impacted by the accuracy and completeness of data gathered from diverse sources. False positives and false negatives can result from poor data quality, which can be expensive for financial firms and regulatory agencies.
2. Data security and privacy issues are raised by the usage of private and sensitive financial information for AML purposes. Making sure that data collection, storage, and processing comply with data protection laws and regulations is crucial.
3. Compliance with legal standards: Legal rules and regulations for anti-money laundering (AML) are continuously changing, so it can be difficult to stay on top of everything. Regulator rules that are not followed may result in penalties, legal action, and reputational harm.
4. Cost: For companies with limited funding, creating and maintaining an AML as a service platform can be expensive. The cost of data processing, analysis, and storage, as well as the cost of acquiring qualified personnel, may be high.
5. Low uptake: Financial institutions could be reluctant to use new AML technology, especially if they have already devoted money to using AML solutions that are already on the market. For a new AML as a service firm, trust-building and establishing a customer base might take some time.
6. False positives: AML detection systems have a tendency to produce many false positives, which can make further investigation time-consuming and interrupt important lawful transactions.

Scoop:

1. To use geographical data in collaboration with mules, suspicious account identified
2. To make use of web data from suspected user to take legal action.
3. Using call data log and mapping along with them transaction network on another side to draw connection between sender and receiver.
4. To collaborate with E-commerce platform for monitoring watchlist and spending pattern.

8. Specific actionable conclusion

To adhere with the main goal to stop and prevent money laundering activities we have touched some area as following:

1. Concentrate on creating and implementing cutting-edge machine learning algorithms to efficiently and accurately detect money laundering activities.
2. Set up a reliable system for data gathering and processing that can cope with enormous amounts of transaction data coming from many sources in real-time.
3. To safeguard sensitive financial information and stop data breaches, implement cutting-edge data privacy and security solutions like differential privacy and blockchain.
4. Provide a user-friendly reporting system and dashboard that offers regulatory agencies, financial institutions, and other stakeholders actionable information and analytics on money laundering compliance.
5. Use Hadoop and Spark, as well as cloud computing and data lake services, to effectively handle and analyse huge amounts of transaction data.
6. Work together with regulatory organisations, financial institutions, and other parties to make sure the AML as a service platform complies with applicable laws and requirements.
7. Maintain a constant eye on and upgrade the AML as a service platform to remain abreast of new developments in AML trends and technology.
8. To expand the capabilities and offers of the AML as a service platform, develop strategic alliances and collaborations with other businesses and start-ups in the AML and financial technology sectors.

Reference:

European Parliament. (2019). "Study on the Evaluation of the EU Framework for the Prevention of Money Laundering and Terrorist Financing." Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634441/IPOL_STU(2019)634441_EN.pdf

FATF. (2018). Money laundering / terrorist financing risks and vulnerabilities associated with gold. Financial Action Task Force. Retrieved from https://www.fatf-gafi.org/media/fatf/documents/reports/Risks-and-Vulnerabilities-Gold.pdf

Financial Action Task Force. (2020). "Money Laundering and Terrorist Financing Awareness Handbook for Tax Examiners and Tax Auditors." Retrieved from https://www.fatf-gafi.org/media/fatf/documents/publications/methodsandtrends-tax.pdf

Kshetri, N., & Voas, J. (2018). Assessing the role of Big Data in tackling financial crime and compliance management. Journal of Organizational Computing and Electronic Commerce, 28(3), 226-244. doi: 10.1080/10919392.2018.1478584

National Crime Agency. (2021). "National Strategic Assessment of Serious and Organised Crime 2021." Retrieved from https://www.nationalcrimeagency.gov.uk/who-we-are/publications/506-national-strategic-assessment-of-serious-and-organised-crime-2021/file

United Nations Office on Drugs and Crime. (2020). "Global Study on Smuggling of Migrants." Retrieved from https://www.unodc.org/unodc/en/data-and-analysis/glosom.html