Technologies for Anti-Money Laundering and Financial Crime

Author: Neharika Joshi

Acknowledgement

To begin with, I would like to express my deepest appreciation to those who provided us with the opportunity to attend this module.

Many thanks go to the course leader, Dr. George Samakovitis who has invested his full effort in guiding the students in achieving the goal. Also, many thanks go to the course tutor Mr. Antreas Pogiatzis.

Lastly, I would like to thank my beloved husband, Mr. Deep Khadka and my fellow classmates for providing me with insightful remarks and improving my work for this module.

Table of Contents

Table of Figures

1. Introduction to Anti-Money Laundering

The policies that enforce tracking every transaction and monitoring if there is any unusual activity which is against the law is known as Anti-money Laundering. It aims to prevents illicit resources to enter the financial system which are fueled by criminals and its main goal is to track where the funds exactly came from. Money launderers often gather their funds from activities like smuggling, drugs trafficking, insider trading, embezzling money, scams through technology and many more.

Lately, money laundering is emerging as a huge financial problem hence, it is essential to train the employees of a financial institution on how to deal with and locate money laundering on their surveillance. If any suspicious behavior is detected then, they are legally required to report the activity to proper authority. Nowadays, machines and software are now being used to detect possible illegal scenarios that people are unable to detect.

2. The Ecosystem

This report is about how the Anti-money laundering service will be conducted and is targeted more towards banks in association with Payment Service Providers and ID verification providers and the center of it is the service for Anti-Money Laundering. The main goal here is to detect any fraudulent transactions happening throughout whether the account that was made a long time ago had sudden transactions for high amounts or alarming remittance from high-risk countries. Currently, European Commission recognized 16 different countries as high-risk: Afghanistan, Bosnia and Herzegovina, Guyana, Iraq, Lao PDR, Syria, Uganda, Vanuatu, Yemen, Ethiopia, Sri Lanka, Trinidad and Tobago, Tunisia, Pakistan, Iran, and Democratic People's Republic of Korea (The Association of Accounting Technicians, 2019).

3. Background

To manage the complex hurdles in like Anti-Money Laundering, it is critical to learn Artificial Intelligence, particularly Natural Language Processing and Machine Learning. NLP is already gained a lot of attention towards the levels of AML that has already being employed at various levels of AML regulations. Furthermore, Sentiment analysis is a popular method which has been used successfully in a variety of fields such as analyzing various factors and individual consumers. The ability to collect data from real time and apply sentiment analysis is a remarkable achievement. Investigators can choose high-quality data based on definite, reasonable, and rational emotional expressions and views (Thi, et al., 2020).

Also, customers can be segmented effectively and also find anomalies using advanced analytics tools especially unsupervised machine learning algorithms which will be derived from real-time financial behaviors. The model can generally view the customers for understanding the fundamental behavioral patterns of customers in an organization. In addition, the model can categorize the customers by integrating high-risk features, only using cash or frequent international transfers.

3.1 Current Scenario for UK

In context on UK, there are a lot of criminals trying to launder money through innocent peoples. According to the report (UK Finance, 2021), the Covid-2019 pandemic was a huge safe haven for criminals to commit frauds and launder their money. Furthermore, about £1.6 billion was halted by the UK financial industry due to unauthorized fraud attempts in 2020. The number of scams crated to rob people of their money increased during the pandemic through phishing texts and emails, tax evasion emails, e-commerce or auction scams and many more. As a result, UK Finance is urging the government to include fraud and financial crimes in the upcoming Online Security Bill. Basically, huge technology companies should address the flaws in their platforms and make it even more efficient to save their customers to be exploited by the criminals (UK Finance, 2021).

The UK exited the European Union on 31st January 2020 and ended the transition process on December 31st,2020. The both parties agreed on the Trade and Cooperation Agreement (TCA) on December 24, 2020 governing the upcoming ties between them. In case of the third-country parties which means the countries residing outside the EU before Brexit but now, even EU nationals are considered third-country entities. So, moving forward even EU has the third-country criteria when it comes to business deals (The Law Society, 2021). As a member of EU, UK applied the first five AML directives but right before the transition period ended, EU created sixth AML directive which took effect all over EU. Although, UK government decided not to follow the sixth directive and decided to stick with the most recent. UK believes that they are already compliant with the Directive's initiatives, and already goes much further in regards to any violations and verdicts set out as per the Directive (Glaser & Roberts, 2021).

3.2 Existing Anti-Money Laundering Systems

In the current scenario, AML systems are simple black box system that either identifies a certain transaction as fishy or not fishy by setting some standards. Furthermore, there are a huge number of false positives being detected by the system and the average amount of false positives created is between 95% and 99% (Singla, 2020). This is not the correct process to handle fraudulent behaviors without giving the proper reason of doing so. A proper interaction with the risk analysts is a crucial factor and they should be held accountable for the detecting any suspicious activity in a business. Existing systems provide a large number of inconsistent results, causing banks to waste billions of pounds (£) on nothing. For detecting Money Laundering, financial instructions generally establish caps per transaction and looks for certain patterns which violates the set boundaries. To exemplify, after the investigation of the 9/11 attacks in United States of America, it was discovered than the terrorists transferred small amounts of money every now and then to be undetected by the systems which lead to such heinous crime (Gao, et al., 2006). In today's context, many security organizations believe Machine learning is the supreme answer for anti-money laundering to reduce growing pressure on detecting fraudulent behaviors.

3.3 KYC and Transaction Monitoring

Firstly, for properly identifying a new customer the banks should implement the procedures for KYC. Know You Customer (KYC) is a procedure that helps institutions on identifying who the person before taking them on board. For analyzing the customer's identity, the bankers can use the system to do a credit score, check the search engine history and customer's presence on social media. Particularly, Twitter and LinkedIn which provides a background of what the person is interested in and how they think also, what they do for living and where they come from. Also, the ID verification service providers can come into play for making sure that the identity provided by the customer is all correct. It would also be a plus if the documents provided by the customer can be cross-checked by using some powerful image recognition algorithms to check if there is a possibility of it being fake or not. Machine learning could be very useful for determining whether a customer's identity matches with an entry from risk database. Fortunately, utilizing Natural Language Processing (NLP) methods, cross-checking texts with various data sources is rather simple.

4. Life Cycle

There is total 6 phase while designing an AML service project

I. **Understanding the AML Scenario:**
The initial phase of this cycle is to study the current problems that is arising due to Money Laundering. Furthermore, the techniques that is currently in use to tackle Money Laundering is essential. Some popular methods are Blockchain Technology, Machine Learning, Graph Data Analysis and many more.

II. **Data Exploration:**
This phase begins with collecting the data and getting acquainted with the data. Furthermore, it is essential to determine what kind of data is necessary, identifying the data quality issues and gaining insights to the data. Also, looking for intriguing subsets in order to produce assumptions about unlabeled data. KYC process is also done in this process.

III. **Data Preparation:**
All efforts required to generate the final data set to be used in the modelling phase are concentrated in the data preparation phase. The work required to generate the dataset which will be used for training the model is concentrated in this stage. Generally, data preparation tasks are completed multiple times like selecting, cleaning, creating, merging, and modifying data for modelling purposes are among the responsibilities.

IV. **Train the Model:**

There are numerous algorithms to pick from, so experimenting with different methods might be a useful way to begin training the model. Hyper-parameter adjustment may also help the model perform better. Certain approaches have certain data format criteria, which leads to frequent returns to prior steps in search of a solution to improve the model.

V. **Evaluate the Model:**

To ensure that the strategic aim is reached, it is critical to assess and thoroughly analyze the procedures taken to create the final model before final deployment. Furthermore, it is critical to determine whether there are any other factors that have not yet been addressed. In order for the final model to be deployed, it is also necessary to evaluate the model using appropriate metrics and determine whether the outputs are efficient.

VI. **Deployment of Model:**

For the final phase, the developed models must be introduced and implemented within the company and ultimately shared for consumers to use in a usable format.

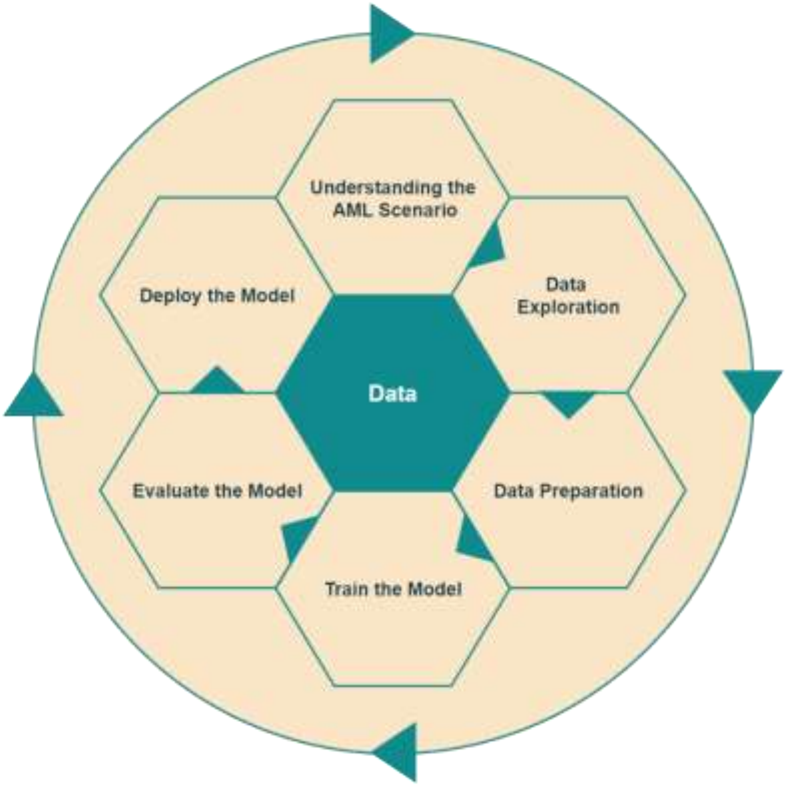

Figure 1: Lifecycle of Designing AML Service

5. Data Perspective

“More the data, better the results” is one of the rules in Machine Learning which means that it is necessary for the model to congest data as possible with diversity as well. It is essential to feed huge amount of data to the system as it will elevate the probability of locating more patterns. Hence, the system will be more capable for understanding the pattern and relationship of the data itself. To exemplify, if a system is provided with existing transactional data with variety of customer patterns and habits then it is likely for it to produce more efficient result. It is always ideal to provide data from diverse resources. Furthermore, every customer should have a different profile since, the characteristics of a single transaction may imply a high risk for one consumer while being a normal transaction for another. As a result, relying solely on transactional data to make decisions is insufficient also, the ML-based model is self-sufficient to decide whether the data given is valuable or not. Data quality should also be taken into account as, there is a popular saying “garbage in, garbage out” which means that if inadequate data is provided, the results will also be inadequate. Hence, data cleaning is crucial.

Any information that a money laundering officer examines in order to complete a task or investigate a questionable case. Transactional data, user profile data, and behavioral data are all vital for the system. Numerous banks create massive amounts of transaction data every day through their banking systems. This kind of data is produced in a large amount and generated really fast containing all types of data. Most banks are attempting to keep up with the velocity at which data is generated and analyzing it in order to gain important insights or possibly using analytical methods to identify financial crimes. Even though banks have massive amount of data being stored, most of it goes unused and unmined in many common cases. Furthermore, the data can mostly be unstructured which doesn’t have designed a prior defined schema. Other publicly available data sources like IP addresses, social network, company links, news articles etc. should be explored and integrated into the system. Hence, automating the collection of both internal and outer sources can aid to improving the efficiency of the suspicious activity alerts. In addition, most of the datasets produced contains 98% of genuine transactions whereas only 2% are suspected as fraudulent behavior thus, resampling of the data is necessary or else the model can be inaccurate while detecting money laundering.

6. Data Privacy

The anti-money laundering (AML) checks mandated by Fifth Money Laundering Directive(5MLD) would inevitably necessitate the acquisition of much more personal data from consumers and employees than previously. Documents issued by authorized parties, such as passports or citizenship, are necessary to establish proof of identity, and this data is crucial for executing AML processes, particularly for ID Verification institutes. However, if this information enters into the wrong hands, it raises the risk of identity theft. Also, information about an individual's net worth should be protected since if it is made public, the person may face threats or even financial loss; therefore, personal data security precautions must be performed. The AML Directive and recommendations from the Financial Action Task Force (FATF) urge that such checks be conducted in a risk-sensitive manner (Kenna, 2021). The people whose data is being used are now coming forward to filing legal compensatory damages against the organizations that have been exploiting their private data. It is rather simple to win the lawsuit, as they don't need to show that they have gone through a financial loss but a "sense of powerlessness" over the data being used in enough.

Sensitive data will be stored in a private cloud that is extremely secure. Customer data and financial transaction records will be utilized to train the machine learning model and to improve its performance by feeding it new data on a regular basis. The data will be retained in semi-structured format (CSV, JSON) files for at least 5 years. In addition, the data will be appropriately encrypted with a key for each entity, making it more difficult to decrypt even if it falls into the wrong hands. Furthermore, the data subjects will be given adequate transparency information about what is being done with their data, such as ensuring that their privacy statement mentions that their data will be used to conduct legal checks and may be shared with legal departments (Hewson, 2021).

7. Comparison of Human Review vs. Machine Learning Review

Traditionally, AML officers were appointed to inspect every transaction made by the customer for clearing out if there is any type of money laundering. AML has gotten more complicated than ever; authorities are pushing for automated systems. Thankfully, machine learning-based AML systems can do all of the required activities automatically and efficiently. It used to take a long time to examine each transaction, and it was really stressful. However, when Machine Learning is used, the system minimizes personnel time and filters out the results in a relatively short time by crunching numbers.

If the model is properly trained, the accuracy for detecting suspicious activities will be practically perfect. However, it would be even great if there was a manual review for suspicious activity that the system detected. So, if it was a false alarm, it could be classified as ordinary, or if it was truly worrisome, further investigation could be conducted.

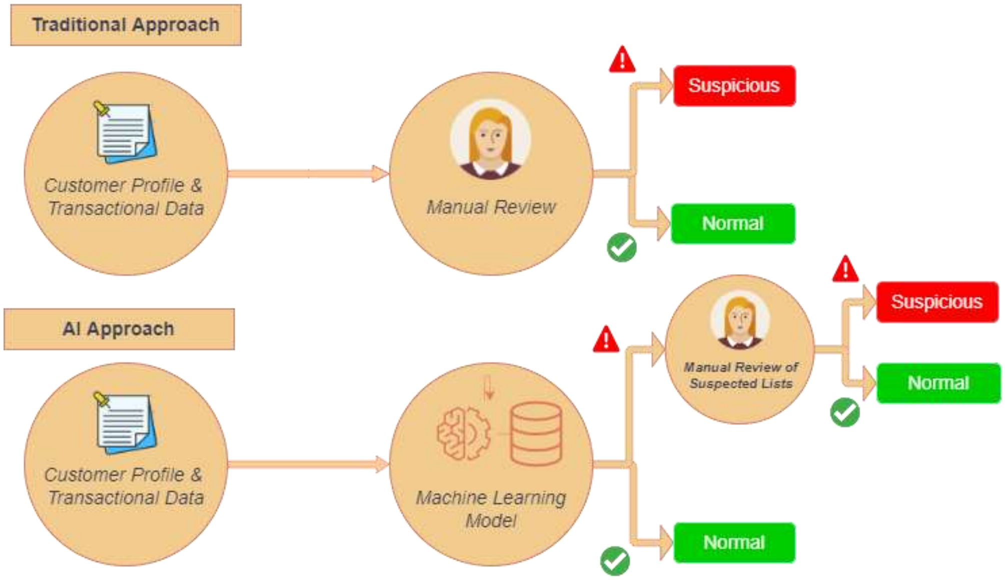
Traditional Approach
Customer Profile & Transactional Data
Manual Review
Suspicious
Normal
AI Approach
Customer Profile & Transactional Data
Machine Learning Model
Manual Review of Suspected Lists
Suspicious
Normal
Normal

Figure 2: Comparison of Traditional & ML system

8. Methodology

Putting all the load in human analysis was in the past, it should be considered for a model to be built by using Semi-Supervised Machine Learning which analyzes each transaction. After the client is on board, their transaction monitoring can also be a great factor to catch up to what the account is being used for. If any activity seems fishy, the transaction can be halted and then required AML officers can investigate further.

When detecting new scenarios, Semi-supervised machine learning comes into play by feeding more data sources and adding features to enable the formation of more intricate cases in order to effectively recognize emergent and uncommon behaviors below the established threshold. It is crucial to consider the recent rise of financial crimes attempting to benefit from vulnerable people due to COVID-19 pandemic, this approach allows robust and accurate detection of unusual activities in the vast uncertainties.

There will exist two model for the analysis which is mentioned briefly in Section “Architecture of the System” below. All the model is trained through a deep learning semi-supervised method called multi-channel convolutional neural network (CNN) which categorizes the sentiment and further used for analyzing the articles. Furthermore, the tweet analysis is also done using a CNN model.

9. Architecture of the System

The proposed system adapts the framework used in (Han, et al., 2018). The system uses a deep learning-based NLP model. It will contain an open standard application layer called Advanced Message Queuing Protocol (AMQP) that is utilized for routing and queueing notes among the services in a manner where it is secure and efficient.

Figure 2 depicts that user Interfaces (UI) are connected to the reporting service which retrieves the information from the database. Furthermore, if there is a new transaction being created then the reporting service can call the main system for further analysis like Sentiment analysis, Tweet Analysis, News article analysis through AMQP.

The database layers contain the information of real time transactions, customer account data, client profiles, KYC data and data from sources twitter & news articles.

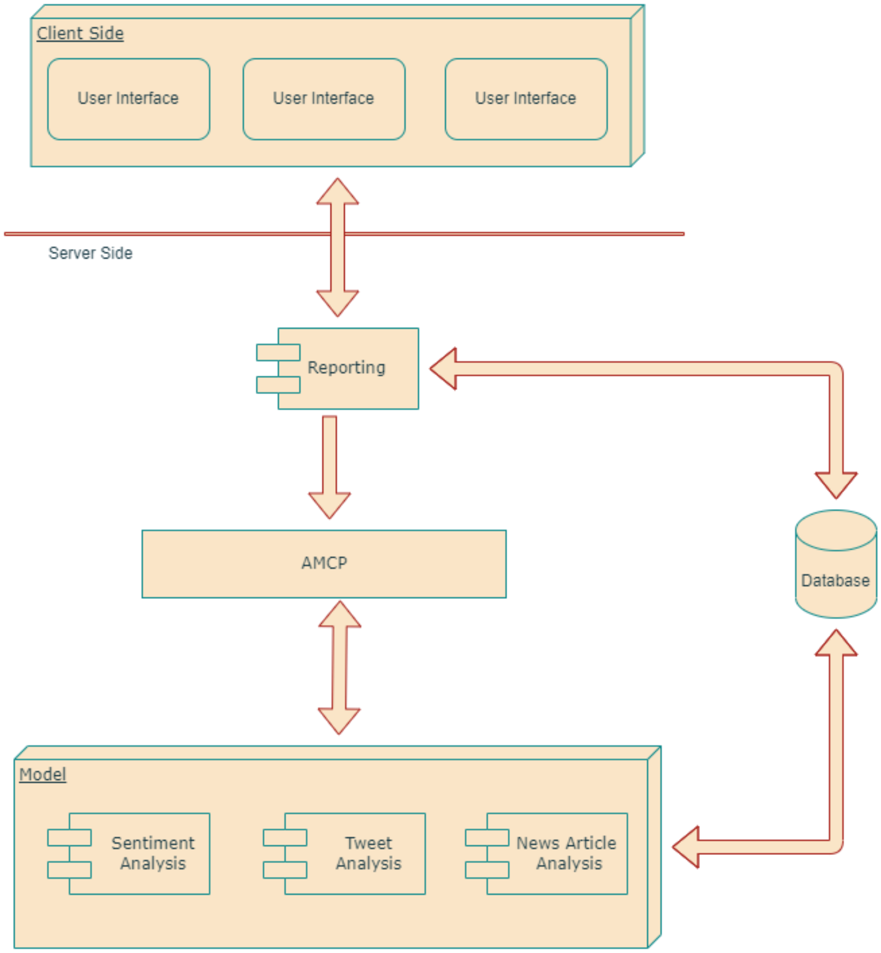

Figure 3: Architecture of the System

9.1 Sentiment Analysis

It is a technique for texts to determine whether the data is positive, neutral or negative using Natural Language Processing. Human express their feelings through texts and verbal communication, so when analyzing feedbacks or reviews of a product sentiments are useful. In our case, Sentiment Analysis will be used for the data retrieved from social media and new article to detect whether texts are positive or negative. If it is negative then the AML team may further investigate on the matter.

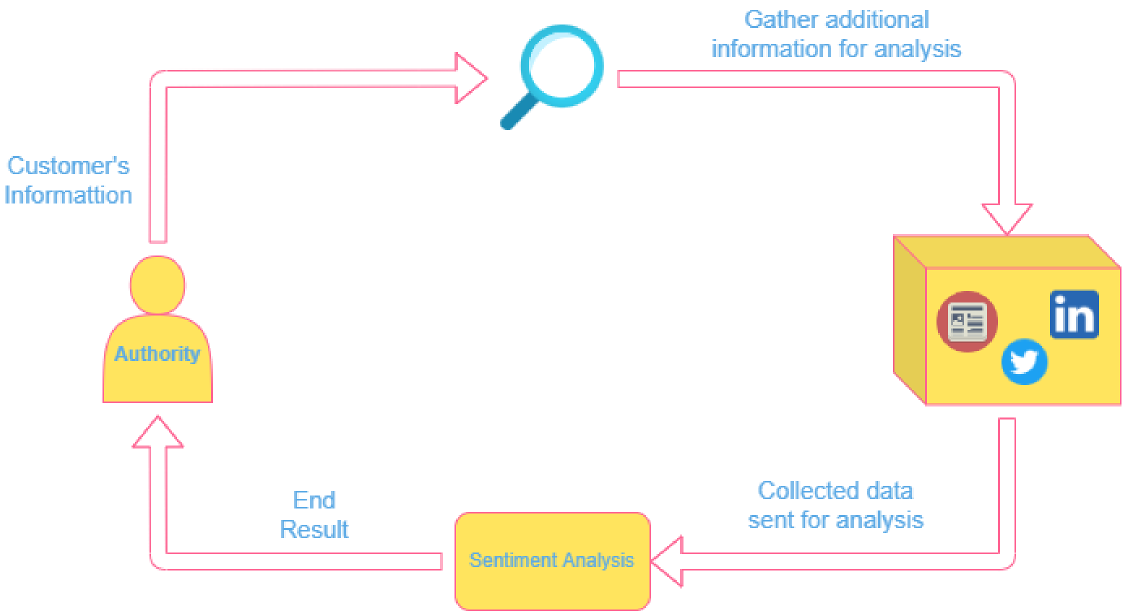
Gather additional information for analysis
Customer's Informattion
Authority
End Result
Sentiment Analysis
Collected data sent for analysis

Figure 4: Sentiment Analysis

Firstly, the customer’s information is fed to the system where it does a brief search on the internet and gather keywords for searching the social media and news article sites for any match. Furthermore, sentiment analysis of the acquired data is conducted and notified to the Authority (in this case bank or the PSP).

Conclusion

To sum up, this report provides a semi-supervised machine learning model for identifying money laundering in financial institutions such as banks and other financial institutions. This model will be more efficient than supervised machine learning models, as per the research. The technology mentioned above is decentralized and extremely adaptable, allowing it to be scaled and adjusted as needed. Every piece in this system is a subset that makes creating, modifying, deploying, and using numerous processes in the same unit easier.

References

Gao, S., Xu, D., Wang, H. & Wang, Y., 2006. *Intelligent Anti-Money Laundering System.* s.l., IEEE.

Glaser, C. & Roberts, K., 2021. *UK Anti-Money Laundering Legislation in a Post-Bexit Landscape.* [Online]
Available at: https://www.jdsupra.com/legalnews/uk-anti-money-laundering-legislation-in-5915630
[Accessed 2 May 2022].

Han, J. a. H. et al., 2018. *NextGen AML: Distributed Deep Learning based Language Technologies to Augment Anti Money Laundering Investigation.* Melbourne, Association for Computational Linguistics.

Hewson, K., 2021. *Data protection and 5MLD.* [Online]
Available at: https://www.shlegal.com/news/data-protection-and-5mld
[Accessed 4 May 2022].

Kenna, A. M., 2021. *Anti-Money Laundering, KYC and Data Protection.* [Online]
Available at: https://www.trilateralresearch.com/anti-money-laundering-kyc-and-data-protection/
[Accessed 06 May 2022].

Singla, S., 2020. *How can machine learning detect money laundering?.* [Online]
Available at: https://medium.datadriveninvestor.com/artificial-intelligence-machine-learning-for-anti-money-laundering-ca896d9fe419
[Accessed 02 May 2022].

The Association of Accounting Technicians, 2019. *Are you being duped? 10 signs of money-laundering.* [Online]
Available at: https://www.aatcomment.org.uk/aatpowerup/anti-money-laundering/are-you-being-duped-10-signs-of-money-laundering/
[Accessed 02 May 2022].

The Law Society, 2021. *Anti-money laundering after Brexit.* [Online]
Available at: https://www.lawsociety.org.uk/en/topics/brexit/anti-money-laundering-after-brexit
[Accessed 2 May 2022].

Thi, M. H., Withana, C., Quynh, N. T. H. & Vinh, N. T. Q., 2020. *A Novel Solution For Anti-Money Laundering System.* s.l., s.n.

UK Finance, 2021. *Fraud - The Facts 2021: THE DEFINITIVE OVERVIEW OF PAYMENT INDUSTRY FRAUD,* s.l.: s.n.