Risk Analysis in Money Laundering A Case Study

Joana Filipe Oliveira Marques
joanamarques@tecnico.ulisboa.pt

Instituto Superior Técnico
Lisbon, Portugal
November 2015

Abstract — **Money laundering has a high worldwide impact that distorts and seriously spoils the economic system, because it can completely deteriorate countries' political and financial systems. Risk management therefore becomes a crucial element for the detection of practices that may indicate money laundering. In this context, the main goal of this paper is to develop two models based on data mining techniques that claim to provide the basis of a detection system of suspected cases of money laundering - the Risk Assessment Model of New Entities and Clients, and the Risk Assessment Model of Client Behavior. The first model results from the need to obtain a client's risk level in a first contact, in order to support the decision making by the bank. For its development, a Regressive Model of the risk assigned to entities based on recommendations from several international institutions is firstly derived, because there is no historical characterization of entities, and then a risk rating is assigned to each entity and client using logistic regressions. The second model groups a set of clients with similar characteristics in terms of transactions, applying clustering methodology. For the development of clusters it is necessary to analyze all transactions of each client. This allows detect when a client has a deviant behavior, which may indicate a suspicious transaction of money laundering.**

Key words: ***Money Laundering, Risk, Risk Management, Data Mining Techniques***

I. Introduction

Risk management is a key strategy present in various business sectors. It deals with the uncertainty and the impact associated with future events that might be envisaged [1]. One of the main problems faced by a banking institution is money laundering. This is a crime that consists of the concealment of certain assets unlawfully origin so that they appear to be from a legitimate source. The money laundering process is composed by three phases: placement, accumulation and integration. The first phase consists of placing the capital from criminal acts into the financial system, whereas the second phase ends in the integration of that capital in the legitimate economy by moving the capital to other financial institutions [2].

The awareness of this problem has meant that, in recent years, there has been an increase on anti-money laundering measures, and monitoring them allows the bank to protect its reputation and the system's integrity. This research focuses on the risks of money laundering in the banking institution XY[1] in Angola. In this country, though there are measures to prevent money laundering, some key points are not in agreement with the international procedures. However, Angola's efforts to put into practice preventive measures have increased. In September 2014 Angola became president of the Eastern and Southern Africa Anti-Money Laundering Group (ESAAMLG) which belongs to the Financial Action Task Force (FATF). The purpose of these institutions is to combat money laundering and cooperate with other international organizations [3]. The National Bank of Angola (BNA) is the responsible for the financial regulations of the country and its main focus is to ensure the preservation of the value of the national currency and participate in the definition of monetary, financial and foreign exchange policies [4]. The study of money laundering risk should be based on three main types of risk: geographic and country risk, entities and clients risks and, lastly, product and transactions risk.

There are several international organizations fighting for an anti-money laundering regime. Some of the most important are:

- Financial Action Task Force (FATF) – This institution was set up in the framework of the Organization for Economic Co-operation and Development (OECD) and has 36 governmental members from several countries. Its responsibility is to develop a standard international normative for combating money laundering. It is also responsible for the creation of the Financial Information Unit (FIU), which constitutes the link between financial institutions and legal entities [5];
- United Nations (UN) – this organization provides tools and conventions to address money laundering [6];
- Egmont Group of Financial Intelligence Units (FIU) – this group is composed of FIUs and recognizes the benefits inherent in the development of a network of financial units that stimulates national cooperation [7];

[1] In order to preserve confidentiality, this banking institution will be here and after known as XY.

- Basel Committee on Banking Supervision – It provides guidelines for banks on various issues, including money laundering [8];
- Wolfsberg Principles – It consists of a series of anti-money laundering principles [9];
- Eastern and Southern Africa Anti-Money Laundering Group (ESAAMLG) – It aims to establish regulations for all countries of the region and ensure the implementation of effective crime-fighting systems. This organism works under the FATF and the UN [10].

The crime of money laundering becomes more sophisticated with the advancement of technology and it is increasingly challenging for the banking institutions to develop measures that can be effective in terms of prevention. It is difficult to statistically evaluate this kind of illegal operations because they can occur in several countries simultaneously, enforcing on the criminal different policies on prevention of money laundering. For this reason it becomes difficult to understand the magnitude of this problem and it is difficult to create measures and strategies in a global level to fight money laundering.

However, it is relatively easy to see the way that money laundering affects the economy's security and social stability [11]. In an economic level it causes a decrease in tax revenues and makes the federal tax revenue more difficult to raise; to the developing countries the consequence is even greater due to their fragile financial system [12]. At the microeconomic level the problem is that "front companies" carry out fictitious activities to mix illicit income from illicit activities with legitimate funds, to hide the criminal revenue [13].

The goal of this paper is to provide a thorough understanding of the risks of money laundering to which a bank is exposed in order to support the development of models for identification, analysis and management based in alerts that can prevent and mitigate these risks. Hence, two risk assessment models are developed: (1) *Risk Assessment Model of New Entities and Clients*, to support bank's decision accepting new entities at first contact; (2) the *Risk Assessment Model of Client Behavior* so that the bank could monitor the transactions of its clients and their suspicious actions with regard to behavioral deviations in relation to the expected. In order to better understand the structure of this paper, it is necessary to clearly define the concepts of *entity* and *client*. An entity is a person with certain characteristics such as profession, nationality and country of residence, whose identification may be made by the Identity Card. A client is the product / behavior that one or more entities have to the bank. The identification number is a client account. An entity may be associated with one or more clients, as well as a client may be associated with one or more entities.

The end result of this work will cover entities and clients of the XY bank, with the possible acknowledgement of risk of money laundering, and thus, making easier and safer decisions regarding those same clients and entities. The methodology adopted is as follows:

1. Define the universe and select the variables that will be part of the models. In this context both entities and clients, privates and corporates, will be considered for each one;
2. Develop a Risk Assessment Model of New Entities and Clients which will have two stages – first it is necessary to build a Regressive Model of Risk Assessment, which aims to create a historical characterization of the entities of the XY bank based on its profession, country of residence and nationality risks, and after that develop a Risk Assessment Model of New Entities and Clients using the *software* Prospero that will be able to predict the risk of a particular entity and client;
3. Develop a Risk Assessment Model of Client Behaviour, through the analysis of conduct and transactions of the XY bank clients, so that it is possible to group clients with certain characteristics (clusters) and evaluate them when they are deflecting from their corresponding cluster. Grouping clients is done through a decision tree.

II. Literature Review

i. Money Laundering

Money laundering is the act of concealing the true origin of funds derived by criminal sources. This act tends to appear as from a legitimate source so that there is no suspicion. When there is money laundering it is assumed that there was a criminal process at its root, and this could be trafficking illicit substances (weapons, organs, drugs, etc.), tax evasion, corruption or terrorism. There are several reasons that drive criminals to commit money laundering. The first reason is that any criminal act requires an investment to cover the operation. This investment needs to be hidden in order to camouflage the source of wealth and ensure that illicit proceedings are not used to denounce that crime. Subsequently, the products resultant of crime, become subject to apprehension and investigation and it is necessary to cover the origin of the funds that appear legitimate [11].

ii. Politically Exposed Persons (PEPs)

The FATF believes that PEPs are individuals who are associated with public functions (heads of state or of government, senior policy frameworks, government, judicial or military officials, and senior executives of companies and key employees of political parties) and people who are senior members of international organization in foreign or national countries. Nevertheless, relatives or people who have financial or business relationships with PEPs are also considered PEPs.

There is a consensus on this revised context, in that most countries adopt this treatment methodology on these individuals. However, in Angola this methodology was not considered and they only considered foreign politically exposed persons who exercised political functions until a year ago. This is somewhat problematic because, using this methodology, an Angolan individual can be a PEP worldwide, except in his own country, limiting the prevention and mitigation of risk of money laundering of the country's capital [5].

iii. Compliance and diligence

The Compliance Officer is responsible for the implementation of a money laundering prevention program, and he is also

responsible for handing over information and communication susceptible to money laundering operations to the Financial Information Unit and other competent authorities. The compliance officer mandatory functions are [3]:

- To obtain the approval of the anti-money laundering program and terrorist financing;
- Monitor the performance of established policies and procedures under prevention of money laundering offenses implemented by the financial institution;
- Manage and monitor the implementation of controls related to the prevention of money laundering crimes;
- Centralize and analyze internal communications received;
- Communicate suspicious operations to FIU;
- Receive requests from FIU or any other competent authority and, if applicable, respond to the information requested;
- Prepare an annual report of the compliance risk assessment retrieved by the financial institution as well as gather the effectiveness of the implemented measures for the prevention the anti-money laundering.

iv. Risk, threat, vulnerability and consequence

In recent decades, some standards and legislation have been adopting different risk settings. In many cases the risk is confined to the concepts of damage and loss, however, Australia and New Zealand have developed a standard that defines the risk not only in terms of something that might happen, but also relate it to the impact on the objectives of an organization. The standard AS / NZS 4360 [14] defines risk as "the hypothesis of something happening that will have an impact on a particular goal." This standard was adopted throughout the world and its widespread acceptance led to the development of the first global standard for risk management, ISO 31000 [15]. This was published in 2009, with a revised set of definitions that are accommodated in ISO / IEC Guide 73 [16] in "Vocabulary for Risk Management". According to ISO / IEC Guide 73 [17], the risk is "the effect of uncertainty in some objectives." Both the old and the latest settings put the concept of risk in what all organizations wish to achieve: one's goals. I.e. an organization must define its goals, however, to reach them it will have to deal with internal and external factors that are often impossible to control, generating uncertainty.

A threat has the potential to harm assets such as information, procedures and systems, and hence the organizations. In the context of money laundering, a threat also includes the crime agents, terrorists and those who facilitate criminal activities. Vulnerability can be defined as a property of the system, which is susceptible to threats, i.e. everything that can be exploited by threat or that can support or facilitate the criminal activity. The consequence is the impact when a threat is executed, which affects the institutions, economy and society [18].

v. Risk factors of a financial institution

a. Geographical or country risk

The country risk is directly related to sanctions, embargoes and non-cooperation with anti-money laundering policies which often links with terrorist acts, corruption or other criminal activities. These countries are easily identified by organizations such as the UN and the FATF [9].

b. Client Risk

Regarding the risk associated with the client, it is often difficult to have the perception of this risk of a new client and this risk is only realized after the beginning of the movement of capital. However, there must be a profile associated with each client in relation to their behavior. The client risk profile depends on the perceived risk by the institution. For instance, a client who deposits or transfers a high volume of capital and does not fit the risk profile associated, should be treated as a high risk client, despite not having this profile and vice versa. Another example is the duration of the relationship between the client and the bank: long-term relationships provide frequent contact with the client, allowing a better understanding of the client and consequently representing a lower risk of money laundering [9].

c. Risk products / services

The risk associated with the products/services is related to the types of products and/or services offered by the financial institution. The institution understands that higher risks are taken with implementing, for example, a new product or service such as international transfers between banks and transfers that require anonymity.

vi. Risk variables on money laundering

There are some variables that can increase or decrease the risk perceived by the bank of a client or a transaction and these could be, for example: the main goal of a new bank account or relationship, the amount deposited in a particular account, the regulation level of the country from which a client wants to open an account and the duration of the relationship between the client and the bank [9].

vii. Risk-based approach vs regulation-based approach

An approach to the problem of money laundering based on risk gives financial institutions more freedom regarding the application of rules and their exceptions. On the other hand, an approach on money laundering based on regulations encompasses most government decrees that impose rules and guidelines that must be applied. A rules-based approach, in theory, will be more efficient. However, the laws sometimes are too vague, dubious and financial institutions are not able to distinguish transactions generated from crimes and those which did not. This resulted in the thought that this approach is not enough in the fight against money laundering [19]. The fact that money laundering may have a risk-based approach requires that an acknowledgement of risks is held as well as a prehension of risk assessment and a development of strategies to meet them [20].

viii. Cost-benefits analysis

The cost of the application of anti-money laundering measures may be compared to its benefits. However, as mentioned above, it is not easy to translate in a straightforward manner this cost/benefit. Since resources are always limited, it is necessary

to decide if the cost covers the associated benefit [21]. The policies have direct costs, but also have collateral damage with unwanted effects for the economy. Indirect costs are supported by all clients, and not only by criminals. From the point of view of management, investment in the prevention of money laundering is in line with risk management strategies of the bank. It is a trade-off that banks are able to afford not to take certain risks [22].

ix. Models

The first step in the investigation of money laundering suspicious activity is the analysis of databases. This analysis are critical in order to understand the relationships among the data that needs to be analyzed. This analysis is not at all effective as there are several constraints such as cost, human resources, time, etc. This analysis require technology systems that allow dealing with a large volume of data and enables the automatic identification of pathways that lead to the detection of suspected cases of money laundering [23].

The linear regression is the most basic data mining approach. A dependent variable (output) and a number of independent variables (input) is defined. The equation has the following format: $Y = \beta 0 + \beta 1X + e$, where "X" is the explanatory or independent variable, "e" is the random error, "β0" and "β1" are unknown parameters that the model is intended to estimate and "Y" is the variable explained or dependent. This equation is a line that fits the best data set, which will be useful for forecasting purposes. The goal is to validate and quantify the trends of previous cases of money laundering and transpose this trend to current cases. This technique revolves around the foresight of probabilities, that is, what we have previously observed in the data can easily be transposed to the current data [24]. In specific situations the application of a simple linear regression occurs from the necessity to involve more than one independent variable. In this case, a multiple linear regression, the corresponding equation is: $Y = \beta 0 + \beta 1X1 + \ldots + \beta kXk + e$ and the explanation of each of the parameters is precisely the same of simple linear regression. Logistic regression is a model that is very common for cases containing categorical variables, which can easily be applied to money laundering. Categorical variables are those that can be described, for example, yes/no or male/female. This technique is used, for instance, to assess the financial transaction records that have an interest in being investigated. The result of the application of the logistic regression is a series of graphs that allow to find similarities between the sets of selected variables. The cluster analysis is a technique that can be used to investigate large volumes of data and the relationship between them. It is used when it is necessary to group similar data attributes into categories in order to automatically detect the relationship between data groups, without someone expert. From the similarities between historical data, conclusions about the current data can be drawn. This analysis is useful for this research as it ensures a deeper understanding on the aggregates and it can statistically isolate relations between suspected cases, in a chronologically order [25]. The software "Prospero" is used in the construction of most models of this paper. The methodology used by the software is based on Random Forest Method developed by Leo Breiman and Adele Clutler [26], i.e., it is an essentially continuous learning method used in classification method cited above, and in some regressions. With this method it can control the stability of the model, and is possible get quite accurate results [27]. For the development of risk assessment models of new entities and clients was used various iterations of logistic regressions based on the previous methodology (Random Forest Method). For the development of risk assessment models of client behavior were used the Self Organizing Maps algorithm, enabling the grouping and classification of data. This algorithm has the ability to organize complex data in clusters, according to similar attributes [28].

III. SOLUTION FOR THE PREVENTION OF MONEY LAUNDERING

Two models will be exposed which propose a solution for the prevention of money laundering: the Risk Assessment Model of New Entities and Clients, and the Risk Assessment Model of Client Behavior. The first one will be performed for the entities and clients of the XY bank and it is a predictive model based on statistical regression using the software Prospero. First, was assigned a level of risk to current entities of the bank according to their profession, country of residence and nationality. Second, a regressive model of risk assessment was developed from this classification. Third and last step, was developed a risk assessment model for new entities with the software Prospero. Through the use of several variables that characterize each entity and client (privates or corporates), it is possible to obtain an equation that is able to evaluate each in terms of their level of risk. The Risk Assessment Model of the Client Behavior is a model based on clusters, also using Prospero software, which with Random Forest and the method of Self Organizing Maps allows to group clients according to their behavior in terms of transactions. This model allows each cluster to have an associated risk, and those with higher risk are targets of special attention. Whenever there is a deviation from the expected behavior, there is a warning of suspected money laundering. In Figure 1 you can see a simplified form of what is proposed:

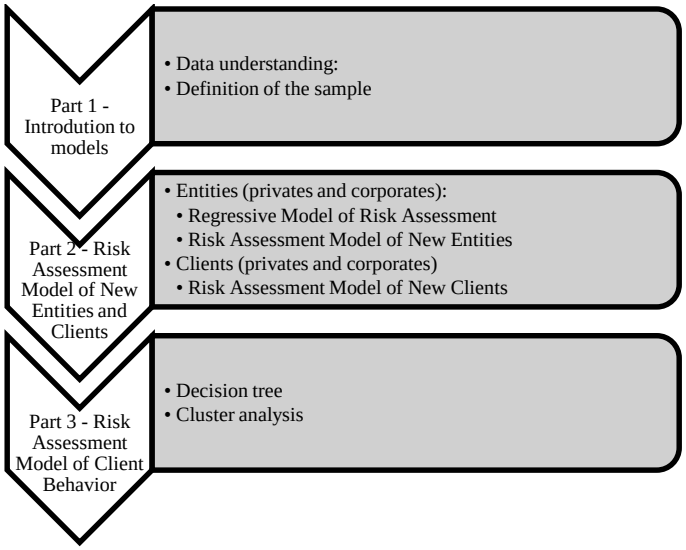

Figure 1- Summary diagram of the proposed solution

At this stage, three components are introduced: data understanding, the Risk Assessment Model of New Entities and

Clients and the Risk Assessment Model of Client Behavior. In the first part, a detailed description of the data sample that will be used in the models is shown. As previously mentioned, the distinction between entities and clients will be made as well as between privates and corporates. The second part corresponds to the preparation of the Risk Assessment Model of New Entities. For its development, it will be necessary to perform a Regressive Model of Risk Assessment, because there is no historical characterization of the entities. This means that a model was carried out to define a level of risk that will be the dependent variable of the Risk Assessment Model of New Entities. Three variables were used to build the regression in Microsoft Excel: the profession, nationality and country of residence. These variables were chosen because, when mentioned to the XY bank, they were considered as required fields upon registration of a new entity. Later in the Risk Assessment Model of New Entities there will be more variables used, with Prospero software, that will have the same significance in relation to the three already mentioned. This way it will be possible to get an equation able to perform a risk classification from 1 to 5. The last step was repeated, but this time with some variables that characterize. In the third part will be to develop the Risk Assessment Model of Client Behavior, which will also use the Prospero software for the development of clusters. The aim is to group clients with similar behaviors, forming several clusters with an associated risk, and whenever a client has a behavior that is not in line with its cluster, it will be an indication of money laundering.

i. Data understanding

a. Definition of the sample

For the risk assessment model the sample is composed of 6656 private entities, 6081 corporate entities, 49717 private clients and 6925 corporate clients.

ii. Risk Assessment Model of New Entities

The Risk Assessment Model of New Entities is used to, at the time of establishment of relationship between the XY Bank and the entity, define the level of risk that it represents. As mentioned previously, this model has a predictive basis and can evaluate new entities of the Bank. To generate the model that have the capacity to evaluate a new entity in terms of risk, one needs a historical data so that the prediction of future events can be made. Therefore, it is necessary to develop a regressive model for the entities, where the goal is to create a historical characterization of entities in regards to its level of risk, since it does not exist in the XY bank. In the end, it is possible to differentiate the XY bank entities as to their level of risk.

a. Selection of variables

The variables of interest to this model, according to the presented sample for private entities are: nationality, country of residence and profession. And for corporate entities: headquarters, country of residence, economic activity code (CAE) and secondary economic activity code (CAE2).

b. Regressive Model of Risk Assessment

To design this model, Microsoft Excel was used to develop a multiple linear regression. The final risk for each private entity is given by a risk level of 1 to 5, where the entities that are located between level 4 and 5 are considered high risk and labelled as bad cases. The risk levels of the values are rounded nearest integer and to zero decimal places.

- Private Entities

For this model, 6656 private entities were used, detailing for each one the risk of nationality, country of residence and profession. For each one of the entities, and based on the risks of the variables mentioned above, a risk is assigned. This risk is variable and has been allocated manually for being an expected risk to the entity. Thus, the risk assigned is the dependent variable of the linear regression and the other three variables (nationality, country of residence and profession) are the independent variables.

The equation calculated for the sensitivity model for the private entities was the following:

$$\text{Risk Level} = 0.3286 * \text{Nationality} + 0.1995 * \text{Country of Residence} + 0.5004 * \text{job} \quad (1)$$

The meaning of the variables in equation 1 is as follows:

- Nationality = Risk of the entity's Nationality;
- Country of Residence = Risk of the entity's Country of Residence;
- Profession = Risk of the entity's Profession.

The validity of the established equation is quite high because the indicators obtained in the regression have shown it. An R^2 of 0.9908 was achieved, meaning that the variables used to calculate the regression explain 99.1% of the risk level obtained. The T-Stat values show that the variables are significant for a confidence interval of 95%. The tStat values are shown in the following table:

Table 1 - tStat values for private entities

Variable	Value
Nationality	11.05
Country of Residence	6.23
Profession	73.43

The values observed in the table above indicate that all variables were relevant to the model, and the profession is the most significant. The results distribution of the regressive model for private entities for the selected sample are as follows:

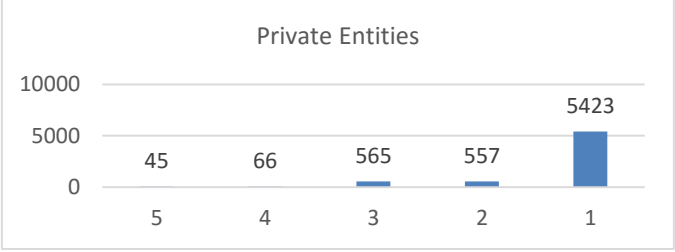

Figure 2 - Distribution of the results of the regressive model for private entities

The percentage of bad cases is about 1.67% which corresponds to the anticipated when taking into account the particular

characteristics of entities existing in the database of the XY bank.

- Corporate Entities

For this model 6081 corporate entities were used, specifying for each one the risk of headquarters, country of residence, CAE and CAE2. For each of the entities, and based on the risks of the variables mentioned above, a risk is assigned. Thus, the risk assigned is the dependent variable of the linear regression and the other four variables (headquarters, country of residence, CAE and CAE2) are the independent variables.
The equation calculated for the regressive model for the corporate entities was the following:

$$\text{Risk Level} = 0.2729*\text{Country of residence} + 0.6664*\text{CAE} + 0.0943*\text{CAE2} + 0.1624*\text{Headquarters} \quad (2)$$

The meaning of the variables in equation 2 is as follows:
- Country of Residence = Risk for the entity's Country of Residence;
- CAE = Risk for the entity´s CAE;
- CAE2 = Risk for the entity´s CAE2;
- Headquarters = Risk for the entity's Headquarters.

Highly acceptable results were obtained, as it can be seen with the R^2 and T-stat parameters. R^2 had a value of 0.8957, meaning that the variables used to calculate the sensitivity model for corporate entities explain about 89.6% of the risk level achieved. Also the T-stat values show that the variables are significant for a confidence interval of 95%. The tStat values can be seen in the following table:

Table 2 - tStat values for corporate entities

Variable	Value
Country of Residence	56.49
CAE	212.74
CAE2	18.77
Headquarters	24.44

The values observed in the table above indicate that all variables were relevant to the model, and the CAE is the most significant. The results distribution of the regressive model for corporate entities, for the selected sample are as follows:

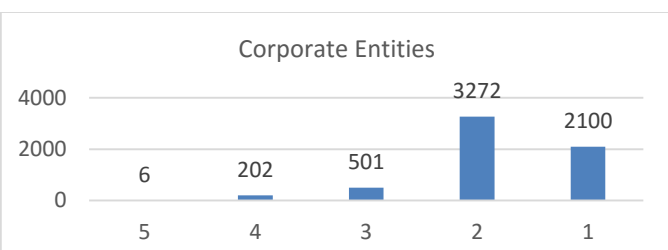

Figure 3 - Distribution of the results of the regressive model for corporate entities

The percentage of bad cases is an expected 3.4% when taking into account the characteristics of corporate entities existing in the database.

c. Risk Assessment Model of New Entities

Having obtained the historical characterization of the entities through the regression model of risk Assessment, the tools are now available to develop the Risk Assessment Model of the New Entities.

- Private Entities

For this predictive model other variables were added in order to add significance to the model. The new variables are: birth date, gender, education, married status, matrimonial regime, annual income, identification type and professional activity.

Testing phase

At this stage several models will be tested to find the one that best suits the original purpose. After preparing all variables for them to be trialed and using the Prospero software, it is possible to start testing the risk model.
As the dependent variable, the risk level obtained in regressive model will be used for it is the variable that we want to set.
After several tests, the model chosen was the one with the following characteristics:
Variables – nationality, professional activity, country of residence, birth date, gender, education and annual income.
Test quality – R^2=98%; Accuracy=99.98%; AuROC=1; K/S Statistic=0.9997; GINI index=99.77%.
The equation for the predictive model of particular entities was the following:

$$R\_Ent\_I = 1/(1+\exp(4.59732-SUM)) \quad (3)$$

Equation 3 provides the level of risk of a new private entity of the XY bank. The value of SUM is calculated with a combination of conditions of the independent variables and their respective coefficients.
When any new entity has a risk of 0, assigned by default, it means that that entity has not been assessed by the model on that occasion. These entities will be assessed with a risk level between 1 and 5.
Because the main goal is to find the bad cases, with the software "Prospero" it is possible to define them in terms of percentage, given previous results. Thus, all values above 57.34% for this model are defined as bad cases, i.e., with risk 4 or 5. It can be seen the following results:

Table 3 - Risk Results for Risk Assessment Model of New Private Entities

Risk Level	Class	Number of cases
1	0 - 10.01%	1883
2	10.02 - 10.08%	3540
3	10.09 - 57.34%	1123
4	57.35 - 91.27%	104
5	91.28 - 100%	6

The number of bad cases is 110 i.e., which means that each one of these entities achieved a score above 57.35% which corresponds, in this model, to classes 4 and 5.

As demonstrated, these 110 cases represent 1.7% of the total sample, meaning that in approximately 100 entities, 2 are at high risk.

- Corporate Entities

In order to add significance to the model the following variables were added: type of company, date of incorporation, social capital and type of identification.

Testing phase

The method was the same as for private entities. As the dependent variable, the Risk Level obtained in sensitivity mode will be used for it is the variable that we want to set.

After several tests, the model chosen was the one with the following characteristics:

Variables – country of residence, headquarters, CAE, CAE2, type of company and social capital.

Test quality – R^2=70%; Accuracy=99.95%; AuROC=1; K/S Statistic=0.9856; GINI index=98.56%.

The equation for the predictive model of particular entities was the following:

$$R\_Ent\_E = 1/(1+\exp(1.99104-SUM)) \quad (4)$$

Equation 4 provides the level of risk of a new corporate entity of the XY bank. Like the previous model, the value of SUM is calculated with a combination of conditions of the independent variables and their respective coefficients. Similar to what was done to the model of new privates entities, and through the "Prospero" software, was set the percentage of bad cases. In this way, all values above 48% in this model are defined as bad cases. For the sample used, the following results were obtained:

Table 4 - Risk Results for Risk Assessment Model of New Corporate Entities

Risk Level	Class	Number of cases
1	0 - 9.69%	245
2	9.70 - 17.08%	2572
3	17.09 - 47.99%	3045
4	48 - 65.83%	201
5	65.84 - 100%	18

The number of bad cases is 210 i.e., which means that each one of these corporates achieved a score above 48% which corresponds, in this model, to classes 4 and 5.

As expressed, these 210 cases represent 3.6% of the total sample, meaning that in approximately 100 entities, 4 are at high risk.

iii. Risk Assessment Model of New Clients

The Risk Assessment Model of New Clients Model is used to, at the time of establishment of the relationship between the XY Bank and the client, define the level of risk it represents. This model was developed using the same techniques of the previous model.

a. Selection of variables

Private and corporate clients: maximum risk associated entities, entity's risk first holder, minimum risk associated entities, the difference between the maximum risk and minimum of associates, number of different relationships that make up the client, number of associated entities that are not holders, product, product subscription channel and segment type.

- Private clients

The software is the same used in the preceding models for entities

Testing phase

After several tests, the model chosen was the one that had the following characteristics:

Variables - entity's risk first holder, minimum risk associated entities, the difference between the maximum risk and minimum of associates, number of different relationships that make up the client, number of associated entities that are not holders, product and product subscription channel.

Test quality – R^2=79%; Accuracy=99.73%; AuROC=0.9953; K/S Statistic=0.9997; GINI index=98.76%.

The equation for predicting the model of private clients was the following:

$$R\_Cli\_I = 1/(1+\exp(1.73224-SUM)) \quad (5)$$

Equation 5 provides the level of risk of a new private client of the bank. Like the previous model, the value of SUM is calculated with a combination of conditions of the independent variables and its respective coefficients. Through the same method used for the risk Assessment model of new entities, was obtained with Prospero software that values above 41.25% for this model are defined as bad cases. It can be seen in the following table:

Table 5 - Risk Results for Risk Assessment Model of New Private Clients

Risk Level	Class	Number of cases
1	0 - 30.25%	44853
2	30.26 - 30.4%	3109
3	30.41 - 41.24%	1169
4	41.25 - 81.4%	551
5	81.41 - 100%	35

The number of bad cases is 586 i.e., which means that each one of these entities achieved a score above 41.25% which corresponds, in this model, to classes 4 and 5.

As demonstrated, these 586 cases represent 1.18% of the total sample, meaning that in approximately 100 entities, 1 is at high risk

- Corporate clients

The software used to run the models is identical to the one used previously.

Testing phase

After several tests, the model chosen was the ones that had the following characteristics:
Variables - entity's risk first holder, minimum risk associated entities, the difference between the maximum risk and minimum of associates, number of different relationships that make up the client, number of associated entities that are not holders, product and product subscription channel.
Test quality – R2=98.5%; Accuracy=100%; AuROC=1; K/S Statistic=1; GINI index=98.65%.
The equation for the guessed model of particular clients was the following:

$$E\_Cli\_E = 1/(1+\exp(4.06744-SUM)) \quad (6)$$

Equation 6 has the capability to show us the level of risk of a new corporate client of the bank. Like the previous model, the value of SUM is calculated with a combination of conditions of the independent variables and its respective coefficients.
Like previous model, the following results were obtained to calculate the bad cases:

Table 6 - Risk Results for Risk Assessment Model of New Corporate Clients

Risk Level	Class	Number of cases
1	0 - 30%	1778
2	30.08 - 30.41%	4545
3	30.42 - 79.80%	400
4	79.81 - 88.01%	156
5	88.02 - 100%	46

The number of bad cases is 202 i.e., which means that each one of these corporates achieved a score above 79.81% which corresponds, in this model, to classes 4 and 5.
As exhibited, these 202 cases represent 2.9% of the total sample, meaning that in approximately 100 entities, 3 are at high risk.

iv. Risk Assessment Model of Client Behavior

For the construction of the behavioral models it was necessary to use a different sample than the ones used in previous models (risk assessment models) which had a predictive basis. For these models it is fundamental to analyze the related transactions made by the clients in order to be able to assess their behavior. Therefore, all transactions for 22,600 private clients and 3736 corporate clients of the XY bank in the period of one year were used.

For the segmentation of these models, particularly for the creation of clusters, variables such as the counter, client number, operation code, transaction currency, etc., were used.

a. Private clients

For clustering clients all transactions were used, and through various calculations made by Prospero software several clusters of clients were identified, as well as the respective risk cluster.
After several tests, it was possible to obtain 32 different clusters, and the cluster which had the most clients was number 30, with 2610 clients. It is also possible to verify that the highest risk cluster is number 1, which has about 97% probability of risk (money laundering), which is included in risk class 5. It is also possible to establish that the cluster with lower risk is number 32 with about 0.0004% probability of risk. The accuracy of the model is 89.63%
All these client groups have been formed by software "Prospero" using the Self Organizing Groups methodology already mentioned above, through a decision tree that defines the similarities between them and which has the following format:

(min_ScaleCoef > 0.0242)
yes->(sum_MXVLRT_debit > -478.5372)
yes->(AKZ_max_MXVLRT_total > -0.5001)
yes->(max_ScaleCoef > 4.9873)
yes->(max_MXVLRT_debit > -1.9110)
yes->(min_MXVLRT_credit > 51.0275)
yes->(cluster: 12)

Taking this example, if the amount of transactions below 1,000 USD is greater than 0.0242, if the debit transaction value is greater than the -478.5372, if the maximum transaction value exceeds in -0.5001 kwanzas, if the total of transactions below 1,000 USD exceeds 4.9873, if the maximum debit transaction value is greater than -1.9110, and if the minimum credit transaction value is also greater than 51.0275, then the client will be placed on cluster number 12.

b. Corporate clients

Similar to private clients, in the clustering stage of clients all transactions were used, and through various calculations made by the p.Modeler Prospero software several clusters of clients were identified, as well as the respective risk cluster.
After the tests, 23 different clusters were obtained, and the cluster that had the most clients was number 18, with 615 clients. It is also confirmed that the cluster of increased risk is number 1 with a 99% risk probability, which includes risk class 5, and the cluster of lower risk is number 23 with 0.01% probability of risk.
Similar to the clusters for private clients, all clusters have been formed through a decision tree that categorizes similar clients in each group.

IV. Conclusions

Money laundering is a reality which impacts worldwide. It distorts and seriously impedes the economic system and in extreme cases may completely corrupt the countries' political and financial systems, affecting their development. Thus, several international organizations have been cooperating in order to regulate and create prevention practices, which can be applied in a comprehensive and global manner. In an increasingly interconnected world, the negative effects of these activities are global, and their impact on the integrity and financial stability of countries is widely felt.
So, this investigation aims to solve the adaptation of the XY bank to internationally standardized rules for the prevention of money laundering by developing models which make it possible to create alerts in the cases of higher risk. With these models it would be possible to assess entities and corporates by assigning a risk to each of them, set between 1 and 5; those that

are located in risk levels 4 or 5 are designated as bad cases. With this solution, the XY bank will be able to comply with the instructions of the National Bank of Angola regarding the prevention of money laundering practices. The development of the proposed models includes a process of selection and evaluation of multiple variables, in order to obtain the best possible adhesion. The variables used for models relating to entities are: nationality, country of residence and profession, to private entities; headquarters, country of residence, IAC and IAC2 to corporate entities. In turn, the variables used for the related client models are as follows: maximum risk associated entities, entity's risk first holder, minimum risk associated entities, the difference between the maximum exposure and associated entities minimum, number of different relationships that make up the client, number of associates who do not hold, product, product subscription channel and segment type (all of them to private and corporate clients). The first model, the Risk Assessment Model of New Entities and Clients aims to evaluate an entity or a client in a first contact with the bank. Initially, since there is a historical characterization of privates and corporate entities, it developed a Regressive Model of Risk Assessment based on recommendations of international institutions. This model is based on multiple linear regression using Microsoft Excel. The result was two equations able to characterize the risk of private and corporate entities. The first equation for private entities, obtained an R^2 of about 99%, meaning that the variables used to calculate the regression explain 99% of the achieved level of risk. The second equation, corporate entities, obtained an R^2 of about 90%. Having been brought up a historical characterization of private entities and corporates in terms of risk level, the conditions are met to develop the Model of New Risk Assessment Entities. For this model, we used the software "Prospero", allowing the performance of various tests in order to obtain the best results. Two equations, which incorporate several variables able to evaluate a new private entity and a new business entity as to the risk that they pose to a banking institution, were then obtained. Two representative risk equations were obtained with an accuracy level of 99.98% and 99.95% for private entities and corporate entities, respectively. From the results obtained, and according to the universe of data available, 1.7% of bad cases of particular entities and 3.6% of cases of bad corporate entities were obtained. Having already evaluated entities, it is possible to evaluate clients, and these have features of associates and will be considered in the model. In this model, we also obtained two representative equations of private clients and corporate clients. The first equation was obtained with a 99.72 % accuracy (private clients), resulting in 1.2% of bad cases, and the second equation obtained a percent accuracy of 100%, with 2.9% of bad cases.

The Risk Assessment Model of Client Behavior aims to group clients according to their behavior in the bank, through transactions carried out. To develop this model the software "Prospero" was also used, by creating clusters. After several iterations 32 different clusters to private clients were obtained, classified as risk levels 1 to 5, with an accuracy of about 90%. For corporate clients 23 different clusters were obtained, these also ranked from 1 to 5, with an accuracy of about 85%.

Thus, taking into account the limitations encountered, the results obtained were extremely positive with regard to the objective. However it is possible to improve them with the use of a larger sample, which enables integration of more variables in the models.

These models are intended to be integrated in a web application designed to be used by employees of the bank, where they can perform all the required functions of entities and clients management. This enables the financial institution to perform a decision-making based on relevant statistical data, avoiding errors and costs for the bank.

V. References

[1] M. Crouhy, G. Dan and R. Mark, "The Essentials of Risk Management," McGraw-Hill, 2006, pp. 37-125.

[2] Basel Committee on Banking Supervision, Sound management of risks related to Money laundering and financing of terrorism, 2014, pp. 3-17.

[3] Banco Nacional de Angola, "Avisos, Directivas, Instrutivos e Lesgislação.," [Online]. Available: http://www.bna.ao/.

[4] Banco Nacional de Angola, Guia sobre a implementação de um programa de prevenção do branqueamento de capitais e do financiamento do terrorismo, 2013, pp. 3-35.

[5] Financial Action Task Force, "Publicações FATF.," [Online]. Available: http://www.fatf-gafi.org/.

[6] Nações Unidas, "United Nations - Office on Drugs and Crime," [Online]. Available: http://www.unodc.org.

[7] Egmont Group, "Egmont Documents.," [Online]. Available: http://www.egmontgroup.org/.

[8] Basel Committee on Banking Supervision, "Sound management of risks related to money laundering and financing of terrorism.," [Online]. Available: http://www.bis.org/bcbs/.

[9] The Wolfsberg Group, "Wolfsberg Standards.," [Online]. Available: http://www.wolfsbergprinciples.com/.

[10] Eastern and Southern Africa Anti-Money Laundering Group, [Online]. Available: http://www.esaamlg.org/.

[11] Banco Mundial, "Guia de Referência Anti-Branqueamento de Capitais e de Combate ao

Financiamento do Terrorismo," in *Segunda Edição e Suplemento sobre a Recomendação Especial IX*, 2004.

[12] G. S. Becker, "Crime and Punishment: An Economic Approach," *Journal of Political Economy,* Vols. 0-87014-263-1, pp. 1-54, 1974.

[13] D. Masciandro, "Money Laundering Regulation: The Microeconomics," *Journal of Money Laundering Control,* vol. 2, pp. 49-58, 1998.

[14] AS/NZS (2004), *Risk Management, AS/NZS 4360:2004, Standards Australia/Standards New Zealand..*

[15] ISO (2009), *Risk management - principles and guidelines on implementation, International standard, ISO 31000:2009,* Geneve: International Organisation for Standardisation.

[16] ISO (2009), *Risk management – Vocabulary, ISO/Guide 73:2009,* Geneve: International Organization for Standardization.

[17] ISO/IEC (2008), *Information technology — Security techniques — Information security risk management, ISO/IEC 27005:2008,* Geneve: International Organization for Standardization, International Electrotechnical Commission.

[18] T. Aven, Risk Analysis, Wiley, 2008.

[19] L. d. Pellegrina and D. Masciandaro, "The Risk-Based Approach in the New European Anti-Money Laundering Legislation: A Law and Economics View," *Review of Law & Economics,* vol. 5, pp. 931-952, 2008.

[20] D. Masciandaro, "Money Laundering: The Economics of Regulation," *European Journal of Law and Economics,* vol. 7, pp. 225-240, 1999.

[21] R. C. Watkins, K. M. Reynolds, R. F. DeMara, M. Georgiopoulos, A. J. Gonzalez and R. Eaglin, "Exploring Data Mining Technologies as Tools to Investigate Money Laundering," *Journal of Policing Practice and Research: An International Journal,* pp. 163 - 178, 2003.

[22] S. Kannan and K. Somasundaram, "Selection of optimal mining algorithm for outlier detection - An efficient method to predict/detect money laundering crime in finance industry," *Elysium Journal,* vol. 1, pp. 1-13, 2014.

[23] N. Le Khac, S. Markos, M. O'Neill, A. Brabazon and M. Kechadi, "An investigation into Data Mining approaches for Anti Money Laundering," in *International Conference on Computer Engineering and Applications*, Singapore, 2009.

[24] ACAMS, "The Third EU Directive – Customer due diligence and risk-based approach," 2010.

[25] I. Stancu and D. Rece, "The Relationship between Economic Growth and Money Laundering - a Linear Regression Model," *Theoretical and Applied Economics,* pp. 3-8, 2009.

[26] Breiman, L., Cutler, A., "Random forests," Berkeley, 2005. [Online]. Available: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.

[27] R. T. J. F. Trevor Hastir, "The Elements of Statistical Learning," in *Data Mining, Inference, and Prediction, Second Edition*, Springer, pp. 587-597.

[28] A. Ultsh and H. P. Siemon, "Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis," in *Proceedings of the International Neural Network Conference*, Paris, France, 1990.