RISK MANAGEMENT MAGAZINE

Vol. 16, Issue 3
September – December 2021

EXCERPT

https://www.aifirm.it/newsletter/progetto-editoriale/

Money laundering transaction detection with classification tree models

Paolo Giudici, Giulia Marini

Money laundering transaction detection with classification tree models

Paolo Giudici (University of Pavia), Giulia Marini (Intesa SanPaolo)

Article submitted to double-blind peer review, received on 9th September 2021 and accepted on 3rd November 2021

Abstract

The detection of money laundering is a very important problem, especially in the financial sector. We propose a mathematical specification of the problem in terms of a classification tree model that ”automates” expert based manual decisions. We operationally validate the model on a concrete application that originates from a large Italian bank. The application of the model to the data shows a good predictive accuracy and, even more importantly, the reduction of false positives, with respect to the ”manual” expert based activity. From an interpretational viewpoint, while some drivers of suspicious laundering activity are in line with the daily business practices of the bank’s anti money laundering operations, some others are new discoveries.

Keywords: Classification Trees, Automatic laundering detection, Predictive accuracy

1. Introduction

Money laundering embraces all those operations that disguise the illicit origin of capital flows, giving them a semblance of legitimacy, to facilitate the subsequent reinvestment in the ”legal” economy.
Money laundering is a very important problem for the society. It is estimated that 2-5% of global GDP is laundered annually, with an overall recovery rate of illicit assets at just 1.1% in Europe, according to Europol.
Given the central role of financial intermediaries such as banks, insurance companies and asset managers in the management of capital flows, authorities and regulators focus their surveillance on them. The detection of money laundering thus becomes of uttermost importance for financial intermediaries, whose daily activities should be monitored to avoid money laundering transactions.
To achieve this aim, financial intermediaries have established internal functions targeted towards the preventive measurement of anti money laundering risks, with the mission of marking as “suspicious” the transactions with the highest risk. The identification of a transaction as suspicious may have several consequences, from the the bank account of the originator of the transaction being frozen, to the opening of a legal case.
Against this background, a financial intermediary should develop a data driven statistical model to detect suspicious transactions that, while correctly identifying all possible laundering cases, avoiding a type I error, does not lead to an excessive number of susspicious cases, which may lead to a reputational loss for the intermediary (type II error).
The literature on the topic of money laundering stems from the context of fraud detection. For a general review see, for instance, the contributions of Bonini et al. (2019) and Chen et al. (2014) or, from a statistical viewpoint, Bolton and Hand (2002), Sudjianto et al. (2010) and, from a computer science viewpoint, Luo (2014) and Wang and Yang (2007).
Fraud detection methods are suited to model situations in which the cases of actual fraud are known, and the aim of statistical methods is to discover the most frequent patterns that have led to the fraudolent cases, in an unsupervised setting, for which the variable to predict is unknown.
Financial intermediaries, however, are not interested only in estimating the probability of a laundering fraud but also in minimising the probability of an unnecessary identification of a transaction as a suspicious one.
Our aim is to tackle this problem and provide a model to classify each financial transaction as being ”suspicious” or ”not suspicious”. From a statistical viewpoint, our response variable to predict is not the event ”laundering”, which may also be very rare, and may take long time before being ascertained; but, rather, the event ”suspicion of laundering”.
The latter event is measurable, as the money laundering function of the intermediary has, by law, to identify, every day, which are the ”suspicious” transactions.
Given the difference in magnitude between amount of suspicious transactions and the amount of actual laundering transactions, our proposed model can be a supervised learning model, rather than an unsupervised one, in which the variable to be predicted is the suspicious flag event.
We remark that classifying transactions as ”suspicious” or ”non- suspicious” with expert-provided training labels is an objective which is different from the actual money laundering detection in banking transactions. Although different, it is worth pursuing, as the activity of providing expert based labels of ”suspicion” is customarily done by banks of all size, in a labor-intensive labeling of training data which may not always be practical and that, for prudentiality reasons, typically over estimates actual money laundering cases. From an operational viewpoint, the over estimate can lead to a false positive rate of about 90% (as reported by an Italian bank) which implies a cost which is economic and reputational, as the account of a suspicious customer may be temporarily frozen.
Our proposed statistical method, which generates ”automatic” suspicion rules from the ”manual” expert based activity, can thus be built and optimised not only as a tool aimed at accurately predicting expert’s labels but also to reduce the false positives generated by the experts: cases detected as suspicious when they are not.
The only paper, to our knowledge, in line with our approach, is the recent paper by Jullum et al. (2020). Similarly to them, not only we propose a novel model to detect suspicious laundering activities, but we also validate it on on a concrete application, that originates from a large bank. While Jullum et al. (2020) considered data from DNB Bank, in Norway, we consider data from UBI Bank, in Italy.
Our added contribution to Jullum et al. (2020) is a model that distinguishes companies from individual customers, and provide different predictive rules for them. Furthermore, by using, as a response variable, the ”suspicion” that a certain transaction is a laundering, as declared by the responsible people within the bank, we contribute to ”automatise” human decisions by means of an

artificial intelligence method. We thus contribute to the literature on explainable AI models (see for example, Giudici and Raffinetti, 2020) by comparing human learning with automated learning.

2. Methods

The detection of money laundering suspicions can be embedded into the field of credit scoring models, common in the credit risk literature, following the seminal papers of Altman (1968) and of Merton (1974). For a review see e.g. Resti and Sironi (2007)

The most commonly used model for credit scoring applications is logistic regression, based on generalised linear models (see e.g. Mc- Cullagh and Nelder, 1989; Agresti, 1992), specified by a Bernoulli response random variable and a logistic link function, which relates a function of the expected value of the response variable to a linear combination of the available predictor variables. In our context, the event "suspicious laundering", as previosuly defined, can be represented by a Bernoulli random variable with a parameter that indicates the probability that a transaction of a given customer is suspicious.

Logistic regression models are based on a link function that corresponds to the log-odds of the probability and its complement, with respect to the explanatory variables. The predicted probability of money laundering can then be obtained converting the estimated log-odds regression coefficients using the logistic function.

To improve the predictive accuracy of scoring models, we employ a machine learning model and, specifically, a classification tree, instead of a logistic regression model, similarly to Jullum et al. (2020).

Differently from logistic regression, a tree model is non parametric and does not require the assumption of a distribution for the response variable. A tree model is essentially an algorithm which partitions the available data recursively in subgroups (determined by the values of the explanatory variables). The partitioning stops when a further split of the sample does not significantly improve the within group homogeneity. For a review on tree models, and a comparison with logistic regression, also on a credit scoring case, see Giudici (2003).

Machine learning models are typically better in predictive accuracy, with respect to logistic regression models, and our work on UBI Bank data confirm this intuition. On the other hand, machine learning models may be more difficult to explain, being usually "black boxes". Differently from Jullum et al. (2020) we compare alternative tree models, but select only one among them: the most accurate in predictions, instead of taking an ensemble model, such as a random forest, which may not be so transparent. We also compare our results with the experience of the money laundering professionals of the Bank, and this provides a very good practical validation of the usefulness of the model.

Most of our application work has focused on feature engineering and variable selection. We have first built, from the original database of all customers of the bank, two databases: that of companies and that of individuals. This because they have different features, corresponding to different degrees of accountability: higher for companies, lower for individuals.

For both databases we have then randomly split the sample into a "training" sample (80% of the data) and a "validation" sample (20% of the data). As the observed percentage of cases is low (around 3%) we have under sampled the not suspicious cases to obtain a balanced sample.

For both databases we have taken all the continuous available data (summarising aggregate monthly volumes of transactions) as they were. Differently, for the qualitative data (describing the "de- mographic" status of companies or of individuals) we have chosen to aggregate them in the grouping that is the most predictive of the target variable. We have then associated a binary dummy variable to each group.

We have then compared alternative tree models, calculating for each of them, on the same test set, the Area Under the ROC Curve as a measure of predictive accuracy, as well as the False Positive and True Negative rates. When the AUROC were similar, we chose the model with the lowest False Positive rate and, possibly, the lowest complexity.

3. Application

The main problem of the Bank is to identify in advance customers

with suspicious operations, to be reported for suspected money laundering activities. The identification should target the highest number of fraudolent cases but keeping unnecessary reports to a minimum. To this aim we build a predictive statistical model which, based on the reports submitted in the past (taken as a response variable) and on customers' transactions (taken as explanatory variables) recognizes automatically suspected customers.

The considered data consists of all transactions automatically alerted by the Information Systems of the bank in the years 2019-2020. Each transaction has been mapped to a specific customer ID. For the same customer we collected information on whether the Anti Money laundering function of the Bank have flagged the customer as "suspicious" or not during the year 2020. For the response variable, we consider only the year 2020, as the decision to flag a customer as suspicious is usually taken looking at its transactions in the past year.

The total number of considered customer alerts is equal to 96624 among which 723 have been flagged as "suspicious". About 53% of the customers are companies and the remaining 47% individuals. The bank transactions of each customer have been classified according to the common standards in use into a series of binary variables. Besides transaction variables, each customer has "demographic" variables, such as the economic sector of activity, and the legal nature, for the companies; the market segment, profession and age, for individuals. The total number of considered variables are: 46 for individuals and 184 for companies.

Once data has been prepared for the application of classification tree models, it has been split into a 80% training set, on which to build models, and a 20% validation set, on which to calculate model predictions, and select the best model, comparing the

predictions with the actual values, in the ROC curve model comparison setting. Precisely, we have selected the tree model with the highest value of the AUROC curve, as suggested by Hand, Mannilla and Smyth (2001). In case of ”similar” values of the AUROC, we have chosen the model with the lowest value of the false positive rate, as one of our main objectives is to reduce the over identification of cases. When also false positive rates were similar, we have chosen the model with the most parsimonious model, for the sake of explainability and interpretability.

For the companies, the best selected model contains 10 final classes (leaves of the tree), with an AUROC equal to 0.81. and a false positive rate of 0.43. A model with 15 classes has a higher AUROC of 0.84 and a better false positive rate of 0.35: however, the management of the Bank has decided that the increased complexity and cost of its possible implementation exceeds its limited accuracy advantage.

The selected model is described in Figure 1. The relative impor- tance of the variables that have been selected is described in Table 1. The importance of a variable is defined by the reduction in the residual sum of squares reduction determined by the variable splits in the tree.

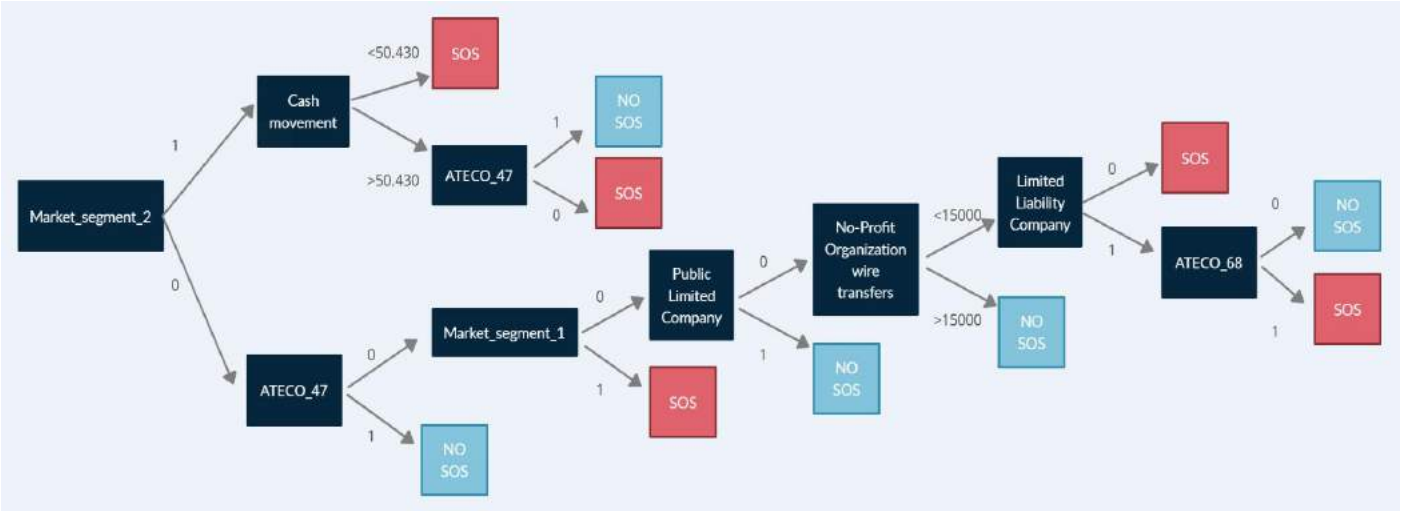

Figure 1: Final tree model - Companies.

ID	Variable	Relative importance
1	Medium size companies	1
2	Small size companies	0.93
3	Retail sector (excluding cars and motorcycles)	0.85
4	Limited liability companies	0.75
5	Real estate business sector	0.63
6	Non profit organisations	0.61
7	Public limited companies	0.53
8	Cash movements	0.52

Table 1: Final tree model - Companies.

From Figure 1 and Table 1 note that the most important variables that determine whether a company performs suspicious operations, in terms of money laundering, have mainly to do with the size of the company, its sector of activity, and its legal form: all ”demographic” variables. The only relevant transaction variable is the amount of cash movements. While the result on cash is in line with the daily practice and experience of the bank, is was somewhat unexpected that ”cash” is the only transaction amount that matters. The findings on the important demographic variables are instead new discoveries, not used in routinely applications. This concerns in particular the relevance of the retail sector: a driver of suspicious activity which is likely related to the type of customers of the bank, concentrated in an Italian region specialised in retail manufacturing. Another unexpected finding concerns

the high risk attributed to non profit organisations: again, there is a large presence of these institutions in the regions where the bank operates.

For the individuals, the best selected model contains 15 final classes (leaves of the tree), with an AUROC equal to 0.77, slighlty inferior than for the companies, and with a false positive rate of 0.70, rather higher.

A model with 20 classes has a lower AUROC of 0.69 and a better false positive rate of 0.42: the management of the bank has decided that the increased complexity and cost of its possible implementation exceeds its limited accuracy advantage.

The chosen model is in line with the fact that the ”demographic” information on the companies are more standardised. The selected model is described in Figure 2. The relative importance of the vari- ables that have been selected is described in Table 2. The importance of a variable is defined by the reduction in the residual sum of squares reduction determined by the variable splits in the tree.

From Figure 2 and Table 2 note that the most important variables that determine whether an individual performs suspicious operations, in terms of money laundering, have to do with the market segment of the individual, its profession, and its age: “demographic” variables. In addition, three transaction variables are important: the relevant amounts (in cash); the account balance (in a month) and the geolocalised wire transfers (for example to ”offshore” destination).

The results on the transactions are in line with the experience and every day practice of the bank professionals, and in particular those that concern relevant cash amounts and the wire transfers. The account balance is somewhat unexpected as the current practice considers the total of positive and negative amounts, separated, rather then their difference.

On the other hand, the results concerning the ”demographic” status variables are more unexpected. In particular, the importance of the Market segment dummy variables: lower mass, top private, upper mass; indicate that all those categories are at risk, and only the ”medium” segments have a low risk. The results also indicate the need to work on the individuals for which the segment is ”Not available” . The results about the high risk from entrepreneurs and senior people are more in line with the intuition. Last, the fact that the ”consumers” variable is relevant depends on its importance, particularly for the bank.

We finally remark that our results, for both companies and indi- viduals, have been thoroughly discussed with the management of the bank, finding their agreement on the above conclusions.

ID	Variable	Relative importance
1	Account balance	1
2	Relevant amounts	0.677
3	Geo localised wire transfers	0.631
4	Lower mass market segment	0.405
5	Consumers	0.403
6	Entrepreneurs	0.377
7	Top private	0.309
8	Age greater than 65	0.257
9	Market segment NA	0.243
10	Upper mass market	0.221 height

Table 2: Final tree models - Individuals

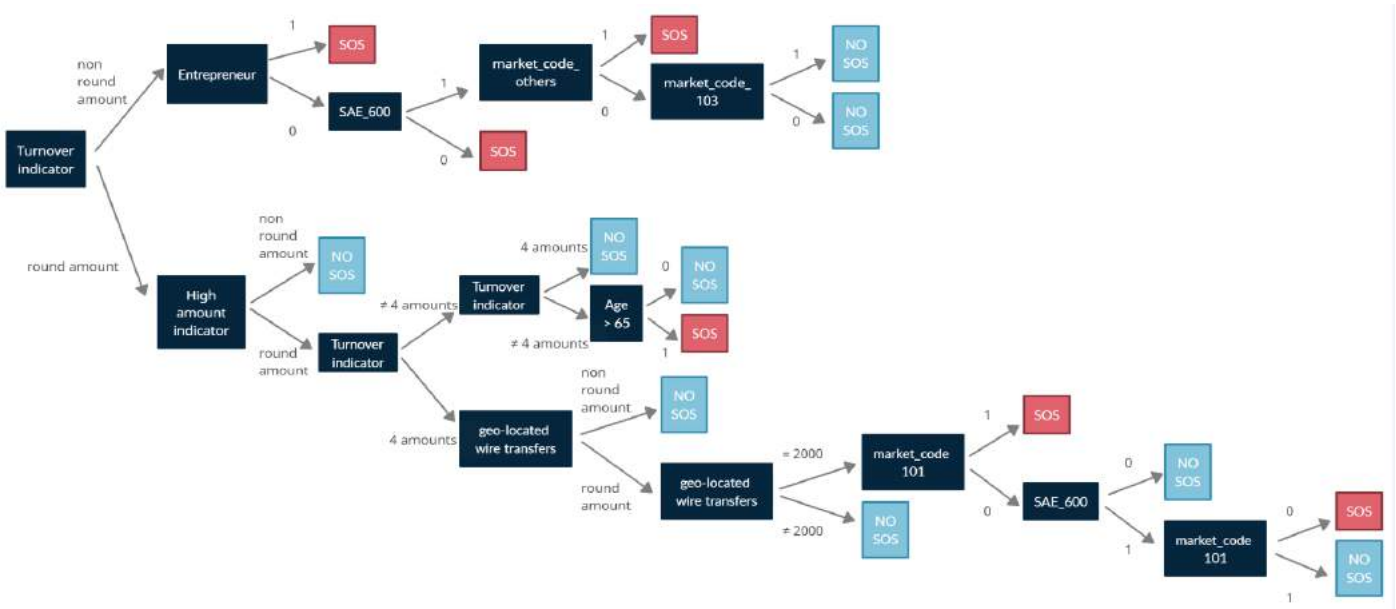

Figure 2: Final tree models - Individuals

4. Conclusions

The work in this paper shows how the "manual" activity of flagging bank transactions as suspicious money laundering activities, under constant scrutiny by supervisors, can be automatized with a save in cost and predictive accuracy. The analysis was applied on a single bank, albeit large, and it would be extremely interesting to apply to other banks, to reinforce the previous conclusions.

The application of artificial intelligence techniques reveals, once more, its potential and, in particular, the gain in predictive accuracy. This gain, however, should be balanced against negative aspects such as a more limited explainability of the results, and a possible lower robustness with respect to classical regression models.

Acknowledgements

The Authors thank Valerio Pisoni, Paolo Francesco Griffo and Roberta Vitali from the Anti Financial Crime Measures and Analytics function of UBI Banca, for the computational work behind the paper.

References

1. Agresti A. (2002). Generalised linear models. Wiley, New York.
2. Altman E. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of finance, 23 (4), 589-609.
3. Bolton, R.J. and Hand, D.J. (2002), "Statistical fraud detection: a review (with discussion)", Statistical Science, Vol. 17, pp. 235-249.
4. Bonini, S., Caivano, G., Cerchiello, P., Giribone, P.G. (2019). L'applicazione di machine learning e predictive analytics nel risk management. AIFIRM position paper n.14, 2019
5. Chen, Z., Van Khoa, L.D., Nazir A., Teoh, E.N., Karupiah, E.K. (2014). Exploration of the effectiveness of Expectation maximization algorithm for suspicious transaction detection in anti-money laundering. ICOS 2014–2014 IEEE conference on open systems, pp 145–149.
6. Giudici, P. (2003). Applied data mining. Wiley, 2003.
7. Giudici P., Raffinetti, E. (2021). Shapley-Lorenz explainable AI models, Expert Systems with applications, vol. 167
8. Hand D.J., Mannila H., Smyth P. (2001) Principles of Data Mining, MIT Press.
9. Jullum, M, Loland, A., Huseby, R.B., Anonsen, G., Lorentz, J. (2020) Detecting money laundering transactions with machine learning. Journal of money laundering control, volume 23 (1),
10. Luo, X. (2014). Suspicious transaction detection for anti-money laundering. International Journal of Security and its Applications 8(2):157–166
11. McCullagh, P. and Nelder, J. (1989). Generalised linear models.
12. Chapman and Hall, New York.
13. Merton R.C. (1974). On the pricing of corporate debt: the risk sructure of interest rates. Journal of Finance 2, pp. 449-471.
14. Resti A., Sironi A. (2007) Risk management and shareholders' value in banking. Wiley.
15. Sudjianto, A., Nair, S., Yuan, M., Zhang, A., Kern, D. and Cela Dıaz, F. (2010), Statistical methods for fighting financial crimes, Technometrics, Vol. 52 No. 1, pp. 5-19.
16. Wang, S.N., Yang, N.G. (2007). A money laundering risk evaluation method based on decision tree. In: Machine learning and cybernetics, Hong Kong.