Bankrobotics: Artificial Intelligence and Machine Learning Powered Banking Risk Management

Prevention of Money Laundering and Terrorist Financing

Alexandra Prisznyák
University of Pécs
alexandra.prisznyak@gmail.com

Summary

Based on a country study related to money laundering and terrorist financing, the Financial Action Group downgraded Hungary's compliance with Recommendation R15 (use of new technologies). At the same time, between 2020 and 2021, the Magyar Nemzeti Bank imposed fines on several commercial banks operating in Hungary for shortcomings on complying with money laundering and terrorist financing regulations. As a gap-filling analysis, the study examines supervised (classification, regression), unsupervised (clustering, anomaly detection), and hybrid machine learning models and algorithms operating based on highly unbalanced dataset of anti-money laundering and terrorist financing prevention of banking risk management. The author emphasizes that there is no one ideal algorithm. The choice between machine learning algorithm is highly determined based on the underlying theoretical logic and additional comparative. Model building requires a hybrid perspective of the give business unit, IT and visionary management.

Keywords: Artificial Intelligence, Machine Learning algorithms, banking risk management, Anti-Money Laundering and Counter Financing Terrorism, supervised/unsupervised methods

JEL codes: C45, C80, G21, G32, O33

DOI: https://doi.org/10.35551/PFQ_2022_2_8

With the famous question *'Can machines think?'* in his paper titled 'Computing Machinery and Intelligence', Turing kicked off the development of artificial intelligence and the related technologies in 1950. Artificial Intelligence (AI), as the European Commission (2018) defined it *'refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals'*. As an umbrella term, AI covers the concept of Machine Learning (ML), too (ECB, 2020). ISO/IEC 38505-1:2017 defines machine learning as a process that facilitates the making of predictions for the future, based on existing data and by using algorithms. ML can be defined as a subset of AI that is able to perform pre-programmed tasks on the basis of large volumes of data (hereinafter as: Big Data) and through the learning process of software programs and algorithms (EBA, 2020; European Parliament, 2020).

In relation to the application of artificial intelligence in banking areas, FSB (2017) points out the following:

- *front office:* sales and trading support solutions (strong customer identification, chatbots, contracting, market trading, modelling, impact analysis, other),
- *middle office:* credit analysis, scoring activity, customer rating, customer profile building, other,
- *back office:* risk management activities of banks (stress tests, prevention of money laundering and terrorism financing, compliance, anomaly detection, model validation, other).

Szikora and Nagy (2020) finds that the application of AI in banking processes in most cases covers customer rating, credit assessment, customised financial services, detection/prevention of fraud and corrupt practices, due diligence of contracts, as well as legal due diligence.

The author's primary objective is to review machine learning methods, techniques and algorithms used in the area of digitalization to prevent money laundering and terrorism financing in banks, and to draw the conclusions related to their application and comparison.

SITUATION IN HUNGARY

With digital technologies gaining ground during the Covid-19 pandemic and because of the continuously increasing number of financial crimes in the digital era, the spread of artificial intelligence, machine learning and related technological solutions – such as Advanced Analytics (AA, hereinafter: advanced analytic instruments) – presented a challenge to the banking sector. Consequently, artificial intelligence and machine learning appeared in the fields of Anti-Money Laundering (AML), Counter Financing Terrorism (CFT), fraud prevention and compliance (with legislation), too, and it is developing at an increasing speed (Van Wegberg, Oerlemans, Van Deventer, 2018; Johari et al., 2020).

The prevention of the use of financial systems for the purposes of money laundering and terrorism financing is treated as an issue of high priority by several international organisations, such as the UN, the Financial Action Task Force (hereinafter: FATF), the European Union, the Council of Europe and its expert committee, the MONEYVAL, and a number of international organisations – including the IMF, the World Bank and the Basel Committee on Banking Supervision. Regarding the prevention of money laundering and terrorism financing, Directive EU 2015/849 can be considered as a benchmark, and it covers the Customer Due Diligence (CDD) examinations, including the KYI (Know Your Intermediary!) and the KYC

(Know Your Customer, hereinafter: Know Your Customer!) policies (BIS, 2001). In order to ensure compliance with legislation, Article 22 of Commission Delegated Regulation (EU) 2017/565 supplementing Directive 2014/65/EU prescribes the operation of a customised and independent compliance function that corresponds to the organisational structure and the extended service, and the operation of a risk-based monitoring system. The referenced EU Directives are supplemented with several other legal regulations, directives and recommendations at international and national levels. The transposition of EU-level and international legislation (directives, regulations, recommendations) into Hungarian legislation is implemented with the amendment of Act CXXXVI of 2007 on the Prevention and Combating of Money Laundering and Terrorist Financing and other related legislation. Implementation in Hungary is supported by other legislation in force and the recommendations of the Magyar Nemzeti Bank (National Bank of Hungary, Hungarian acronym: MNB). Hungarian regulations in force against money laundering can be found in the Criminal Code and in Act LIII of 2017 on the Prevention and Combating of Money Laundering and Terrorist Financing.

The Paris-based Financial Action Task Force (hereinafter: FATF) founded in 1989 as an intergovernmental organisation assists the anti-money laundering and counter-terrorism financing activities of 39 countries. Hungary is loosely connected to it through its membership in the European Commission. For the prevention of money laundering and terrorism financing the objective of FATF is to establish international cooperation, work out recommendations according to its risk-based approach and to implement these at global level. Compliance with FATF recommendations (R40, +9 special recommendations) is regularly evaluated for each FATF member (including Hungary) (FATF, 2012, 2021, 2021b).

In line with the FATF recommendations, Hungary has successfully improved the related regulatory environment (FATF, 2021a; Tóth, 2018) over the past decade. Based on the country report also accepted by FATF, Hungary achieved an improvement in the period of 2019-2021 regarding the examined recommendations (R40), which is presented in *Table 1* (FATF, 2021a).

FATF recommendations are rated as follows: (C) compliant, (LC) largely compliant, (PC) partly compliant, (NC) non-compliant. As far as the recommendations are concerned, the ultimate goal is to improve recommendations of PC rating to at least LC/C ratings, otherwise the plenary session proposes a Compliance Enhancing Procedure (CEP).

On the basis of the country rating of 2019, the following changes occurred in 2021:

- the following recommendations are considered compliant: R4, R9, R20, R29, R30;
- the following recommendations are considered largely compliant: R1, R2, R3, R5, R6, R7, R10, R11, R12, R14, R16, R17, R19, R21, R22, R23, R25, R26, R27, R28, R31, R33, R34, R35, R36, R37, R38, R39, R40;
- area to be improved: R13, R15 (use of new technologies), R18, R24, R32.

The author points out that in relation to the application of artificial intelligence and the related technologies, the downgrading (from C to PC) of Recommendation 15, 'New technologies' (R15) supporting the prevention of money laundering and terrorism financing is an important change compared to 2019. The objective of Recommendation 15 is to encourage the implementation of new technologies in banks to prepare them for the new types of money laundering and terrorism financing methods of the digital era, increasing

Table 1

HUNGARY'S COMPLIANCE WITH FATF RECOMMENDATIONS IN 2019–2021

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
2019	LC	LC	LC	C	LC	LC	LC	PC	C	LC
2021	LC	LC	LC	C	LC	LC	LC	PC	C	LC
	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
2019	LC	PC	PC	LC	C	LC	LC	PC	LC	C
2021	LC	**LC**	PC	LC	**PC**	LC	LC	PC	LC	C
	R21	R22	R23	R24	R25	R26	R27	R28	R29	R30
2019	LC	LC	LC	PC	LC	LC	LC	LC	C	C
2021	LC	LC	LC	PC	LC	LC	LC	LC	C	C
	R31	R32	R33	R34	R35	R36	R37	R38	R39	R40
2019	LC	PC	LC	LC	LC	LC	LC	LC	LC	LC
2021	LC	PC	LC	LC	LC	LC	LC	LC	LC	LC

Source: own edition, based on FATF (2021a)

thus the efficiency of the risk management activities of financial institutions. As we can see, in the respect of recommendations for the use of new technologies, Hungary's performance is lower than expected.

In 2020 and 2021, in the auditing of the efficiency of anti-money laundering controls required for activities involving significant cash turnover (risk identification, management, process regulation, internal audit), the Magyar Nemzeti Bank imposed high fines on commercial banks registered in Hungary if deficiencies were identified *(Figure 1)*.

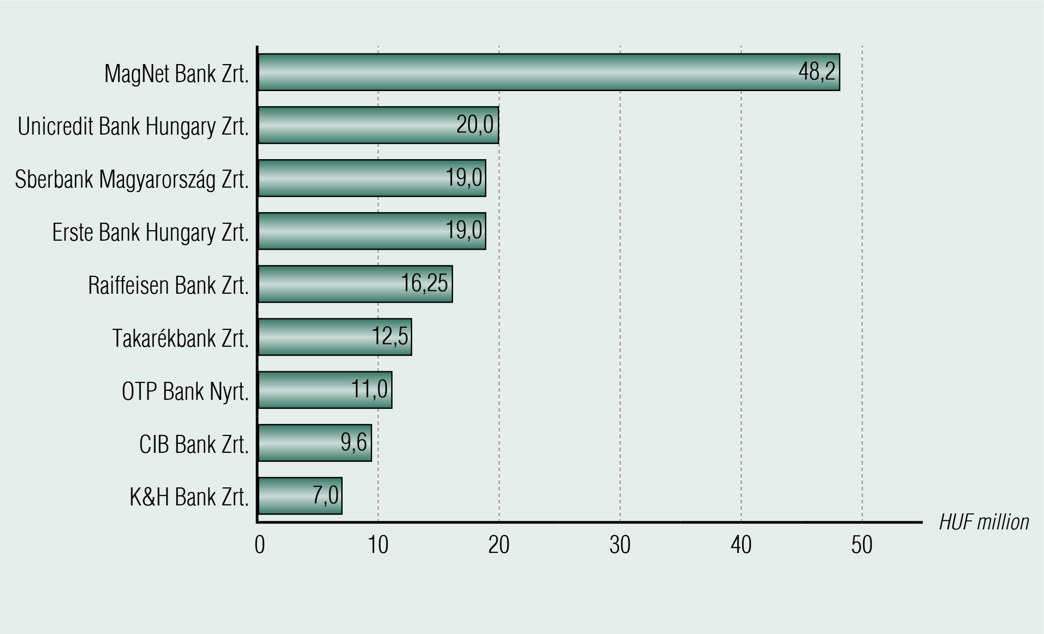
A high number of underlying reasons can be traced back to errors found in process organisation, information analysis and reporting obligations, which errors could have been significantly mitigated by an interaction between human and artificial intelligence.

GENERAL PROCESS OF AML AND CFT EXAMINATION AND THE POSSIBILITIES OF MACHINE LEARNING

In order to avoid the management of anonymous accounts and accounts opened under fictitious names, financial institutions operate systems to check the identity of customers when the business relation is established and when individual transactions are initiated. The alert system performing CDD identifications of various extents and frequencies per risk category supports the verification of the identity of customers on the basis of external and internal databases, other submitted documents and independent source documents. In line with the FATF recommendations the high volume of customer data usually stores information that allows for

Figure 1

EXTENT OF MNB SUPERVISORY FINES FOR VIOLATION OF MONEY LAUNDERING PREVENTION LEGISTLATION IMPOSED ON COMMERCIAL BANKS OPERATING IN HUNGARY IN 2020–2021



Source: own edition based on www.mnb.hu/sajtoszoba/sajtokozlemenyek

the reliable and continuous identification of customers (CDD) and the examination of the performed transactions.

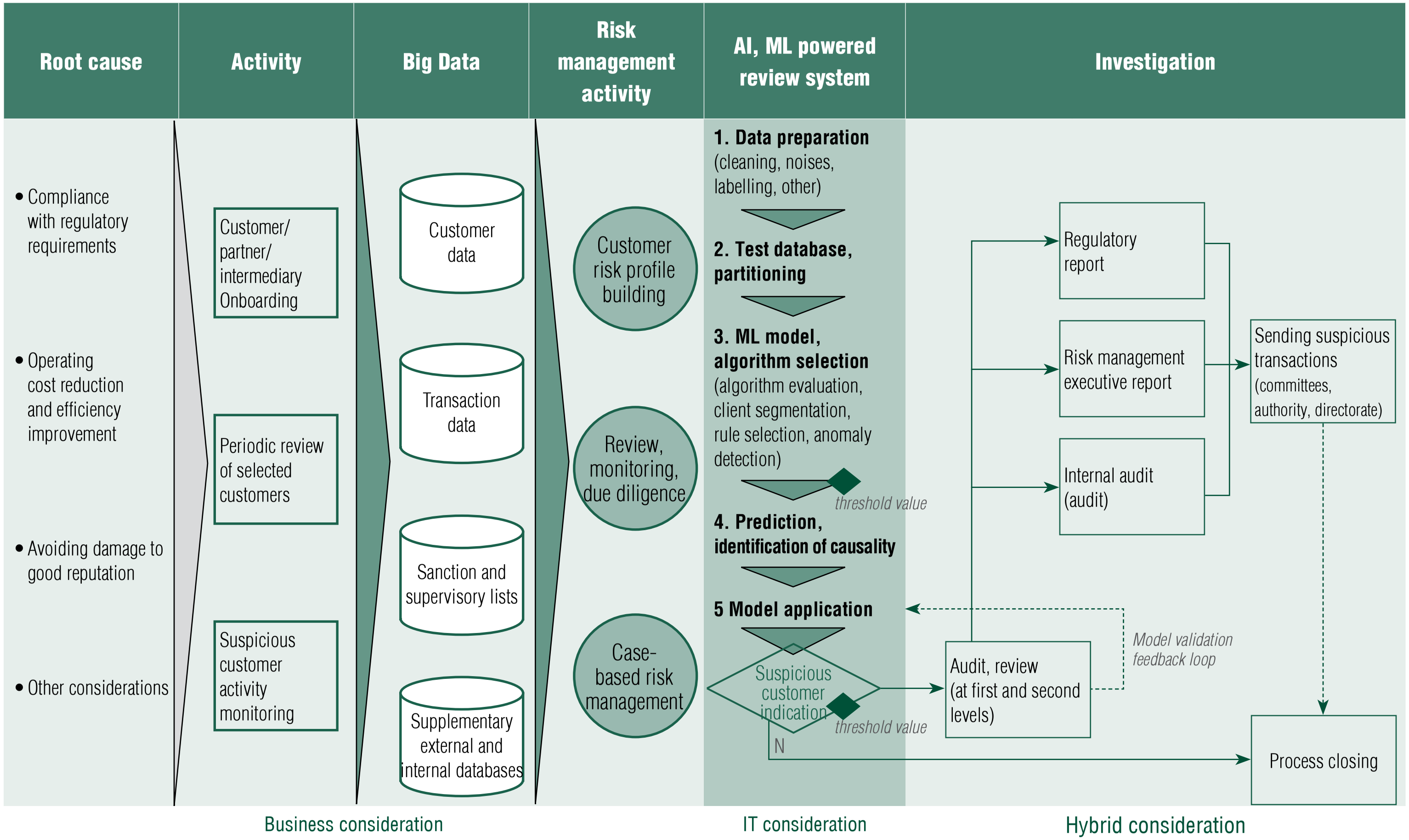
Alarm models recognising patterns indicating money laundering activities are able to associate a probability of the occurrence of money laundering through machine learning and by using customer data, transaction and other background information. If the system labels a given event as 'suspicious', it is delegated to a higher expert level for further examination in the process (*Figure 2*). If the further expert analysis determines that the given case is a suspicious event that should be forwarded to the competent authority, the case is reported so that reporting and other compliance requirements are met (Jullum, Løland, Huseby, 2020).

The scope of applied attributes (information on the regulatory side, open external sources, expert knowledge, history data of financial crimes) may be extended with other additional information on the basis of the professional position of the model builder, in order to improve the model (Rocha-Salazar, Segovia-Vargas, Camacho-Miñano, 2021; Rouhollahi, 2021; Chen et al., 2018). Typical attributes used for machine learning model building in the area of anti- money laundering area are illustrated in *Table 2.*

The proper selection of the scope of attributes and the removal of redundant and irrelevant attributes form an important phase of ML model building. The removal of irrelevant data may improve the learning accuracy of the algorithm, may reduce the time required for calculation, and may contribute to the more accurate understanding of relations.

Figure 2

MACHINE LEARNING BASED MONEY LAUNDERING AND TERRORIST FINANCING (GENERAL) CONTROL PROCESS


Root cause
Activity
Big Data
Risk management activity
AI, ML powered review system
Investigation
• Compliance with regulatory requirements
• Operating cost reduction and efficiency improvement
• Avoiding damage to good reputation
• Other considerations
Customer/ partner/ intermediary Onboarding
Periodic review of selected customers
Suspicious customer activity monitoring
Customer data
Transaction data
Sanction and supervisory lists
Supplementary external and internal databases
Customer risk profile building
Review, monitoring, due diligence
Case-based risk management
1. Data preparation (cleaning, noises, labelling, other)
2. Test database, partitioning
3. ML model, algorithm selection (algorithm evaluation, client segmentation, rule selection, anomaly detection)
threshold value
4. Prediction, identification of causality
5 Model application
Suspicious customer indication
threshold value
N
Audit, review (at first and second levels)
Model validation feedback loop
Regulatory report
Risk management executive report
Internal audit (audit)
Sending suspicious transactions (committees, authority, directorate)
Process closing
Business consideration
IT consideration
Hybrid consideration

Source: author's own figure

Table 2

TYPICAL ATTRIBUTES USED FOR ML MODEL BUILDING IN AML AND CFT

Attribute category	Változó
Customer-related attributes	Customer type (legal entity, private person, other) Customer segment, Politically exposed person (PEP) Age, Nationality, Source of incoming funds (income), Used product type, Economic activity, Time elapsed since the customer's entry Business data (ownership structure, shares)
Transaction-related attributes	Transaction type, Names of customers involved in the transaction (sending, receiving party) Transaction frequency, Transaction date and time, Transaction amount, Currency, Average amount, Target bank, Transaction code, Branch (customer) type, Transaction statement
Other attributes	Product/service type Geographical area (exposure) Representative of legal entity
Networking	Interconnected business relation network of clients

Source: the author's own compilation on the basis of the literature processed

OPTIMISATION PROBLEM: ALGORITHM SELECTION AND COMPARISON

In the case of the traditional analysing techniques applied for the prevention and detection of money laundering and terrorism financing, financial institutions face a number of problems: high IT costs, resource-intensive analysis techniques, high ratio of false positive hits, inflexibility and lack of dynamism in the rules identifying criminals' behaviour patterns (Rocha-Salazar, Segovia-Vargas, Camacho-Miñano, 2021). The underlying reasons for the development seen in traditional methods include the improvement in computer performance, the spread of the application of artificial intelligence and data mining, the development of artificial intelligence and related technologies, and the tightening of the relevant legislation (Watkins et al., 2003).

Through the supervised and unsupervised reinforcement learning methods, machine learning provides an efficient analysing solution, which in addition to the structured data processed by the traditional relational databases (document-oriented/NoSQL databases), further enhances the scope of data available for the analysis. The key question in selecting the algorithm to be used in machine learning model building is which algorithm are able to use the values of individual variables for correctly estimating the values of other variables.

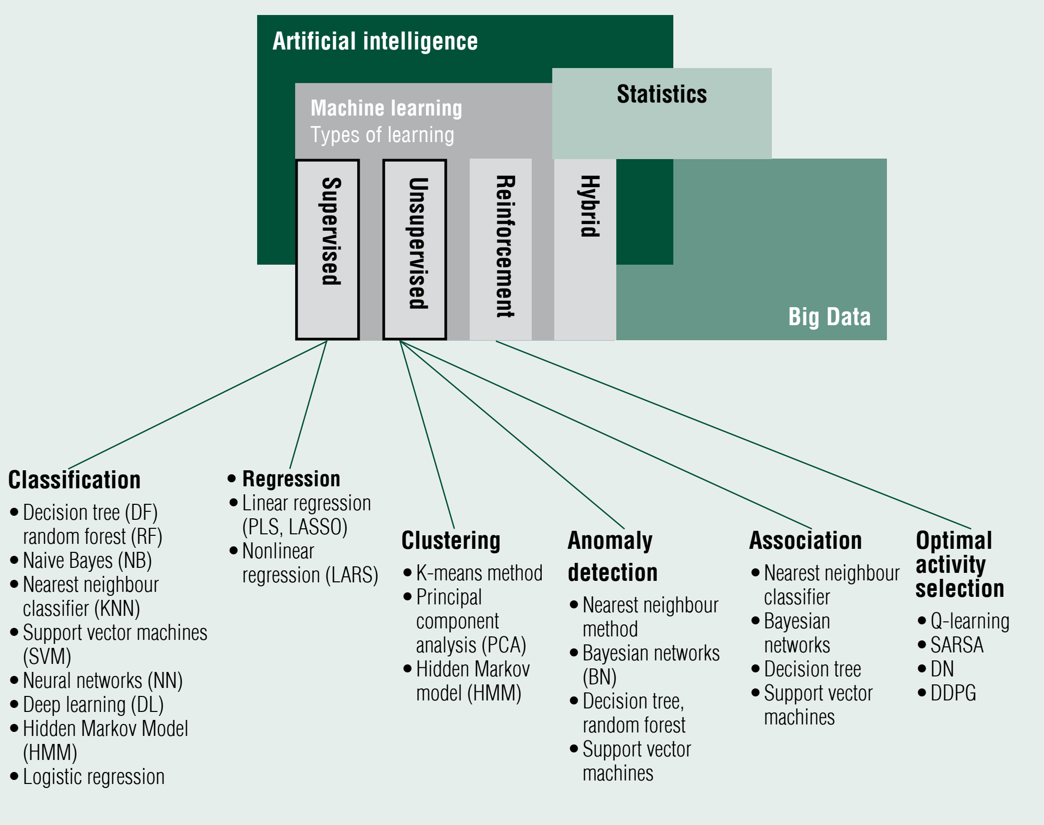
The grouping of machine learning algorithms belonging to the umbrella term of artificial intelligence is illustrated by *Figure 3.*

The size of the available datafile is important when the model is learnt, and the method of learning should be selected on the basis of that and the scope of possible algorithms should be selected accordingly (Savage et al., 2016; Zhang, Trubey, 2019; Chen et al., 2018).

In the case of supervised machine learning the algorithm learns through a teaching dataset that has labels (for instance, based on confirmed past cases of money laundering and terrorist financing). (1) classification and

Figure 3

AI UMBRELLA TERM, ML, TEACHING METHODS AND ALGORITHMS


Artificial intelligence
Machine learning
Types of learning
Statistics
Supervised
Unsupervised
Reinforcement
Hybrid
Big Data
Classification
• Decision tree (DF) random forest (RF)
• Naive Bayes (NB)
• Nearest neighbour classifier (KNN)
• Support vector machines (SVM)
• Neural networks (NN)
• Deep learning (DL)
• Hidden Markov Model (HMM)
• Logistic regression
• Regression
• Linear regression (PLS, LASSO)
• Nonlinear regression (LARS)
Clustering
• K-means method
• Principal component analysis (PCA)
• Hidden Markov model (HMM)
Anomaly detection
• Nearest neighbour method
• Bayesian networks (BN)
• Decision tree, random forest
• Support vector machines
Association
• Nearest neighbour classifier
• Bayesian networks
• Decision tree
• Support vector machines
Optimal activity selection
• Q-learning
• SARSA
• DN
• DDPG

Source: own figure, based on EBA (2020) report

)

(2) regression must be classified as supervised learning methods. The model (algorithm) taught on the basis of patterns taken from the used database is able to accurately predict the classification of a new, yet unknown object (input, *X*) (output, *Y*), or the features and characteristics it will have. (*Y*) = *f (X).*

In the case of supervised learning suspicious transactions are typically labelled by experts or with labelling methods (e.g. Snorkel model).

Labelling example based on certain transaction types, transaction amounts, target countries:

ase when Transaction_Type = 'Payment transfer' and Transaction_Amount > x then 'True'

when Country_Code in ('Ethiopia', 'Kenya') and Transaction_Amount > x then 'True' else 'False' and as Suspicious

where the countries listed were selected from the FATF blacklist, and *x* is the limit value of the means of payment denominated in the given currency, defined from risk threshold aspect.

In the case of unsupervised learning the algorithm automatically sets up classes by creating relations, association links and decision strategies not known in advance, based on the relations and patterns recognised in the database. Unlike in supervised learning, the criteria, based on which the given objects of the sample database are grouped, are not specified in advance. In other words, the objects are not labelled in advance, they are labelled by the algorithm on the basis of the patterns detected (hidden) in the database. Typical unsupervised methods: (1) cluster analysis, (2) anomaly detection. Following the specification of the given business problem the user may make a decision for the selection of the algorithm on the basis of the following: range of available data (data volume), the dataset's structure, the ratio of outstanding data (anomalies) within the dataset population (number of cases). After the building of the model the decision may furthermore be supported by evaluation tools. The various machine learning algorithms require different input data formats and attribute selections. The different ML processes induce the fact that the learning process and the data usage of the given algorithm are different for the solution of the given problem *(see Tables 3, 4, 5).*

Based on *Tables 3, 4 and 5*, we can say that the typical algorithms used in the classification method are as follows: logistic regression, nearest neighbour method, (artificial) neural networks, Naive Bayes, decision tree, (and its variations: XGBoost, pGBRT, FP-growth, random forest model), SVM, Bayes logistic regression, Bayesian network. While in the case of regression procedures: the Maximum Likelihood logistic regression was applied on the examined sample from literature. In the cluster analysis, the application of the following algorithms can be observed: K-means method, neural networks, Neural Gas, SOM algorithm. In the course of anomaly detection, the iForest algorithm was used in the examined sources.

Tables 3, 4 and 5 indicate that the performance assessment of ML models is usually realised on the basis of a performance indicator or through expert validation (performance indicator column).

Considering the fact that the algorithms and results indicated in tables 3, 4 and 5, the applied databases and model building procedures (data cleaning, parametrisation, other) are different, it is not possible to compare the results of the algorithms shown in the tables, because that would assume the possibility of running on the same database. However, they give us a picture of the wide spectrum of solutions used

Table 3

SUPERVISED MACHINE LEARNING METHODS AND ALGORITHMS

Author(s), year	ML algorithms	Applied database	Performance indicator	Result
Zhang and Trubey (2019)	SVM, ANN, Maximum Likelihood/ Bayesian logistic regression, decision tree, random forest	Real data of American financial institutes	AUC, ROC curve	Best performance: ANN algorithm. SVM and RF have a better performance than logistic regression. SVM performs well in the classification of linearly non-separable groups.
Wang, and Yang (2007)	Decision tree	Customer transaction data	False positive ratio	The decision tree has limited efficiency (it does not identify each suspicious transaction properly).
Chen and Guestrin (2016)	XGBoost, pGBRT	Contains 4 public data sets	Runtime (execution time), accuracy	The runtime and the resource requirement of XGBoost are lower than those of pGBRT, and offer satisfactory performance.
Wei et al., 2012	ContrastMiner	Database of Australian bank	Alarm and detection ratio	In the case of large and unbalanced data files, the ContractMiner algorithm significantly improves accuracy.
Khan et al (2013)	Bayesian networks	Real financial institution dataset	False positive ratio.	'Suspicious' mark related to customer behaviour pattern for higher level of professional examination.
Kannan and Srinath (2017)	TBOD and AROMLD algorithm	Real bank database	Sensitivity, accuracy, specificity, runtime (execution time)	TBOD is more accurate, however, its calculation complexity and execution time reduce the total performance. As an alternative, the runtime of AROMLD is lower.
Jullum, Løland, Huseby (2020)	XGBoost	Anonymous banking data	Brier-score, AUC curve, PPP	The application of XGBoost results in increased performance (AUC, PPP, TPR)
Patil, Dharwadkar (2017)	Artificial neural networks	Publicly available German borrower database	Average square root difference, accuracy	The model has good results in rating accuracy.
Álvarez et al. (2017)	Logistic regression, decision trees, neural network, random forest model	Database containing real transaction data	Accuracy	The elimination of the noise in the data file improves the performance of the classifying algorithm. RF ensures a higher classifying performance than others.
Savage et al. (2016)	K-nearest neighbour algorithm, SVM, RF	Australian Transaction Report and transaction data reported to the Analysis Centre	ROC curve, accuracy, FTP rate	FTP has slightly better performance than SVM. Relation analysis improves the efficiency of the system.
Deng et al. (2012)	Naive- Bayes, SVM	Real transaction data from financial institution	Accuracy, expert validation	The application of the sequential active learning method provided a higher performance than that of the Naive Bayes and the SVM models.
Luo (2014)	FP-growth algorithm	Large-size generated transaction data file	Accuracy	The performance of the model improves with the increase in the number of transactions.
Luo (2014)	FP-growth algorithm	Large-size generated transaction data file	Accuracy	The performance of the model improves with the increase in the number of transactions.

Source: own edition

Table 4

: UNSUPERVISED MACHINE LEARNING METHODS AND ALGORITHMS

Author(s), year	Applied algorithms	Applied database	Performance indicator	Result
Alexandre, Balsa (2018)	*K*-means algorithm	Based on financial institution's real database	ROC curve and banking professionals (validation)	In the rule generation, the J48, JPART algorithm offered the best result. However, the rate of precision is less than expected.
Khac et al. (2010)	Neural networks	Transaction data set of BEP Bank related to investment funds	banking expert validation	In the case of the model, the selection of the parameters is important for the performance and runtime (execution time) of the model.
Drezewski, Sepielak, Filipkowski (2012)	FP-Growth, FPClose, FPMax, Sequence Miner, BIDE, BIDEMax	Database containing bank account statements	execution time	Cluster analysis and the created clusters can be successfully used to detect money laundering.
Rocha-Salazar, Segovia-Vargas, Camacho- Miñano (2021)	*K*-means, Neural Gas, Strict, SOM algorithm	Database of Mexican financial institutions	Calinski-Harabasz-index; abnormality indicator, accuracy?!, ERR, ACC	The integration of additional non-transactional variables into the model improves the accuracy of the predictions and reduces human resources requirements and costs.
Rocha-Salazar, Segovia-Vargas, Camacho- Miñano (2021)	*K*-means, Neural Gas, Strict, SOM algorithm	Database of Mexican financial institutions	Calinski-Harabasz-index; abnormality indicator, accuracy?!, ERR, ACC	The integration of additional non-transactional variables into the model improves the accuracy of the predictions and reduces human resources requirements and costs.

Source: own edition

Table 5

HYBRID MODELS

Author(s), year	ML algorithms	Applied database	Performance indicator	Result
Rouhollahi (2021)	Classification algorithms: logistic regression, nearest neighbour method, random forest, neural networks, Naive Bayes, Anomaly detection: iForest	Structured banking transaction datafile	Accuracy, precision, cover, F1-rate	Highest accuracy: neural network (its runtime is longer than that of other algorithms). Highest value from the aspect of cover: RF. Best result in anomaly detection: iForest. Classification produced a better result than anomaly detection, however the combination of the two resulted in higher precision and lower human resources requirements.

Source: own edition

in the area of money laundering and terrorism financing.

CONCLUSIONS

The author points out that there is no one ideal algorithm. In fact, each prediction made with an algorithm can be considered as an optimisation problem, as the objective is the optimisation of a given target function. The algorithm is selected on the basis of the examination of the problem to be solved. While the objective of the linear regression model is to minimise the (squared) difference between the predictions and the actual value, the SVM algorithm performs linear categorisation on the hyperplane by using a separating plane, and the margin received (the space determined by the hyperplane that is parallel with the separating plane, containing no teaching data points) is as wide as possible. The independent decision trees in the random forest model make their individual decisions on the basis of a random sample, and finally, through a majority vote, they provide the solution to the classification problem. While the Naiv Bayes classifier calculates the probability of the data's belonging to the given class on the basis of the input vector value.

In the case of multiple possible solutions, the measuring and the comparison of the efficiencies of individual algorithms is possible on the basis of the following criteria:

(1) time complexity (time used for teaching),

(2) execution time,

(3) memory/storage requirement (memory needed during run),

(4) possibility of parallel operation (concurrent performance of multiple operations, running on multiple machines),

(5) parametricity,

(6) linearity.

Table 6 contains the comparison along the random forest model, the nearest neighbour method, the SVM, the *K*-means and the linear regression algorithm factors mentioned above.

It is worth pointing out that in the cases of algorithms indicated with 'no' values in the table from the aspect of parallel operation, there are a number of methods to establish the ability to carry out parallel operation.

Based on the comparison of the examined algorithms, we can say that the business objective, the harmony of the underlying theoretical aspects of the algorithm and the quality of the available data are issues that cannot be separated from each other. With the mitigation of the variance error trade off and the improvement in the accuracy of prediction, the combination of the prediction of algorithms (hybridisation, ensemble models) may offer a reliable solution to avoid unilateral analysis. At the same time, we should be aware that it does not automatically result in a model of higher performance and accuracy.

CONCLUSION

The validation of machine learning models applied in the field of preventing money laundering and terrorism financing is not feasible without human resources. However, machine learning models significantly contribute to the freeing up of working hours to focus on more value-added activities and support the quality of work done by employees. There is no one ideal algorithm. The consideration of the hybrid viewpoint originating from the cooperation of the business aspect, the IT area and the visionary management is an indispensable precondition of the integration of ML models into the process and their successful utilisation. The selection among algorithms is supported

Table 6

COMPARISON OF SELECTED ALGORITHMS BASED ON THE FACTORS OF CALCULATION COMPLEXITY

Considerations of comparison	*RF*	*K-NN*	*SVM*	Linear regression	Logistic regression	Naive Bayes	*K*-means
Teaching time complexity	$O(n \times log(n) \times m \times T)$ Slow in the case of lots of observations	$O(k \times n \times m)$ Slow in the case of lots of observations	$O(n^{\wedge}2)$ is low in the case P, while $O(n^{\wedge}3)$ is high in the case of P learning time is long	$O[m^2(n + m)]$ Teaching time is long, execution time is short	$O(n \times m)$ Short and efficient execution time (especially on small database)	$O(n \times m)$ Low teaching time	$O(I \times CL \times n \times m)$ Slower with large dataset
Performance time	$O(DoT \times T)$	$O(n \times m)$	$O(S \times m)$	$O(m)$	$O(m)$	$O(o \times m)$	$O(n \times m{+}CL \times m)$
Storage requirement	$O(DoT \times T)$	$O(n \times m)$	$O(n^{\wedge}2)$ alacsony	$O(m)$ alacsony	$O(m)$ alacsony	$O(o \times m)$ alacsony	$O[(CL \times n)] \times m$
Parrallel operation	Yes	Yes	No	No	No	No	No
Parametricity	Nonparametric	Nonparametric	Nonparametric	Parametric	Parametric	Parametric	Parametric
Linearity	Nonlinear	Nonlinear	Linear/Nonlinear (kernel)	Linear	Linear	Linear	Linear

Note: Abbreviations of factors used in Table 6 are as follows:

n: teaching dataset data volume — *m:* number of data characteristics/dimensions

P: penalty parameter (parameter of penalty for incorrect classification) — *T:* number of decision trees, teaching time is long

k: number of neighbours — *S:* number of support vectors

DoT: depth of decision trees — *CL:* number of clusters

I: number of iterations — *c:* number of classes

Source: own edition

by the underlying theoretical aspects of individual algorithms, as well as other factors used for comparison. It is, however, necessary to emphasize that data preparation works forming a significant part of model building (approx. 80 per cent) may also have a major influence on the results of the received predictions. Depending on the field, the ratio of events connected to the given optimisation problem varies within the population, and depending on that, the modelling party may select the application of different machine learning methods (supervised, unsupervised, reinforcement learning) and algorithms.

LIMITATION

The comparison of typical algorithms applied in the examined field may be carried out as a continuation of this study by using real banking databases. ■

References

Alexandre, C., Balsa, J. (2018). A Multi-Agent System Based Approach to Fight Financial Fraud: An Application to Money Laundering. *Preprints*, 2018010193, http://doi.org/10.20944/preprints201801.0193.v1

Álvarez-Jareño, J. A., Badal-Valero, E., Pavía, J. M. (2017). Using machine learning for financial fraud detection in the accounts of companies investigated for money laundering. Working Papers 2017/07, Economics Department, Universitat Jaume I, Castellón (Spain), https://ideas.repec.org/p/jau/wpaper/2017-07.html

Chen, T., Guestrin, C. (2016). XgBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp. 785-794, http://dx.doi.org/10.1145/2939672.2939785

Chen, Z. et al (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, 57, pp. 245-285, http://doi.org/10.1007/s10115-017-1144-z

Deng, X., V. Roshan, J., A. Sudjianto, Jefi Wu, C. F. (2012). Active Learning Through Sequential Design, With Applications to Detection of Money Laundering. *Journal of the American Statistical Association,* 104(487), https://doi.org/10.1198/jasa.2009.ap07625

Drezewski, R., Sepielak, J., Filipkowski, W. (2012). System supporting money laundering detection. *Digital Investigation*, 9(1), pp. 8-21, https://doi.org/10.1016/j.diin.2012.04.003

Johari, R. J., Zul, N. B., Talib, N., Hussin, S. A. H. S. (2020). Money Laundering: Customer Due Diligence in the Era of Cryptocurrencies. Proceedengs of the 1st International Conference on Accounting. *Management and Entrepreneurship,* (ICAMER, 2019), https://doi.org/10.2991/aebmr.k.200305.033

Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control*, 23(1), pp. 173-186, https://doi.org/doi/10.1108/JMLC-07-2019-0055/full/html

Kannan, S. R., Somasundaram. K. K. (2017). Autoregressive-based outlier algorithm to detect money laundering activities. *Journal of Money Laundering Control*, 20(2), pp. 190-202, https://doi.org/10.1108/JMLC-07-2016-0031

Khan, N. et al (2013). A Bayesian approach for suspicious financial activity reporting. *International Journal of Computers and Applications,* 35(4), https://doi.org/10.2316/Journal.202.2013.4.202-3864

Le-Khac, N. A. et al. (2010). A Data Mining-Based Solution for Detecting Suspicious Money Laundering Cases in an Investment Bank. 2nd International Conference on Advances in Databases. *Knowledge, and Data Applications*, DBKDA, pp. 235-240, https://doi.org/10.1109/DBKDA.2010.27.

Luo, X. (2014). Suspicious Transaction Detection for Anti-Money Laundering. *International Journal of Security and Its Applications,* 8(2), pp. 157-166, https://doi.org/10.14257/ijsia.2014.8.2.16

Patil, P. S., Dharwadkar, N. V. (2017). Analysis of banking data using machine learning. International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). pp. 876-881, https://doi.org/10.1109/I-SMAC.2017.8058305

Rocha-Salazar, J-d-J., Segovia-Vargas, M-J., Camacho- Mi´nanom M-d-M. (2021). Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Systems with Applications*, 169, https://doi.org/10.1016/j.eswa.2020.114470

Rouhollahi, Z., Beheshti, A., Mousaeirad, S., Goluguri, S. R. (2021). Towards Artificial Intelligence Enabled Financial Crime Detection. ArXiv abs/2105.10866, pp. 538-546, https://doi.org/10.1145/3487664.3487740

Savage, D., Wang, Q., Chou, P., Zhang, X., Yu, X. (2016). Detection of money laundering groups using supervised learning in networks. AAAI-17 Workshop on AI and Operations Research for Social Good, Australia, https://doi.org/10.1108/JMLC-07-2019-0055

Szikora, A., Nagy, B. (2020). Mesterséges intelligencia a pénzügyi szektorban. Online: https://www.mnb.hu/letoltes/baksa-szikora-andrea-nagy-benjamin-ai-a-penzugyi-szektorban-final.pdf

Tóth, Z. B. (2018). Magyarország válaszlépései a pénzmosással és terrorizmusfinanszírozással kapcsolatos kihívásokra. [Hungary's Responses to the Challenges Related to Money Laundering and Terrorism Financing.] *Polgári Szemle, [Civic Review],* 14(1-3), pp. 418-427, https://doi.org/10.24307/psz.2018.0832

Wang, S. N., Yang, J. G. (2007). A Money Laundering Risk Evaluation Method Based on Decision Tree. IEEE. The 6th International Conference on Machine Learning and Cybernetics, 1, pp. 283-286, https://doi.org/10.1109/ICMLC.2007.4370155

Watkins, R. C., Reynolds, K. M., DeMara, R. F., Georgiopoulos, M., Gonzalez, A. J., Eaglin, R. (2003). Tracking dirty proceeds: Exploring Data Mining Technologies as Tools to Investigate Money Laundering. *Journal of Policing Practice and Research: An International Journal,* 4(2), pp. 163-178, https://doi.org/10.1080/15614260308020

Wegberg, Van R., Oerlemans, J.-J., Deventer, Van O. (2018). Bitcoin money laundering: mixed results? An explorative study on money laundering of cybercrime proceeds using bitcoin. *Journal of Financial Crime*, 25(2), pp. 419-435, https://doi.org/10.1108/JFC-11-2016-0067

Wei, W., Li, J., Cao, L., Ou, Y., Chen, J. (2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web 16, pp. 449-475, https://doi.org/10.1007/s11280-012-0178-0

Zhang, Y., Trubey, P. (2019). Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection. *Computational Economics,* 54, pp. 1043-1063, https://doi.org/10.1007/s10614-018-9864-z

BIS (2001). Consultative Document. Customer Due Diligence for Banks. Technical report, https://www.bis.org/publ/bcbs85.htm

FATF (2012). 40 Recommendations. Online: https://www.cfatf-gafic.org/documents/fatf-40r

FATF (2019). Hungary's progress in strengthening measures to tackle money laundering and terrorist financing. Moneyval 3rd Follow Up Report Hungary, https://www.fatf-gafi.org/publications/mutualevaluations/documents/fur3-hungary-2019.html

FATF (2021). Opportunities and Challenges of New Technologies for AML/CFT, FATF, Paris, France, https://www.fatf-gafi.org/publications/fatfrecommendations/documents/opportunities-challenges-newtechnologies-aml-cft.html

FATF (2021a). Anti-money laundering and counter-terrorist financing measures Hungary 4th Enhanced Follow-up Report April 2021, http://www.fatf-gafi.org/media/fatf/documents/reports/fur/Moneyval-FUR-Hungary-2021.pdf

FATF (2021b). International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation, The FATF recommendations. FATF, Paris, France, http://www.fatf-gafi.org/recommendations.html

FSB (2017). Artificial intelligence and machine learning in financial services Market developments and financial stability implications, https://www.fsb.org/wp-content/uploads/P011117.pdf

Legislation:

Act LIII of 2017 on the Prevention and Combating of Money Laundering and Terrorist Financing, https://njt.hu/jogszabaly/2017-53-00-00

Commission Delegated Regulation (EU) 2017/565, https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:32017R0565&from=de