

Machine learning methods application for consumer banking

Andrii Kaminskyi^{1,*}, Maryna Nehrey^{2,**}, and Larysa Zomchak^{3,***}

¹Department of Economic Cybernetics, Taras Shevchenko National University of Kyiv, 64 Volodymyrska Str., Kyiv, 01026, Ukraine

²Department of Economic Cybernetics, National University of Life and Environment Science of Ukraine, 15 Heroyiv Oborony Str., Kyiv, 03041, Ukraine

³Department of Economic Cybernetics, Ivan Franko National University of Lviv, 1 Universytetska Str., Lviv, 79000, Ukraine

Abstract. Machine learning (ML) methods are effective tools for analysis of many actual problems in modern banking. Increasing growth of data and rapid digitalization underpin the acceleration of ML implementation. These processes are especially noticeable in consumer banking because banks have millions of the retail customers. The first goal of our research is to form an extended review ML application in consumer banking. From one side we have identified the most developed ML methods, which are applied in this segment (for example different types of regressions, fuzzy clustering, neural network, principal component analysis etc.). From the other side, we point out two multi-purpose tools used by banks in consumer segment intensively, namely scoring and clustering. Secondly, our goal is to present some innovative applications of ML methods to the analysis of each task. This includes several applications for scoring models and fuzzy clustering application. All applications are oriented to make banks business processes more effective. Considered applications were realised on real data from the Ukrainian banking industry.

1 Introduction

Machine Learning (ML) is a dynamically growing class of methods that has many successful applications [1, 2]. One of the areas of productive ML using is modern banking and more widely modern financial institutions. The basic reason for such use arises from digitalization and intensively implementing online technologies in the financial sphere. These processes generate Big Data that can be involved in ML processing. Data handling triggers for further development of existing methods and offers new ideas. We want to emphasize some directions in banking (especially in lending banking) in which ML methods were productively used and perspectives of development are fruitful.

The first direction is a concern to generalized scoring methodology. This methodology is highly implemented in different business processes in banking: estimation creditworthiness of borrowers (credit scoring), identifying potentially profitable customers (marketing scoring), anti-fraud systems (fraud scoring), and so on. Almost all ML methods were used for scoring (first of all credit scoring) construction. Today's package of scoring construction methods includes Multivariate Adaptive Regression Splines, Support Vector Machine (SVM), k-nearest-neighbors method, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and, of course, Artificial Neural Network (ANN). In general, ANN maybe now one of the most popular methods in scoring building tools. We

should highlight paper [3] where presented comprehensive results of the comparison of ML credit scorings with classical expert-based scorings. All types of scoring are generalized for involving new (online) data. It should be complemented that ordering banks them-selves also getting here.

The second direction where ML methods were multi-skilled implemented in banking is clustering. Modern banks operate with millions of customers. Each customer can be characterized by vectors with hundreds, maybe a thousand components (customer characteristics). Applying ML arises strong possibilities for clustering customers, identification of their behavior and, as consequence, improved productivity of banking services. When a bank wants to realize clustering, it needs to identify "hidden knowledge" that will help to divide customers effectively into a set of clusters. Here dominates the concept of K-Means algorithms. It involves k-mean, improved k-mean, k-medians, applied hierarchical clustering. We design fuzzy clustering in this paper. Of course, the task of clustering banks themselves is also treated by such an ML method.

The third direction is the cybersecurity of banks. This direction ever more actual through moving to online interaction with customers. The logic of applying ML methods for cybersecurity tasks may be based on Gartner's PPDR model [4]. This model points out five categories: prediction, prevention, detection, response, and monitoring. A good overview of cyber-risk and cyber-security problems for financial institutions can be found in [5]. The

*e-mail: kaminskyi.andrey@gmail.com

**e-mail: marina.nehrey@gmail.com

***e-mail: lzomchak@gmail.com

spectrum of applying ML in this sphere is described here [6].

We tried to systematize methods of ML in the context of the above mentioned directions. The first two considered spheres have economic nature and the third is more technical (but of course concerns economic consequences!). Our focus was concentrated on the first two spheres which have economic nature and are closely connected with business processes.

Part 2 is devoted to the literature review and describes the application of the most effective methods. Part 3 contains some illustrations of applying ML, which was partially presented in our researches. The conclusion involves a point of view for development.

2 Materials and methods

Machine learning is the concept that a computer program can learn and adapt to new data without human interference [7]. Machine learning is a field of artificial intelligence that keeps a computer's built-in algorithms current regardless of changes in the worldwide economy. The application of machine learning methods in economics and banking discussed in researches [8–11]. Machine learning includes the next classes of methods: Supervised Learning, Unsupervised Learning, Reinforcement Learning, Ensemble methods, Neural Networks and Deep Learning (table 1).

Regression. Logistic regression is a classification method that constructs a binary variable result prediction (1/0, True / False, Yes / No, Good / Bad) for a given set of independent variables.

When constructing logistic regression use dummy variables. The maximum likelihood estimation method is used to estimate logistic regression parameters. Logistic regression calculates the probability of Y when realizing certain values of χ . Log-likelihood is the sum of the probabilities associated with the predicted and actual values of Y . Deviations in logistic regression have a distribution of χ^2 . Performance indicators for the logistic regression model: Akaike Information Criteria; Null Deviance and Residual Deviance; the error matrix as a tabular representation of actual and predicted values; ROC curve (Receiver Operating Characteristic).

Fuzzy clustering. Fuzzy clustering is a class of algorithms in which the distribution of data points for clustering is not "clear" (but supposes "fuzzy"). It is used in the Neural-fuzzy systems to determine fuzzy sets if they are unknown a priori. Fuzzy sets are like projections of clusters on each dimension. An a priori knowledge and cluster analysis combination lets to refine the parameters of the membership function. The disadvantage of this method of determining fuzzy sets is the complexity of their interpretation.

One of the most commonly used methods is the fuzzy c-means (FCM) method. It assigns a fuzzy membership value to each object, based on its distance to the cluster centers. The membership of the data point in that cluster will be higher than its membership in the other clusters if the data point is closer to the center of the cluster.

The FCM method is an iterative procedure of sequentially improving a certain fuzzy initial partition user-defined or automatically generated by a specific heuristic rule. According to the algorithm, on each iteration values of the membership functions of fuzzy clusters and their typical representatives are recursively listed. The FCM method will terminate when a specified a priori finite number of iterations is performed, or when the minimum absolute difference between the values of the membership functions on two consecutive iterations does not fall below some a priori setpoints.

Neural network technologies. Artificial Neural Networks (ANN) are widely used in finance and insurance problems. ANN using is credit scoring construction to good effect. Many different methods are applied in this direction. They are based on the representative types for supervised and unsupervised ANN. The main advantage of ANNs is that dependency between variables does not necessary to characterize. The quality of the ANN applying in credit scoring application can be explained by the Big Data of modern banks.

3 Results and discussion

3.1 Parametric scoring model based on the concept of survival

The classic scoring model is based on the linear function

$$Z = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (1)$$

where x_1, x_2, \dots, x_n – borrowers characteristics; a_1, a_2, \dots, a_n are the weights of characteristics that reflect their significance.

Such models have been implemented at all stages of relationships between borrowers and banks in the consumer lending segment. The illustration presents at figure 1.

Our practice indicates that all these scorings in differing degrees have been applied in Ukrainian consumer banking. This models actively enriches by new data from digitalization and online lending last five years. The data from customer behavior on websites, cells using types, and other data can be included in scorings.

One of the typical characteristics of above-mentioned scoring model is a static state. Below we want to present specific approach that includes dynamic consideration. The logic lies in inclusion time parameter t in scoring coefficients. This model will be "dynamic credit scoring model":

$$h(i, x, t) = \ln \frac{P(i)}{1 - P(i)} = a_0(t) + a_1(t)x_{i1} + a_2(t)x_{i2} + \dots + a_n(t)x_{in} \quad (2)$$

$P(i)$ is the probability that the i -th customer (borrower) will be "Good" in the period from 0 to t ; x_{ij} is the j -th characteristic of the i -th customer, where $i = 1, \dots, k$, $j = 1, \dots, n$; $a_j(t)$ is the coefficient of the model at time t , where $j = 1, \dots, n$; t is a time for estimation.

Table 1. Machine learning methods

Class	Subclass	Method	Researches
Supervised Learning	Regression	Linear regression Polynomial regression Ridge lasso/regression	[12–20]
	Classification	Logistic regression Decision trees Support vector machine Naïve Bayes k-nearest neighbor	
Unsupervised Learning	Clustering	k-means Agglomerative Mean-shift Fuzzy C-means DBSCAN	[21–29]
	Rule Engine	Association rules Eclat Apriori FP-growth	
	Dimensionality Reduction	Principal Component Analysis Partial Least Squares Regression Principal Component Regression T-distributed Stochastic Neighbor Embedding Singular Value Decomposition Mixture Discriminant Analysis Linear Discriminant Analysis	
	Reinforcement Learning	Genetic algorithm Q-learning SARSA Deep Q-Network	[30, 31]
Ensemble		Boosting Gradient boosting machines Bagging Random forest AdaBoost Stacked Generalization	[32, 33]
Neural Networks and Deep Learning	Convolutional neural network	DCNN	[34–42]
	Recurrent neural network	Liquid State Machine (LSM) Long short-term memory (LSTM) Gated recurrent unit (GRU)	
	Generative Adversarial Networks	GAN	
	Autoencoders	Seq2seq	
	Perceptrons	MLP	

Model (2) differentiate customers much more accurately and qualitatively because it will show the likelihood of not just going out of time but getting out of time at some point in time. This will make it possible to build the financial model more accurately. In addition, if you apply this model to application scoring, it is possible to implement a specific strategy for working with customer (including monitoring, reminder system and other).

Let us illustrate this approach to collection scoring construction. This model was elaborated on Ukrainian data and applied to good effect for collection business processes [43]. Model was constructed on data pool of 50,000

debtors. The data include more than 30 characteristics, including the type and status of credit accounts, a partial history of payments to them, social and demographic parameters of debtors, information about their education and employment. Among the many factors, nine non-correlated indicators were identified, with a significant effect on the dependent variable (table 2).

Statistical significance was verified using χ^2 statistics, the Cramer coefficient, and the Information Value. The dependent variable assumed a value of 1 if the debtor made at least 1 payment during 8 quarters period. Value 0 was

Stages of credit granting processes	Identify profitable borrowers	Identify borrowers which repeat loan	Identify fraud-behavior for receiving loan	Differentiate borrowers on the base application data	Differentiate borrowers on the base credit history	Identify optimal strategy for collection
Scoring types	Marketing scoring	Lead-generation scoring	Fraud scoring	Application scoring	Behavioral scoring	Collection scoring

Figure 1. Stages of relationships between borrowers and banks

Table 2. Indicators of statistical significance of model parameters for the whole period

Parameter name	Parameters description	χ^2	Kramer coefficient	Information value
x_1 Redemption model	The scheme of repayment 6 months before the purchase of debt portfolio. For example, 101010, where 1 is the payment and 0 is the non-payment. There are 64 such values in total, and they are ordered by probability of the first payment	7980	0.278	0.60
x_2 Non-payment days	The number of days since the last payment was made before buying debt portfolio	2842	0.166	0.50
x_3 Delay days	Number of delays days	6554	0.252	0.41
x_4 Principal and interest	The ratio of principal and interest to the total debt	5187	0.224	0.35
x_5 Term of the loan	Term of the loan at the context number of days after issuance	7373	0.268	0.32
x_6 Paying Principal	The Principal of the loan repaid	3968	0.196	0.30
x_7 Amount of penalties	Amount of all penalties and penalties for late payment	2608	0.159	0.20
x_8 Issue date	Date when loan was issued	2659	0.161	0.13
x_9 Product	Type of credit product (monetary/commodity)	1490	0.120	0.12

used otherwise. Correlation checking was performed using correlation analysis.

The effect of the characteristics influence (table 3) on the probability of debt payment in the dynamics is estimated. At each time moment t , “Information Value” metrics were calculated for each quarter. The probability of the event “Good” occurrence was taken as the dependent variable – the debtor made the first payment within a time period from 0 to t inclusive, (the alternative event “Bad” did not occur), where the parameter t in the model (2) denotes the quarter from 1 to 8, in which this event happened.

Debtors were then assigned scoring points on a scale of 0 to 100. Each group created scoring points: 0–20; 20–40; 40–60; 60–80; 80–100 corresponds to the probability $P(i)$ that the i -th debtor will make his payment in the period from 0 to t , which depends on both the score and the time.

In all models, the normalized R-square – correlation coefficient – close to 1, so the relationship in the models close. Each variable is significant and the model is adequate according to Student’s and Fisher’s criteria. The magnitude of all coefficients decreases with each quarter, except for the constant that compensates for this decrease.

The basic mathematical differences are time changing in weights. This can be used to collection strategies development while optimizing the costs of implementing it.

3.2 Neural network technologies in debt portfolio management

The ANNs usage provides an efficient classification of bank debtors into groups. Debtors are similar to each other in terms of risk characteristics in each group. The economic logic of this is the management of debt collecting strategies applying. Neural networks allow the classification of debtors from the training sample into groups of a more complex geometric shape than the classical linear discriminant function and visualize it. The most valuable property of ANNs is the ability to learn from multiple examples in cases where patterns are unknown and relationships between input and output are not obvious. In such cases, both traditional statistical and expert methods are ineffective.

Our research of the consumer debtors portfolios involves applying ANNs methodology to estimate debtors. The study finds effectiveness of problem structuring:

Table 3. Dynamics of scoring factors over time

Parameter	Coefficient	1	2	3	4	5	6	7	8
Redemption model	$a_1(t)$	0,16	0,09	0,11	0,09	0,10	0,09	0,09	0,09
Non-payment days	$a_2(t)$	0,40	0,35	0,33	0,32	0,31	0,31	0,29	0,29
Delay days	$a_3(t)$	0,25	0,22	0,22	0,20	0,20	0,20	0,19	0,19
Principal and interest	$a_4(t)$	0,26	0,24	0,22	0,22	0,21	0,20	0,20	0,19
Term of the loan	$a_5(t)$	0,20	0,23	0,18	0,18	0,17	0,17	0,16	0,16
Paying Principal	$a_6(t)$	0,22	0,21	0,19	0,19	0,18	0,18	0,17	0,17
Amount of penalties	$a_7(t)$	0,19	0,20	0,17	0,17	0,16	0,16	0,16	0,15
Issue date	$a_8(t)$	0,16	0,20	0,15	0,16	0,14	0,15	0,14	0,13
Product	$a_9(t)$	0,21	0,19	0,18	0,17	0,17	0,16	0,16	0,16
Y-intersection	$a_0(t)$	-7,11	-6,13	-5,45	-5,16	-4,85	-4,71	-4,49	-4,30

- the contact problem with debtors;
- the debtors’ insolvency.

The first problem is raised from the fact that a significant part of debtors is non-contact. This does not allow the application of soft-charging techniques and leads to high-value direct contact or using legal procedures, which are also costly. The practice has shown that the proportion of non-contact debtors approximately 70-80%.

The second problem is that some of the contact debtors refuse to pay for various reasons: lack of funds, unwillingness to pay due to high-interest penalties, etc. The implementation model in practice has shown that 40-50% are contact debtors. The ANN was applying twofold as pictured in the scheme (figure 2).

Based on the AANs using two different scorings were constructed: contact scoring and solvency scoring.

In doing so emphasis is put to such characteristics as socio-demographic (age, marital status, educational level, a region of residence, etc.), professional (employment status, vocational qualification, etc.), loan parameters (amount, interest rate, duration, etc.) and, of course, characteristics of overdue [44].

In both cases, scoring is the first step in applying Self Organizing Maps (SOM) or Kohonen cards to the training sample. Kohonen map is a special type of neural network that allows identifying hidden structures and patterns through learning neural networks. A special algorithm performs clustering based on two-dimensional visualization.

Neural networks technology creates a series of clusters which includes homogeneous debtors.

When using Kohonen maps, there is a problem of choosing between detail and visualization. The increasing details in one of these characteristics lead to the deterioration of the other. Really, more detailed consideration complicates economic analysis leads to the difficulty of visualization. On another side, reducing it can lead to the loss of important patterns. Our analysis showed that in considered cases it is suitable to divide from 6 to 9 clusters. The authors’ experience shows that in most cases, splitting into 7 clusters is optimal.

The next question is concerned with correlation analysis of the debtor’s characteristics. To avoid correlation problems and to include non-dependent characteristics, it

is advisable to select only those with a correlation coefficient not exceeding 0.6. Another way concerns applying principal component analysis (PCA). However, the use of PCA, in this case, maybe difficult due to the complexity of their economic interpretation. After correlation analysis, the optimal levels of influence (significance) of characteristics are determined with the help of classification trees to improve model accuracy. Our analysis of different consumer loan debt portfolios leads to identify the most relevant contact characteristics (table 4).

The obtained scoring allows you to sort the debtors by the level of contact probability: the higher the scoring value, the more likely they are to contact him. Based on contact scoring, the following logic for managing arrears may be suggested. Before working remotely with a new portfolio of debtors, the probability of making contact is assessed by considered scoring. All debtors are allocated by three scoring classes: high-contact, medium-low and low-contact. After the introduction of remote work with the portfolio, the debtors are already factually divided into contact and non-contact. As a consequence, all contact information is being worked on further in debt collection. At the same time, debtors with a high value of contact scoring but no contact, in reality, should be continue elaborated for identifying contacts. It is logically not spending the time for those non-contracted debtors who have low contact scoring. Namely, for non-contact debtors with low contact scoring values, the following strategies may be employed:

- write-off strategy if the debt is not large;
- a strategy for obtaining additional information through a request to the credit bureaus. If there is a significant amount of debt, put a low priority on further work with his debtor. If there are no so much open loans, give high priority;
- transfer to legislative recovery, if the amount of debt is considerable.

Applying an approach based on these strategies creates frameworks for seeking optimal allocational resources. Summing up, we can conclude that ANNs are an effective technique for elaborating strategies of prioritizing collection efforts.

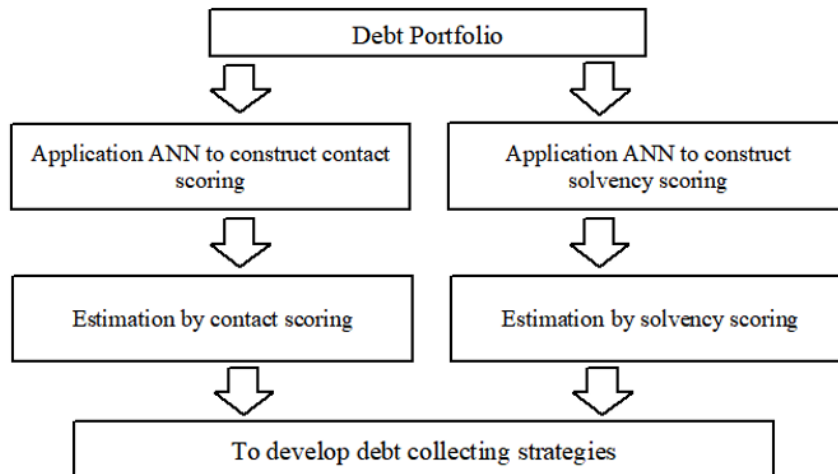


Figure 2. Applying ANN in debt portfolio

Table 4. Contact characteristics

Contact characteristics of the debtor	Medium level of significance debt portfolios
Principal debt and interest / monthly payment	36,55%
Principal debt and interest	22,25%
The time from the last payment to the beginning of the charge	13,04%
Debtor's age	6,54%
The amount paid by the debtor to late payment	6,33%
Loan amount	5,42%
Total / principal and interest	4,33%
The amount of the last payment made by the debtor	3,9%
Number of payments which were paid	1,11%
Become a debtor	0,41%

3.3 Fuzzy clustering of bank's consumer loan portfolio

The intensive development of consumer lending last decades leads to the fact that today banks possess hundreds of thousands or millions of borrowers in their credit portfolios. These are really Big Data. Credit portfolios involve different segments: mortgages, car loans, unsecured consumer loans, credit cards, and others. One of the crucial objectives consists of clustering borrowers. Typically, this corresponds to lead-generation. The objective is to find different clusters of borrowers who successfully paid previous loans. Such different clusters include borrowers with different characteristics, different behaviors, and preferences. So, it is logical to construct a corresponding marketing strategy. The clustering can be done by various approaches which involve choosing economic parameters for clustering basis and mathematical techniques.

Our research in clustering large credit portfolios of banks leads to forming an approach to applying ML for clustering. The first component of our approach corresponds to the identification of basic economic indicators for clustering. The second component is the fuzzy clustering application.

First component realization. We have chosen three indicators at the framework of the first component. They are some unification of risk and profitability indicators. By our logic, these indicators can be calculated for borrowers which have closed loans. The bank is planning to propose some lead-generating product. These indicators are:

- score of credit bureau(s);
- amount of loans which borrower was granted;
- level of overpayment for previous loans.

The economic logic of such indicators is the following.

The score of the credit bureau provides information about the risk level of the borrower. A lower score value (high risk) identifies a “bad” borrower which is not paid or overloaded. It may be logical to exclude such borrowers from further consideration or apply a special approach constructed for the corresponding cluster. A high score value (low risk) identifies a “good” borrower who paying off his/her loans. It will be a nice borrower for lead generation but they are not high profitably. Really, such borrower pays “accurately and timely”. The average score value corresponds to the borrower which periodically hits in overdue but then pays all debt with a penalty, fees, and others. In reality, it is more profitable. Of course, it is average consideration. Different types may be in this category.

The second indicator is the loan amount. The basic economic logic here: low amount generates low profit for the bank but more often overpayment.

Third our indicator reflects the profitability of the borrower. We were following conceptually approach of Storbacka [45] for separate customers and provided extension it for borrowers. According to our approach borrowers can be divided into four classes:

- A – borrowers with high overpayments,
- B – borrowers which pay “accurately and timely”,
- C – there was some payments but the amount of the loan was not paid,
- D – FPD (First payment default and no any payments).

Of course, A-type is more profitable. D, in any case, should be excluded.

Second component realization. The abovementioned estimations of borrowers arise tasks for clustering. This clustering should consider the risk of insolvency, profit from an overpayment, and the amount of loan. Clustering is economically significant because it is possible to construct some (marketing) strategy for each cluster. The consideration classical C-means approach leads to some “strict” separation. Very often it may be not optimal for customers. Because some borrowers can “fuzzy” belong to different clusters with different values of membership functions. So, we applied fuzzy C-means clustering. The conceptual difference between approaches is illustrated below figure 3.

The economic benefits are construction more effective lead-generating strategies for such customers. Really, the classical C-mean clustering will be 3 clusters (if we specify 3), but fuzzy C-mean clustering provides 7 clusters. This allows forming a more adequate set of strategies.

So, fuzzy clustering leads to a more advanced approach for cluster creation which adequate forming strategies of lead-generation.

4 Conclusion

Machine Learning has been successfully developing in many spheres. Sometimes, the applications are impressive. Modern banking really excellent “test site” for different ML methods and techniques. This is mainly due to Big Data and understanding the practical importance of applying ML. Really, the average retail bank typically has an intensive inflow of customers and very many customers in the portfolio. Each customer as from inflow as from portfolio can be characterized by hundreds of indicators. This is Big Data with hidden patterns of customer behavior and preferences.

Tools for business development in consumer banking are scorings and clustering. It is absolutely logical to apply ML and AI for the solution of objectives to construct effective scorings assessment and clustering procedures. The results are multifaced. Our paper illustrated a couple of solutions for consumer banking based on Machine learning applications. All these solutions were implemented

and indicate effectiveness. The huge growth of data according to digitalization in banking and developing fintech produce new objectives and new spheres for applying ML and AI.

Machine learning methods are self-developing areas of researches with the synergetic interaction between them. This fact has been considered in the mentioned examples.

References

- [1] A. Kiv, V. Soloviev, S. Semerikov, H. Danylchuk, L. Kibalnyk, A. Matviychuk, *CEUR Workshop Proceedings* **2422**, 1 (2019)
- [2] A. Kiv, P. Hryhoruk, I. Khvostina, V. Solovieva, V. Soloviev, S. Semerikov, *CEUR Workshop Proceedings* **2713**, 1 (2020)
- [3] L. Munkhdalai, T. Munkhdalai, O.E. Namsrai, J.Y. Lee, K.H. Ryu, *Sustainability* **11**, 699 (2019)
- [4] P. Ganapathi, D. Shanmugapriya, *Handbook of Research on Machine and Deep Learning Applications for Cyber Security* (IGI Global, 2019)
- [5] A. Bouveret, *Cyber risk for the financial sector: A framework for quantitative assessment* (International Monetary Fund, 2018)
- [6] C. Chio, D. Freeman, *Machine learning and security: Protecting systems with data and algorithms* (O’Reilly Media, Inc., 2018)
- [7] Investopedia.com, *Machine learning* (2021), <https://www.investopedia.com/terms/m/machine-learning.asp>
- [8] M. Leo, S. Sharma, K. Maddulety, *Risks* **7**, 29 (2019)
- [9] V. Babenko, A. Panchyshyn, L. Zomchak, M. Nehrey, Z. Artym-Drohomyretska, T. Lahotskyi, *WSEAS Transactions on Business and Economics* pp. 209–217 (2021)
- [10] A. Kiv, V. Soloviev, S. Semerikov, H. Danylchuk, L. Kibalnyk, A. Matviychuk, *CEUR Workshop Proceedings* **2422**, 1 (2019)
- [11] N. Volkova, N. Rizun, M. Nehrey, *Data science: Opportunities to transform education* (CEUR-WS, 2019), Vol. 2433, pp. 48–73, ISSN 16130073
- [12] C. Brownlees, C. Hans, E. Nualart, *Journal of Monetary Economics* **117**, 585 (2021)
- [13] S. Tungsong, F. Caccioli, T. Aste, *Relation between regional uncertainty spillovers in the global banking system* (2017)
- [14] F. Shofiyah, A. Sofro, *Journal of Physics: Conference Series* **1108**, 012107 (2018)
- [15] V. Granaturov, V. Kaptur, I. Politova, *Economic Annals-XXI* pp. 52–56 (2015)
- [16] V. Granaturov, V. Kaptur, I. Politova, *Economic Annals-XXI* pp. 83–87 (2016)
- [17] N. Davydenko, A. Buriak, Z. Titenko, *Intellectual Economics* **13** (2019)
- [18] N. Klymenko, O. Nosovets, L. Sokolenko, O. Hryshchenko, T. PISOCHENKO, *Academy of Accounting and Financial Studies Journal* **23** (2019)

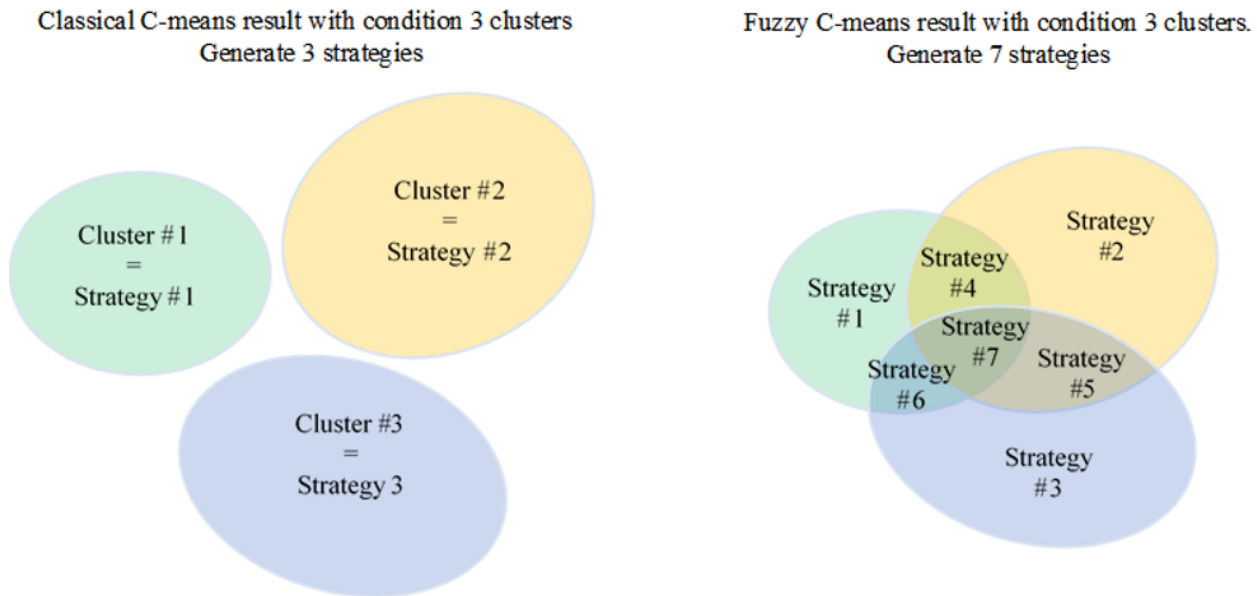


Figure 3. Classical vs Fuzzy C-means clustering

- [19] M. Kuzheliev, D. Zherlitsyn, I.I. Rekunenکو, A. Nechyporenko, G. Nemsadze, *Banks and Bank Systems* **15**, 94 (2020)
- [20] O. Kuzmenko, P. Šuleř, S. Lyeonov, I. Judrupa, A. Boiko, *Journal of International Studies* **13**, 332 (2020)
- [21] F. Butaru, Q. Chen, B. Clark, S. Das, A.W. Lo, A. Siddique, *Journal of Banking & Finance* **72**, 218 (2016)
- [22] H. Ercan, S. Sayaseng, *The cluster analysis of the banking sector in Europe*, in *Economics and Management of Global Value Chains* (2016), pp. 111–127
- [23] L. Fang, B. Xiao, H. Yu, Q. You, *Physica A: Statistical Mechanics and its Applications* **492**, 1997 (2018)
- [24] P. Gogas, T. Papadimitriou, A. Agrapetidou, *International Journal of Forecasting* **34**, 440 (2018)
- [25] H. Danylchuk, O. Kovtun, L. Kibalnyk, O. Sysoiev, *E3S Web of Conferences* **166** (2020)
- [26] O. Kuzmenko, S. Kyrkach, *Banks and Bank Systems* **9** (2014)
- [27] N. Khrushch, P. Hryhoruk, T. Hovorushchenko, S. Lysenko, L. Prystupa, L. Vahanova, *CEUR Workshop Proceedings* **2713**, 239 (2020)
- [28] Y. Hrabovskiy, V. Babenko, O. Al’Boschiy, V. Gerasimenko, *WSEAS Transactions on Business and Economics* **17**, 231 (2020)
- [29] K.B. Cyree, T.R. Davidson, J.D. Stowe, *Journal of Economics and Finance* **44**, 211 (2020)
- [30] L. Guryanova, R. Yatsenko, N. Dubrovina, V. Babenko, *CEUR Workshop Proceedings* **2549**, 1 (2020)
- [31] V. Derbentsev, A. Matviychuk, N. Datsenko, V. Bezkorovainyi, A. Azaryan, *CEUR Workshop Proceedings* **2713**, 435 (2020)
- [32] K. Tanaka, T. Kinkyو, S. Hamori, *Economics Letters* **148**, 118 (2016)
- [33] V. Ravi, H. Kurniawan, P.N.K. Thai, P.R. Kumar, *Applied soft computing* **8**, 305 (2008)
- [34] N. Mohammadi, M. Zangeneh, *IJ Information Technology and Computer Science* **8**, 58 (2016)
- [35] M. Alborzi, M. Khanbabaei, *International Journal of Business Information Systems* **23**, 1 (2016)
- [36] S. Saha, S. Waheed, *International Journal of Computer Applications* **161**, 39 (2017)
- [37] A. Matviychuk, I. Strelchenko, S. Vashchaiev, H. Velykoivanenko, *CEUR Workshop Proceedings* **2393**, 485 (2019)
- [38] J. Asare-Frempong, M. Jayabalan, *Predicting customer response to bank direct telemarketing campaign*, in *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)* (2017), pp. 1–4
- [39] W. Kanmani, B. Jayapradha, *International Journal on Recent and Innovation Trends in Computing and Communication* **5**, 293 (2017)
- [40] E.A.E. Dawood, E. Elfakhrany, F.A. Maghraby, *IEEE Access* **7**, 109320 (2019)
- [41] K. Lei, Y. Xie, S. Zhong, J. Dai, M. Yang, Y. Shen, *Neural Computing and Applications* **32**, 8451 (2020)
- [42] A. Petropoulos, V. Siakoulis, E. Stavroulakis, N.E. Vlachogiannakis, *International Journal of Forecasting* **36**, 1092 (2020)
- [43] A. Kaminsky, K. Pisanets, *Formation of market economy in Ukraine* pp. 136–142 (2012)
- [44] A. Kaminsky, V. Sikach, *Modeling and information systems in economics* **84** (2011)
- [45] K. Storbacka, *Journal of Marketing Management* **13**, 479 (1997)