CLUSTERING APPROACH TO ANALYSIS OF THE CREDIT RISK AND PROFITABILITY FOR NONBANK LENDERS

Andrii Kaminskyi[1], Maryna Nehrey[2]

[1]Department of Economic Cybernetics, Taras Shevchenko National University of Kyiv, e-mail: kaminskyi.andrey@gmail.com

[2]Department of Economic Cybernetics, National University of Life and Environmental Sciences of Ukraine, e-mail: marina.nehrey@gmail.com

Abstract. The paper is devoted to application customer profitability analysis to nonbanking lenders which predominantly focus on payday loans. The "Whale curve" has been constructed and special clusters were singled out. The approach based on joint together customer profitability management and credit risk management are considered. One significant effect was marked and grounding that higher risk interconnects with overpayments. The approach of fuzzy clustering was applied as the second approach to clustering. Such approaches may be considered as the basis of loan granting strategies elaborating.

Keywords: Customer Profitability Analysis, Clustering, Risk-Management, Nonbank Lending, Payday Loans, Machine Learning.

1. Introduction

The modern financial system is transforming. The intensive growth of fintech organizations reshapes the landscape of classical financial services. This turns critically on banking. Harvard Business Review Analytic Services surveyed more than 300 executives in classical financial institutions. Sixty-five percent identify as an essential threat by 2022 [1]. Fintech organizations focus on software, algorithms, and technology to propose services similar to banking and other financial services. Very often costs of their services lower cost than traditional financial institutions.

Online lending one of the crucial directions of developing fintech. Such companies have been largely successful last 10 years. They actively implement new technologies in credit risk-management and account for 15-20% of volume banking loans. Customers of online crediting tend to the younger generation (fig. 1).

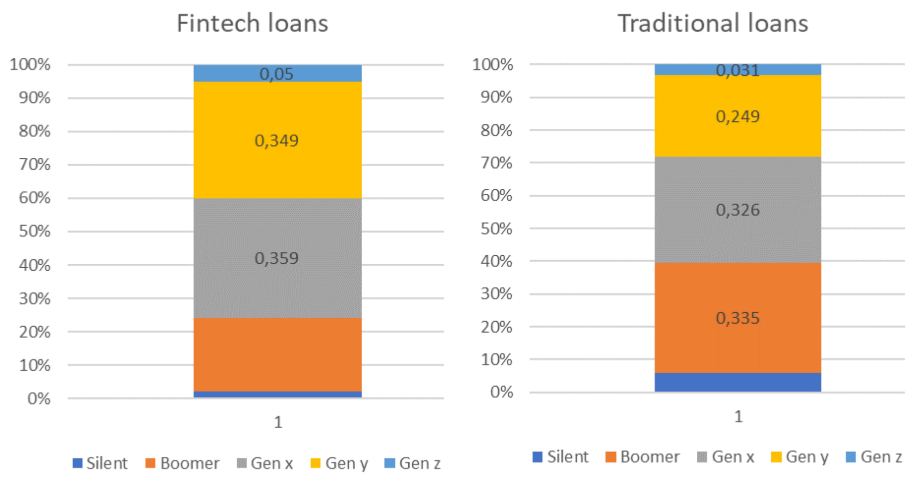

Fig. 1. Loans age structure

Online lenders use different technologies for loan granting processes: verifications, advanced scoring, and others. Most of them specialize in short-term loans (pay-day-loan segment). Credit portfolio becomes large. It is logical to apply customer profitability analysis (CPA) for optimizing strategies of loan granting. Our practical analysis of applying CPA indicated some economic effects. The first effect concerns the necessity to consider CPA jointly with risk management. The second effect concerns the specificity of profit generation at considered credit portfolios. The specificity is a (large) overpayment of some categories of borrowers. The third effect shows itself through the positive correlation between risk and profits.

The paper devotes to presentation customer profitability analysis for nonbanking lenders. Detailed consideration of joined together CPA and risk-management provided economic logic for creation optimal strategy.

2. Materials and methods

2.1. Literature review

The problem of customer profitability analysis has been studied by many scientists. In particular, Pobrić [2] investigates methods of measuring customer profitability under different views. It considers customer profitability as customer group profitability. Storbacka [3] focused on existing client segmentation as a valuable marketing approach. Anandanatarajan [4], considers CRM as the process of acquiring, satisfying, retaining, and growing profitable customers.

A lot of papers are devoted to lending in the aspect of financial market development. Patalano & Roulet substantiate the increase in the level of public and corporate debt and the scale of global credit markets, which is caused by non-bank financial institutions [5]. Chernenko et al. analyzed a random sample of the credit market during

2010-2015 and showed that 32% of all loans were provided by nonbanking institutions [6]. Credit cyclicality for banks and non-banking institutions is studied by Fleckenstein et al. [7]. Kondova & Bandyopadhyay discussed a nonbank lending impact on bank efficiency [8]. The asset pricing model with both bank and non-bank financial institutions was simulated by d'Avernas et al. [8]. Distinctions of dealing with information scarcity between the bank and nonbank financial institutions were analyzed by Han [10]. Eichholtz et al. investigated the influence of local information on pricing for banks and non-banks [11].

Experience of nonbanking lending from different countries presents by the next authors: Bédard-Pagé [12] – Canada, Lee [13] – Korea, Rateiwa & Aziakpono [14] – Egypt, Nigeria, and South Africa, Vasileva [15] – Bulgaria, Soukal et al. [16] – Czech Republic, Sanfilippo-Azofra et al. [17] – Asia and Latin America.

Various theoretical aspects of the use of machine learning in the financial field are studied in the scientific literature. Mathur [18] presented the overview of machine learning in finance. Different aspects of modeling in finance discussed by Damodaran et al. [19], Guryanova et al. [20], Derbentsev et al. [21], Kuzmenko et al. [22], Kiv et al. [23], Sova & Lukianenko [24].

Machine learning approach for credit market modeling and its risk assessment presented in Manasov & Ivanovska [25], Pokorná & Sponer [26], Liberman et al. [27], Babenko et al. [28], Agarwal et al. [29], Venegas [30], Nyangena [31], Papouskova & Hajek [32].

Machine learning algorithms are effectively used to work with the customer base; in particular, clustering methods that use different types of data and make it possible to divide customers into groups/segments and develop an individual proposal for each group. Machine learning application for customer segmentation used in Monil et al. [33], Costea & Bleotu [34], Cuadros-Solas & Rodríguez-Fernández [35].

2.2. Research methodology

The study used two methodologies as a basis. The first in our study is the customer profitability analysis methodology. The methodological approaches of this analysis are widely used in the corporate segment. It is especially effective in cases where the corporation has many products in various segments. At the same time, the main approach of the analysis is focused on considering the income from sales/services provided in the context of the product line and/or segments. In assessing revenues, such parameters as marketing costs, revenues in absolute terms, expenses for subsequent servicing of products sold, the frequency of repeated sales to customers, and others are usually considered. The result of this approach is clustering by the profitability of products and segments. One submission of such clustering is ABC / XYZ structuration by Noche [36]. More detailed of this that ABC focuses on generating income by clusters and XYZ focuses on stability. ABC analysis is based on the classic Pareto rule (20% - 80%) and deals with the share of income in the general ledger. XYZ can be assessed on the basis of risk measures related to the variability of income stream (Table 1).

The key methodological principle of analysis is to create matrix 3X3. The analysis involves assessing profit in each cell of the matrix and leads to elaboration customer relations management for each cell (it can be range from disposal strategy to active growth).

Table 1. ABC/XYZ methodologies comparison

Focus on generating income	Focus on variation income`s stream
A – cluster, which generate 80% value (or close to it)	X – stable income stream
B – cluster which provide 15% value	Y – income stream with middle variation including seasonal fluctuations
C – cluster with only 5% value	Z – completely unstable income

One of the shortcomings of the abovementioned approach related to operations with “blocks” such as products, segments, or others. It may be lost differences in generating income by concrete clients at each cell of the matrix. This particular feature we have considered in our research because, as we illustrate below, borrowers essentially different by profitability. To implement this approach, all clients can be ordered in order of increasing profitability. Moving "left-to-right" along this ordering, you can calculate in cumulative terms the percentage of income that customers bring in relation to total income. The result will be the Whale curve. Typical examples of such curves are given in fig.2.

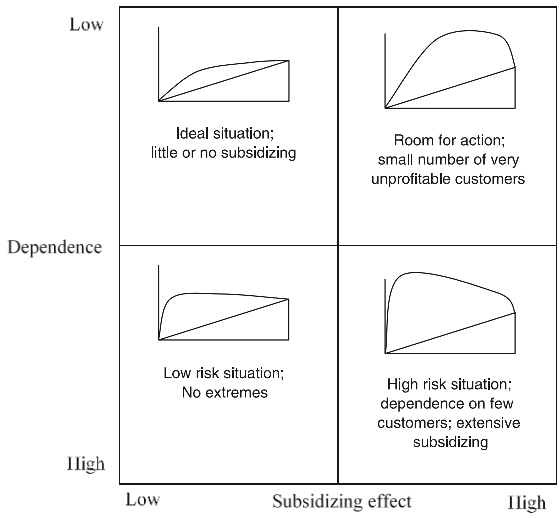

Fig. 2. Whale curves as illustration [Storbacka, 1998]

The second methodology that was used by us is the methodology for constructing risk management used in consumer lending [37], [38]. In a generalized form, it is shown in fig.3.

Fig. 3. Constructing risk management methodology

The methodology includes the following blocks. The first block is the risk analysis of the incoming flow. It can be assessed based on the scoring of the credit bureaus. Then it is possible to get a comparison of the risk level of the incoming flow to the lender and the market as a whole. Also, you can compare the incoming flows by the channels of attracting borrowers and products proposed by lenders. The second block includes checking the borrower against various databases (for example, the database of lost and stolen passports). The third block includes the most dynamically developing borrower appraisal system. The most common model combines credit bureau scoring with application scoring. The combination can be both in matrix form and in the form of a single scoring, which has both application and behavioral characteristics. Monitoring is an important component, which in the segment of short-term loans shows the likelihood of prolongation or obtaining a second loan. The final in the given scheme is the collection activity.

The combined use of customer profitability analysis methodological approaches and credit risk management allowed us to obtain several results presented below.

3. Results and discussion

Our study was based on data on loans issued by several nonbanker lenders in segment Pay-day-loans as on-line as off-line. As part of the initial analysis, we examined a period of one year and estimated the income that clients brought on the borrowing. Credit relations in this segment have specific features in comparison with the banking segment. One of these features is the frequency of cases when the borrower overpays on the loan. This happens for several reasons. In this segment, there is a large percentage of borrowers who have a high risk and quite often experience problems with loan payments. As a result, they use loan rollover, during which they pay interest. Also, in case of delay, they must pay fines and commissions. Thus, in this segment, some borrowers overpay the loan amount several times. At the same time, there is a fairly large number of borrowers who do not pay on loans. A typical dependence of profitability is shown in fig. 4.

The main difference from the banking segment is the increase in the graph on the left. In the banking segment, this curve on the left is flat.

Considering from the point of view customer profitability analysis it is logical to combine profitability analysis with risk analysis. Because here we have high risk and correspondingly high return.

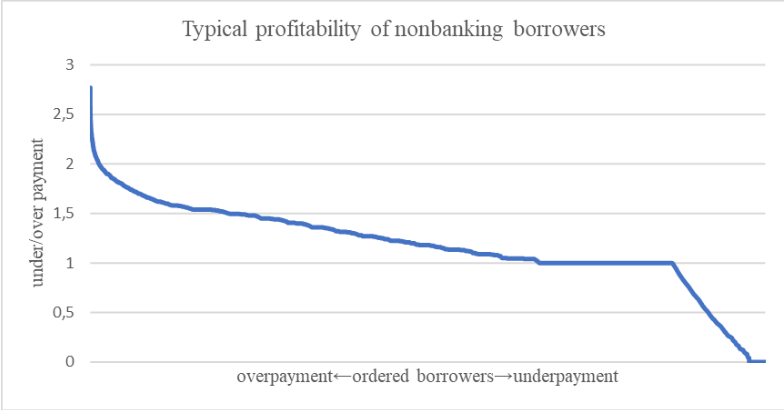

Fig. 4. Typical profitability of nonbanking borrowers

First, our approach includes the expansion of clusters, based on the Whale curve. Analysis of payments specificity we grounded to divide borrowers into 4 clusters: A, B, C, D. Cluster A corresponds to the most profitable customers which provide 100%. Our research confirms the classical Pareto principle 20%:80%. Approximately 20% of borrowers generate 100% of the profit. But of course, in reality it may be quite small percentage of borrowers which generate 100%, as example, 15%, 10% or sometimes 3%. Cluster B involves borrowers who tend to pay "fair and square". There is not so much overpayment. Borrowers from these two clusters are used as "Good" in the credit risk modeling (especially in credit scoring construction).

Borrowers from clusters C and D are typically considered as "Bad" in credit risk management. Our approach includes separation borrowers which pay something (they are in cluster C) and borrowers who "had not paid a penny" (they are is cluster D). The logic of such clustering is effective for two reasons. The first reason is that borrowers from cluster C "better" because they seek to pay. Borrowers from cluster D didn't have any plans to pay. The second reason lies in the strategy of working with borrowers from these clusters. Borrowers which have characteristics D should be carefully rejected at the application stage. The strategy of improving collection procedures should be applied to borrowers from C. Our researches demonstrate the level of recovery of 40%-60% for the borrower from cluster C.

The illustration of our clustering approach for analyse credit portfolio of non-bank lenders is presented in fig. 5.

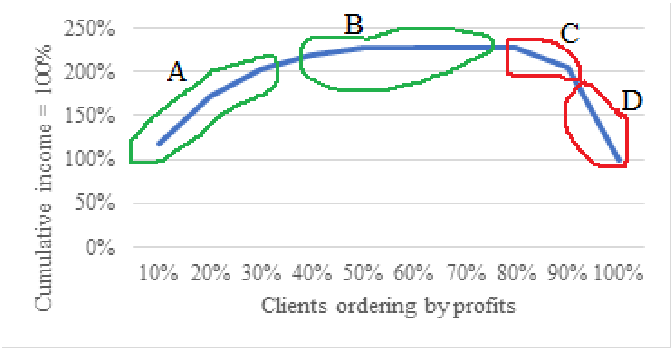

Fig. 5. Clustering credit portfolio for non-bank lender

The basic result of our research lies in linking risk assessment and profitability measurement of nonbanking borrowers (major payday loan borrowers). This result is presented in Table 2. Initially, we have structured borrowers with the help of credit scoring. It was application scoring of lenders which involved as characteristic parameters from borrower`s credit histories (collected by credit histories bureau). Scoring estimates borrower from 0 (high risk) and low risk (1000). There were segmentation borrowers into risk classes presented in Table 2 (with step 100 scores, vertical columns). The cut-off of applied scoring was 400 scores, so there was consideration of borrowers from segment. The bad rate curve is the indispensable part of any scoring. This curve indicates % of "Bads" (C+D) at the borrowers of the corresponding risk class.

Table 2. Correspondence between risk classes and clusters A, B, C, and D

A	Cut-off zone	18,7%	22,0%	20,9%	17,4%	15,7%	11,3%
B		58,1%	65,2%	71,7%	74,8%	75,9%	88,7%
C		19,7%	10,9%	7,7%	5,7%	5,0%	1,5%
D		3,5%	1,9%	3,3%	2,2%	1,8%	1,5%
Net Profit per Borrower, UAH		250	602	647	553	466	429
Bad Rate		23,2%	12,9%	11,0%	7,8%	6,8%	3,0%
Risk classes	≤400	(400-500]	(500-600]	(600-700]	(700-800]	(800-900]	(900-1000]

After that, we estimated the percentage ratio of borrowers from clusters A, B, C, and D into scoring classes. Results are provided in horizontal rows. What are the main results? Borrowers from cluster A constitute a higher percentage ratio at the riskier classes! The percentage ratio of A to the classes of good borrowers is lower.

Borrowers from cluster B demonstrate a monotonic increasing of percentage ratio from high-risk classes to low-risk classes.

Borrowers from cluster D demonstrate a monotonic decreasing of percentage ratio from high-risk classes to low-risk classes. This is natural, but changes are not so much as for C. Percentage ratio of C essentially decreases in this direction.

The main economic effect that was identified in our research is the following. Increasing risk of the borrower interconnects with higher percentage ratio of borrowers from cluster A (which are overpayment). This led to the financial objective: it is important to find optimal correspondence between risk and payments from borrower who essentially overpaid. In other words, it logically finds maximum output from borrowers from different scoring classes.

Analysis of indicated financial objectives necessitates estimation of net profit from one borrower (average) from different risk classes. The results can see in fig. 6.

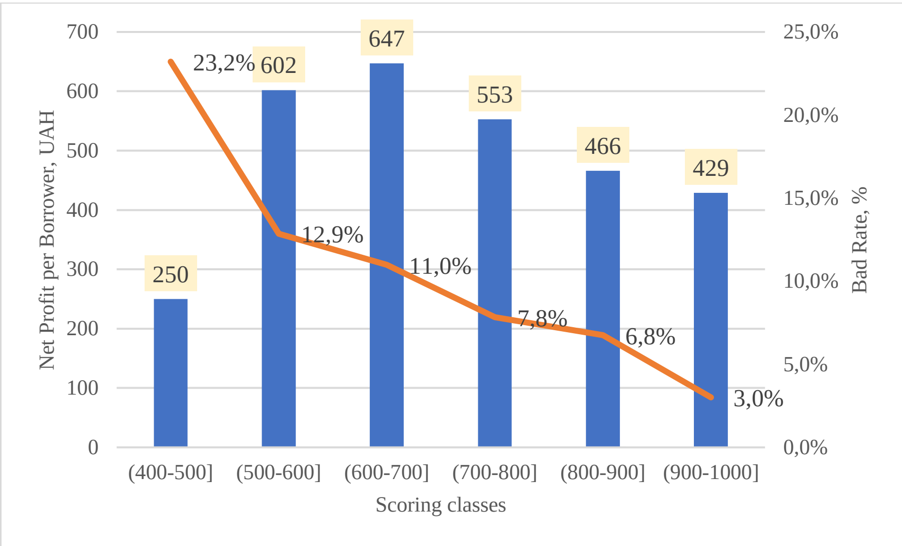

Fig. 6. Net profit per one borrower via credit scoring

The basic conclusion of credit portfolio profit analysis grounded by A, B, C, D clustering leads to the following specificity. Borrowers with high scoring generate not so many net profits. Borrowers with more risky scores generate maximum net profits. Risk classes with low scores (but on the right side from cut-off) demonstrate decreasing in net profits. Because they already include many bad borrowers and overpayments do not cover losses from them. The basic conclusion lies in the fact that most profitable borrowers are involved in more risky classes where overpayments cover losses in maximum.

Another clustering approach that was applied in our research grounded on fuzzy clustering. Fuzzy clustering characterizes by the property that data points can be involved in different clusters. The abovementioned analysis demonstrates that borrowers from the average risk scoring class can generate losses and overpayments. This is one of the basic reasons to choose fuzzy clustering instead of classical “hard clustering”.

It was choosing three indicators for run fuzzy clustering procedures. First is the score of borrowers which indicates risk level. The second indicator corresponds to the net profit level. The third indicator is loan amounts.

The applying of package "ppclust" from R demonstrated the following clustering (fig. 7).

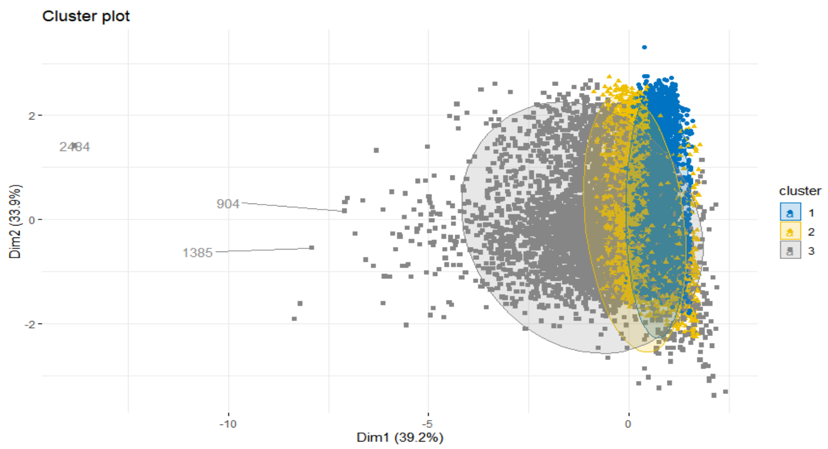

Fig. 7. Results of fuzzy clustering applications

The cluster average characteristics are presented in Table 3.

Table 3. Clustering results

	Size of clusters	Scoring values	Loan amount (UAH)	Net profit (UAH)
Cluster 1	52,08%	610,33	1564,03	270,91
Cluster 2	34,69%	602,90	3428,81	500,69
Cluster 3	13,23%	603,31	5914,17	1813,39

Clusters were also analyzed by cross-sections, which provides a more deeply looking inside (fig.8).

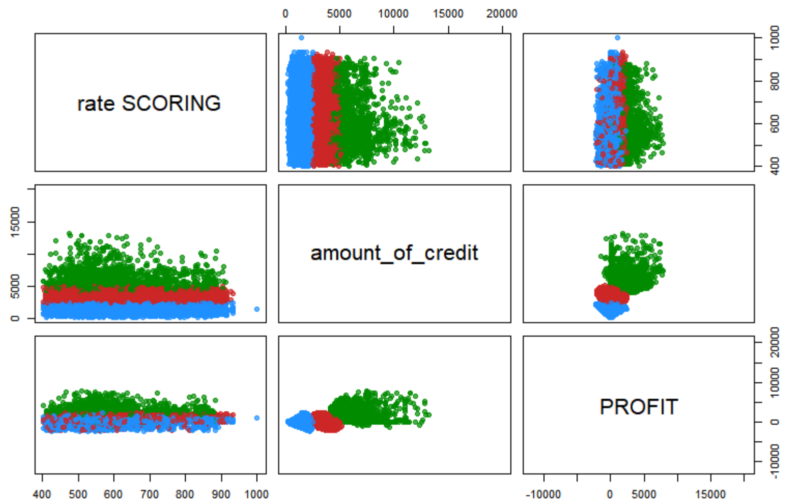

Fig. 8. Clustering results by using pairs of the features

One of the business strategies to develop lending in the considered sphere may be realized by combining risk estimation and net profit-generating by the borrower. Typical high net profit generates borrowers with recurrence loan receive. So, the focus may be concentrated to develop CRM with such a category.

4. Conclusions

Our research identifies the frameworks for customer profitability analysis for borrowers of nonbanking financial institutions. The nature of such borrowers includes high credit risk and at the same time high profit from effect "overpayment". Overpayment has a positive correlation with risk (negative correlation with score value). This provides a new approach to the CPM. It is logical to focus marketing efforts on borrowers from cluster A in parallel with risk assessment. CPM with borrowers from cluster B leads to strategy increase profit. "Bad" borrowers we proposed to divide into clusters C and D. Here should apply strong risk rules for cutting of D borrowers at the application stage.

The prospects of development CPM for this framework we find in the constructing system of estimation borrowers at the application stage. The estimation supposes to assess the probability that the borrower belongs to cluster A, B, C, or D.

Fuzzy clustering which was applied demonstrates differences of clusters. Differences focus on loan amounts and net profits.

One of the crucial applications proposed in article clustering can be used for forming business strategies to lend clients of non-bank lenders. This direction is part of our plans for researches.

References

1. In the Game: Traditional Financial Institutions Embrace Fintech Disruption. Harvard Business Review Analytic Service, 2019. Retrieved from https://hbr.org/resources/pdfs/comm/mastercard/Fintech.pdf.
2. Pobrić, A. (2014). Measuring customer profitability: The applicability of different concepts in practice. Ekonomika preduzeća, 62(3-4), 187-200.
3. Storbacka, K. (1997). Segmentation based on customer profitability—retrospective analysis of retail bank customer bases. Journal of Marketing Management, 13(5), 479-492.
4. Anandanatarajan, D. K. (2019). Customer Reationship Management–A Strategic Tool for Marketing. IJRAR Volume 6, Issue 2.
5. Patalano, R., & Roulet, C. (2020). Structural developments in global financial intermediation: The rise of debt and non-bank credit intermediation.
6. Chernenko, S., Erel, I., & Prilmeier, R. (2019). Nonbank lending. National Bureau of Economic Research.
7. Fleckenstein, Q., Gopal, M., Gutierrez Gallardo, G., & Hillenbrand, S. (2020). Nonbank Lending and Credit Cyclicality. NYU Stern School of Business.
8. Kondova, G., & Bandyopadhyay, T. (2019). The Impact of Non-bank Lending on Bank Efficiency: Data Envelopment Analysis of European Banks. International Journal of Trade, Economics and Finance, 10(5), 108-112.
9. d'Avernas, A., Vandeweyer, Q., & Darracq-Pariès, M. (2020). The growth of non-bank finance and new monetary policy tools. Research Bulletin, 69.
10. Han, J. H. (2017). Does Lending by banks and non-banks differ? Evidence from small business financing. Banks & bank systems, (12, № 4), 98-104.
11. Eichholtz, P., Mimiroglu, N., Ongena, S., & Yönder, E. (2020). Banks, Non-Banks, and the Incorporation of Local Information in CMBS Loan Pricing. Swiss Finance Institute Research Paper, (19-58).
12. Bédard-Pagé, G. (2019). Non-bank financial intermediation in Canada: An update (No. 2019-2). Bank of Canada Staff Discussion Paper.
13. Lee, M. (2018). Non-Bank Lending to Firms: Evidence from Korean Firm-Level Data. The Journal of Industrial Distribution & Business, 9(9), 15-23.
14. Rateiwa, R., & Aziakpono, M. J. (2017). Non-bank financial institutions and economic growth: Evidence from Africa's three largest economies. South African Journal of Economic and Management Sciences, 20(1), 1-11.
15. Vasileva, V. (2019). Development Of Consumer Lending By Non-Bank Credit Companies In Bulgaria. Народностопански архив, (1), 65-76.
16. Soukal, I., Hamplová, E., & Haviger, J. (2021). Effectiveness of Regulation of Educational Requirements for Non-Bank Credit Providers in Czech Republic. Social Sciences, 10(1), 28.
17. Sanfilippo-Azofra, S., Torre-Olmo, B., & Cantero-Saiz, M. (2019). Microfinance institutions and the bank lending channel in Asia and Latin America. Journal of Asian Economics, 63, 19-32.
18. Mathur, P. (2019). Overview of machine learning in finance. In Machine Learning Applications Using Python (pp. 259-270). Apress, Berkeley, CA.
19. Damodaran, S., Kavin, S., Keerthi, K. U., Madhumathi, J., & Mythili, P. V. (2019, November). Empowering MSMEs Through Digital Lending. In 2019 International Conference on Digitization (ICD) (pp. 249-253). IEEE.
20. Guryanova, L., Yatsenko, R., Dubrovina, N., Babenko, V. (2020). Machine learning methods and models, predictive analytics and applications. CEUR Workshop Proceedings, 2649, pp. 1–5.

21. Derbentsev, V., Matviychuk, A., Datsenko, N., Bezkorovainyi, V., Azaryan, A. (2020). Machine learning approaches for financial time series forecasting. CEUR Workshop Proceedings, 2713, pp. 434–450.
22. Kuzmenko, O., Šuleř, P., Lyeonov, S., Judrupa, I., Boiko, A. (2020) Data mining and bifurcation analysis of the risk of money laundering with the involvement of financial institutions. Journal of International Studies, 13(3), pp. 332–339.
23. Kiv, A., Hryhoruk, P., Khvostina, I., Solovieva, V., Soloviev, V., & Semerikov, S. (2020). Machine learning of emerging markets in pandemic times. CEUR Workshop Proceedings, 2713, pp. 1–20.
24. Sova, Y., & Lukianenko, I. (2020, September). Theoretical and Empirical Analysis of the Relationship Between Monetary Policy and Stock Market Indices. In 2020 10th International Conference on Advanced Computer Information Technologies (ACIT) (pp. 708-711). IEEE.
25. Manasov, J., & Ivanovska, L. P. (2018). User preferences for banking services offered by non-banking companies and tech giants. Journal of sustainable development, 8(20), 35-50.
26. Pokorná, M., & Sponer, M. (2016). Social lending and its risks. Procedia-Social and Behavioral Sciences, 220, 330-337.
27. Liberman, A., Neilson, C., Opazo, L., & Zimmerman, S. (2018). The equilibrium effects of information deletion: Evidence from consumer credit markets (No. w25097). National Bureau of Economic Research.
28. Babenko, V., Panchyshyn, A., Zomchak, L., Nehrey, M., Artym-Drohomyretska, Z., Lahotskyi, T. (2021). Classical Machine Learning Methods in Economics Research: Macro and Micro Level Example. WSEAS Transactions on Business and Economics, Vol. 18, 2021, Art. #22, pp. 209-217. https://doi.org/10.37394/23207.2021.18.22.
29. Agarwal, S., Alok, S., Ghosh, P., & Gupta, S. (2020). Financial Inclusion and Alternate Credit Scoring for the Millennials: Role of Big Data and Machine Learning in Fintech. Working Paper, National University of Singapore.
30. Venegas, P. (2018). Risk scoring for non-bank financial institutions. Available at SSRN 3280738.
31. Nyangena, B. O. (2019). Consumer credit risk modelling using machine learning algorithms: a comparative approach (Doctoral dissertation, Strathmore University).
32. Papouskova, M., & Hajek, P. (2019). Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decision support systems, 118, 33-45.
33. Monil, P., Darshan, P., Jecky, R., Vimarsh, C., & Bhatt, B. R. (2020). Customer Segmentation using machine learning International. Journal for Research in Applied Science & Engineering Technology. Volume 8 Issue VI, 2104-2108.
34. Costea, A., & Bleotu, V. (2012). A new fuzzy clustering algorithm for evaluating the performance of non-banking financial institutions in Romania. Economic Computation and Economic Cybernetics Studies and Research, 46(4), 179-199.
35. Cuadros-Solas, J., & Rodríguez-Fernández, F. (2019). A Machine Learning Approach to the Digitalization of Bank Customers: Evidence from Random and Causal Forests.
36. Noche, B. (2014). ABC-/XYZ Analysis Introduction. Universitat Duisburg Essen. Duisburg: Bernd Noche. Retrieved from https://www. unidue. de/imperia/md/content/tul/download/en_ss2015_lm01_le_abc_analysis. pdf. 2014.
37. Kaminskyi, A., Pysanets, K. (2017). Audit of risk management system in consumer lending. Journal Association 1901 "SEPIKE". 18 Edition, 133-140.
38. Kaminskyi, A., Motoryn, R., Pysanets, K. (2017). The effectiveness of the use statistical data of credit histories bureaus in risk management systems. Probability in action Vol. 3, 139-156.