Al-Kharaj: Journal of Islamic Economic and Business
Volume 8 (1), 2026
P-ISSN: 2686-262X; E-ISSN: 2685-9300

Analysis Of Online Loan Regional Clustering in Indonesia in 2024 Based On Outstanding And Default Rate (TWP90) Using K-Means Clustering

Fina Sherli Wewengkang[1], Arief Wibowo[2]
[1,2] Budi Luhur University
Email: 2531600233@student.budiluhur.ac.id, arief.wibowo@budiluhur.ac.id

Keywords:
Online Loans, K-Means Clustering, Default Risk, Outstanding, TWP90

Abstract

The increase in online lending distribution in Indonesia in 2024 was not accompanied by a uniform level of credit risk across regions. This study aims to categorize online lending regions in Indonesia based on outstanding values and 90-day default rates (TWP90) using a quantitative approach based on the K-Means algorithm. Secondary data from all provinces was analyzed using RapidMiner and evaluated using the Davies–Bouldin Index (DBI). The test results showed a DBI of 0.746 at K=2, 0.376 at K=3, and 0.564 at K=4. Although K=2 yielded the lowest DBI, the K=3 model was chosen because it provided a more informative and policy-relevant risk classification. The clustering resulted in three risk clusters: Low Risk, with outstanding values and TWP90 below average; Medium Risk, with values above average; and High Risk, characterized by a very high TWP90 level despite relatively low outstanding values. These findings confirm the effectiveness of K-Means in mapping online lending risks based on regions and support more precise credit monitoring. Keywords: online loans, K-Means clustering, default risk, outstanding, TWP90.

INTRODUCTION

The development of online lending services in Indonesia has experienced rapid growth in line with the increasing use of digital technology in the financial sector. This digital transformation has made financing easier for the public, but on the other hand, it also poses various risks, particularly the risk of default, which varies across regions. According to the Financial Services Authority (2021), the distribution of online loans does not show a uniform pattern, either in terms of outstanding value or default rates, posing challenges for national risk management and oversight. This variation indicates differences in economic characteristics and borrower behavior that require more in-depth analysis.

The risk of default in online loans is a critical issue because it directly impacts the stability of the financial system and the sustainability of the fintech industry. Li et al. (2020) stated that the high default rate on digital lending platforms is generally influenced by heterogeneity in regional characteristics and weak data-based risk mapping. This finding is reinforced by research by Zhang and Chen (2021), which confirms that inaccurately categorizing regions by risk level can lead to ineffective mitigation policies and potentially increase financial losses. Therefore, region-based risk mapping is a strategic step to understand default patterns more comprehensively.

In the context of online loan risk analysis, the use of quantitative data is a relevant approach because it can objectively describe empirical conditions. According to Hair et al. (2020), quantitative approaches are highly effective for identifying complex data patterns and structures, especially when involving large-scale numerical variables. The variables of online loan outstanding value and 90-day default rate (TWP90) are considered representative indicators for reflecting the magnitude of credit exposure and the level of default risk in a region. Research by Khandani et al.

(2021) shows that the combination of loan value and late payment ratio indicators provides a more accurate picture of risk than using a single variable.

Clustering-based data mining approaches are increasingly used in financial risk analysis due to their ability to group data without requiring initial labels. According to Han et al. (2022), clustering techniques allow researchers to objectively identify groups of data with high levels of similarity, making them particularly suitable for exploratory analysis of regional financial data. The application of clustering in online loan analysis is also considered effective for mapping regions with similar risk characteristics, as noted by Rahman et al. (2023), which emphasizes that data-driven clustering can assist regulators and stakeholders in formulating more targeted supervisory strategies.

The K-Means algorithm is one of the most widely used non-hierarchical clustering methods in numerical data analysis due to its simplicity and efficiency. According to Jain (2020), K-Means has good computational capabilities for handling large-scale datasets and producing easily interpretable clusters. In the context of loan risk analysis, K-Means is considered capable of consistently grouping regions based on the proximity of financial characteristics, thus supporting more systematic risk mapping. However, the quality of the clustering results needs to be evaluated to ensure that the resulting clusters truly reflect the optimal data structure.

Cluster quality evaluation is a crucial step in clustering analysis to ensure the reliability of the clustering results. The Davies–Bouldin Index (DBI) is a widely used internal validation index because it measures the balance between intra-cluster cohesion and inter-cluster separation. According to Arbelaitz et al. (2021), the DBI provides an effective measure of how cohesive and separate the resulting clusters are, with a lower DBI value indicating better cluster quality. Singh and Yadav (2022) also recommend using the DBI in financial risk analysis because it supports the selection of the most representative clustering model for the analyzed data.

Based on this background, this study focuses on clustering regions in Indonesia based on outstanding online loan values and 90-day default rates (TWP90) using a clustering-based data mining approach. This study aims to identify regional groups with different risk levels: Low Risk, Medium Risk, and High Risk, thus providing a more structured picture of the distribution of online loan risk in Indonesia in 2024.

METHODS

This study applies a quantitative method using a clustering-based data mining approach to assess online lending risk in Indonesia in 2024. The research begins by identifying issues related to variations in online loan default rates across different regions, followed by establishing research objectives aimed at grouping regions based on their outstanding loan values and 90-day default rates (TWP90). Data collection is carried out to provide the basis for analysis, after which the data is structured and processed to prepare it for quantitative analysis. The subsequent analysis employs the K-Means algorithm to create clusters of regions with similar risk characteristics, and the quality of these clusters is evaluated using the Davies–Bouldin Index to ensure their reliability. The final stage of the research focuses on interpreting the clustering results to categorize regions into Low, Medium, and High Risk groups, as well as formulating conclusions and policy implications based on the findings. All research stages are conducted using RapidMiner software.

The data used in this study consists of secondary data obtained from official publications by the Financial Services Authority (OJK) regarding online loans in Indonesia. The research variables include the outstanding online loan value and the 90-day default rate at the provincial level, representing the total amount of outstanding loans and the level of default risk, respectively. The data is compiled into a numeric dataset, with its format adjusted for further quantitative processing. Data processing ensures that each variable accurately and proportionally represents the risk characteristics of online loans, preparing the dataset for clustering analysis. Data selection was performed to identify the most relevant and representative variables aligned with the research objectives. The outstanding loan variable reflects the exposure of each region to digital credit

distribution, while the 90-day default rate serves as the primary indicator of default risk, showing the proportion of loans that are delayed in repayment. Variables unrelated to the objective of grouping regions by risk were excluded to avoid distorting the analysis and to ensure that the clustering process generates homogeneous and meaningful regional groups.

Data cleaning was conducted to guarantee the quality and reliability of the dataset before analysis. This stage involved removing duplicate records, checking for data completeness, and standardizing numerical formats to maintain consistency across all variables. Invalid or inconsistent data was also corrected to prevent disruption during the analysis. The aim of data cleaning was to minimize errors and biases, ensuring that the resulting clusters accurately and objectively reflect the risk conditions of online loans. The dataset then underwent transformation through normalization to place all variables on a comparable scale before clustering. Normalization was necessary because differences in the value ranges of the outstanding loan and 90-day default rate variables could cause one variable to disproportionately influence the clustering process. Normalization ensures that each variable contributes proportionally, resulting in objective clustering outcomes that better reflect the actual risk conditions of online loans.

A data mining approach was applied to extract hidden patterns and structures from the large and complex online loan dataset. This approach goes beyond describing relationships between variables, allowing the identification of regional grouping patterns based on the similarity of financial characteristics. The use of clustering techniques in data mining provides an objective and systematic method for identifying regional groups with relatively homogeneous online loan risk profiles, forming a strong analytical basis for risk mapping and evidence-based decision-making. Clustering, as an unsupervised learning technique, groups data based on similarity without relying on predefined labels. This method allows the identification of natural structures within the data, particularly when relationships between variables are not fully understood. In this study, a non-hierarchical clustering method was used because it efficiently handles large-scale numerical data and offers strong computational performance. Non-hierarchical clustering, particularly K-Means, is appropriate for exploratory analysis at the national level, producing stable and interpretable clusters based on the characteristics of the analyzed dataset.

The K-Means algorithm was employed to divide the dataset into clusters, aiming to group regions with similar risk profiles. The process begins by initializing centroids, after which each data point is assigned to the nearest centroid. The centroid positions are then recalculated based on the mean of the assigned points. This iterative process continues until the clusters reach stability, ensuring high internal similarity within clusters and clear distinctions between clusters. K-Means is suitable for this study because it allows for efficient, objective, and interpretable clustering of quantitative variables on a national scale. To evaluate the quality of the clusters, the Davies–Bouldin Index (DBI) was used, which assesses the balance between intra-cluster cohesion and inter-cluster separation. The DBI considers how closely data points within a cluster are grouped and how well-separated clusters are from one another. A lower DBI value indicates better cluster quality, reflecting more compact clusters internally and clearer separation between clusters. By using the DBI, the study ensures that the clusters derived from online loan data accurately represent risk levels and are suitable for further analysis and policy recommendations.

RESULTS AND DISCUSSION

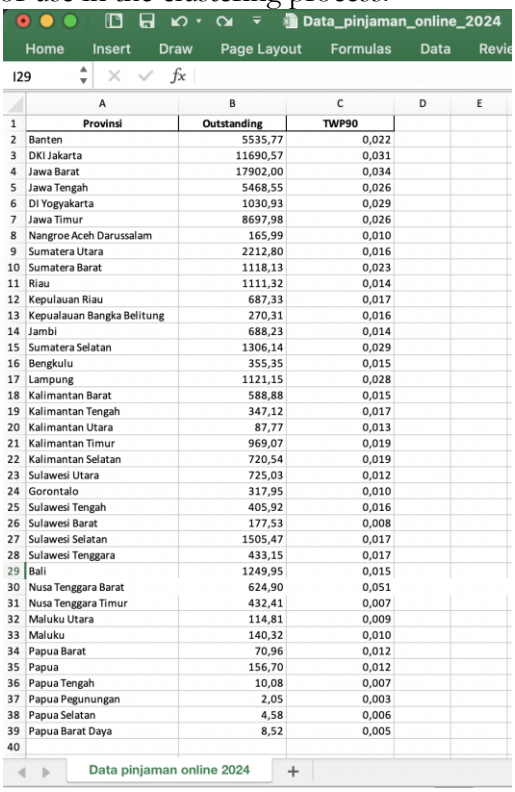
Data Pre-Processing

Data pre-processing is the initial step in this research, aimed at ensuring data quality before the clustering process. The data used is online loan data from 2024, consisting of two main attributes: the outstanding online loan value and the 90-day default rate (TWP90) in each province in Indonesia. This stage includes attribute selection, data completeness checks, and data

inconsistencies. This process is necessary to ensure the data is fully prepared for analysis using the K-Means algorithm and to avoid errors in distance calculations.

Data Preparation

The data preparation stage is carried out after the data is declared clean. At this stage, data transformation and normalization are carried out to ensure each attribute has a comparable scale. Normalization is necessary because differences in the value range between outstanding online loans and TWP90 can affect clustering results. With normalization, each attribute has a proportional contribution to the Euclidean distance calculation. The result of this stage is a standardized dataset ready for use in the clustering process.


Data_pinjaman_online_2024

Home Insert Draw Page Layout Formulas Data Revie

I29

	A	B	C	D	E
1	Provinsi	Outstanding	TWP90		
2	Banten	5535,77	0,022		
3	DKI Jakarta	11690,57	0,031		
4	Jawa Barat	17902,00	0,034		
5	Jawa Tengah	5468,55	0,026		
6	DI Yogyakarta	1030,93	0,029		
7	Jawa Timur	8697,98	0,026		
8	Nangroe Aceh Darussalam	165,99	0,010		
9	Sumatera Utara	2212,80	0,016		
10	Sumatera Barat	1118,13	0,023		
11	Riau	1111,32	0,014		
12	Kepulauan Riau	687,33	0,017		
13	Kepualauan Bangka Belitung	270,31	0,016		
14	Jambi	688,23	0,014		
15	Sumatera Selatan	1306,14	0,029		
16	Bengkulu	355,35	0,015		
17	Lampung	1121,15	0,028		
18	Kalimantan Barat	588,88	0,015		
19	Kalimantan Tengah	347,12	0,017		
20	Kalimantan Utara	87,77	0,013		
21	Kalimantan Timur	969,07	0,019		
22	Kalimantan Selatan	720,54	0,019		
23	Sulawesi Utara	725,03	0,012		
24	Gorontalo	317,95	0,010		
25	Sulawesi Tengah	405,92	0,016		
26	Sulawesi Barat	177,53	0,008		
27	Sulawesi Selatan	1505,47	0,017		
28	Sulawesi Tenggara	433,15	0,017		
29	Bali	1249,95	0,015		
30	Nusa Tenggara Barat	624,90	0,051		
31	Nusa Tenggara Timur	432,41	0,007		
32	Maluku Utara	114,81	0,009		
33	Maluku	140,32	0,010		
34	Papua Barat	70,96	0,012		
35	Papua	156,70	0,012		
36	Papua Tengah	10,08	0,007		
37	Papua Pegunungan	2,05	0,003		
38	Papua Selatan	4,58	0,006		
39	Papua Barat Daya	8,52	0,005		
40					

Data pinjaman online 2024

Figure 1. 2024 Online Loan Dataset in the Data Preparation Phase

Cluster Modeling

In this phase, the pre-processed and normalized data are then grouped into several clusters using the K-Means algorithm with the help of RapidMiner software. The clustering process is carried out iteratively by calculating the distance between data points from the cluster center (centroid) until optimal convergence is achieved. The number of clusters is first determined according to the research objectives, then each data point is allocated to the nearest cluster based on the Euclidean distance calculation. The clustering results are then evaluated using the Davies–Bouldin Index (DBI) to measure the quality of the formed clusters. The DBI value is used as an indicator of the level of intra-cluster cohesion and inter-cluster separation, where a lower DBI value indicates a better cluster structure and better representation of the characteristics of the analyzed data.

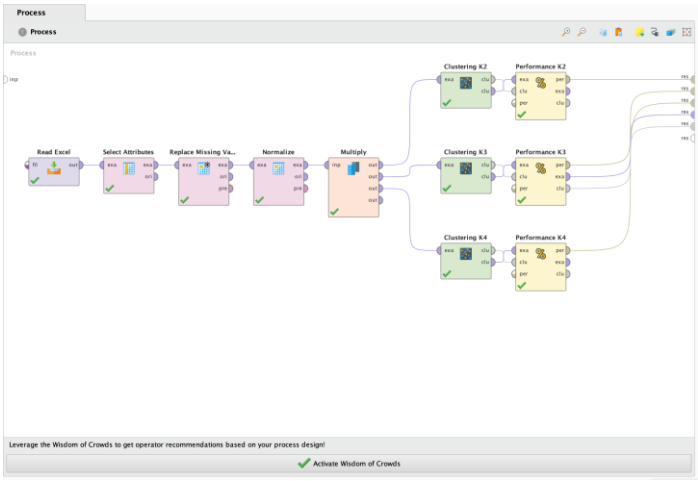


Figure 2. Clustering Process Using the K-Means Algorithm in RapidMiner

Figure 5 shows the flow of online loan data analysis using RapidMiner. The process begins with data reading using Read Excel, followed by Select Attributes to select the outstanding and TWP90 variables. Incomplete data was handled using Replace Missing Values, then normalized to ensure all variables were on a comparable scale. Next, the dataset was duplicated using Multiply to perform several clustering scenarios. The K-Means algorithm was applied with different numbers of clusters (K=2, K=3, and K=4), and the quality of each clustering result was evaluated using Performance (Clustering) based on the Davies–Bouldin Index to determine the optimal number of clusters.

Reporting Research Results

The next stage of this study focused on assessing the performance of the clustering results through the application of the Davies–Bouldin Index (DBI). This evaluative process is crucial for assessing the validity and coherence of the formed clusters, as well as testing the model's capability in optimally representing the data structure. In this phase, we determined the number of clusters to three: 2 clusters, 3 clusters, and 4 clusters. Based on the selected model, regions were then classified into Low Risk, Medium Risk, and High Risk categories. The modeling results were then presented in tables and thematic maps as a basis for interpreting online loan risk patterns across regions in Indonesia in 2024.

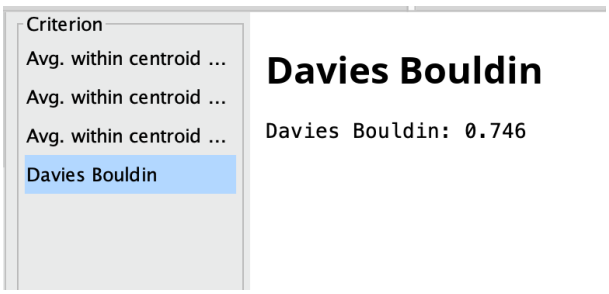


Figure 3. Trial Results on Performance K-2 with a total of 2 Clusters

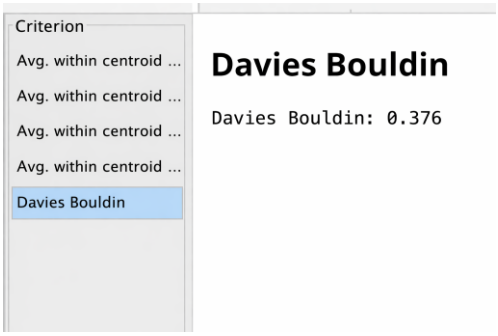


Figure 4. Trial Results on Performance K-3 with a total of 3 Clusters

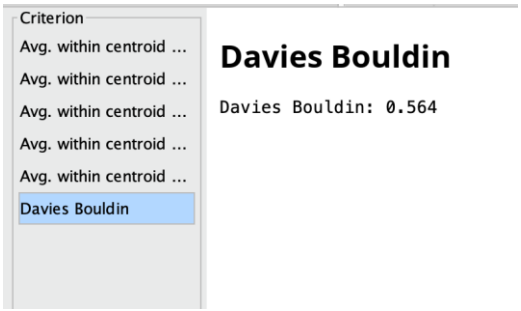


Figure 5. Trial Results on Performance K-4 with a total of 4 Clusters

Discussion

Based on performance evaluation using the Davies–Bouldin Index (DBI), the K=2 scenario produced the lowest DBI value, at 0.746, indicating statistically the best cluster quality due to its high intra-cluster cohesion and optimal inter-cluster separation. Meanwhile, the K=3 and K=4 scenarios produced DBI values of 0.376 and 0.564, respectively, indicating lower cluster quality than K=2. Nevertheless, in the context of mapping online loan risks for policy purposes, K=3 was still chosen because it was able to group regions into Low, Medium, and High Risk categories, thus providing a more informative and applicable interpretation even though it did not produce the lowest DBI value. The following illustrates the results of the Cluster Model.

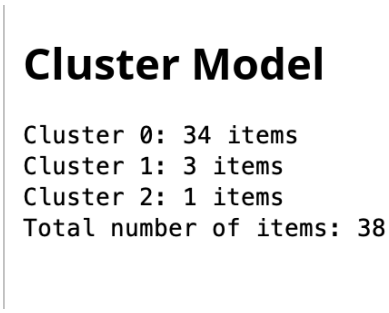


Figure 6. Results of the Cluster Model

Row No.	Provinsi	cluster	Outstanding	TWP90
1	Banten	cluster_0	1.026	0.555
2	DKI Jakarta	cluster_1	2.717	1.503
3	Jawa Barat	cluster_1	4.424	1.778
4	Jawa Tengah	cluster_0	1.008	0.918
5	DI Yogyakarta	cluster_0	-0.212	1.241
6	Jawa Timur	cluster_1	1.895	0.964
7	Nangroe Aceh Darussa...	cluster_0	-0.449	-0.768
8	Sumatera Utara	cluster_0	0.113	-0.071
9	Sumatera Barat	cluster_0	-0.188	0.599
10	Riau	cluster_0	-0.190	-0.315
11	Kepulauan Riau	cluster_0	-0.306	-0.035
12	Kepualauan Bangka Be...	cluster_0	-0.421	-0.116
13	Jambi	cluster_0	-0.306	-0.372
14	Sumatera Selatan	cluster_0	-0.136	1.282
15	Bengkulu	cluster_0	-0.397	-0.242
16	Lampung	cluster_0	-0.187	1.128
17	Kalimantan Barat	cluster_0	-0.333	-0.226
18	Kalimantan Tengah	cluster_0	-0.400	-0.017
19	Kalimantan Utara	cluster_0	-0.471	-0.438
20	Kalimantan Timur	cluster_0	-0.229	0.167
21	Kalimantan Selatan	cluster_0	-0.297	0.161
22	Sulawesi Utara	cluster_0	-0.296	-0.499
23	Gorontalo	cluster_0	-0.408	-0.777
24	Sulawesi Tengah	cluster_0	-0.383	-0.105
25	Sulawesi Barat	cluster_0	-0.446	-0.941

Row No.	Provinsi	cluster	Outstanding	TWP90
26	Sulawesi Selatan	cluster_0	-0.081	0.009
27	Sulawesi Tenggara	cluster_0	-0.376	-0.031
28	Bali	cluster_0	-0.152	-0.242
29	Nusa Tenggara Barat	cluster_2	-0.323	3.546
30	Nusa Tenggara Timur	cluster_0	-0.376	-1.064
31	Maluku Utara	cluster_0	-0.463	-0.806
32	Maluku	cluster_0	-0.456	-0.775
33	Papua Barat	cluster_0	-0.476	-0.559
34	Papua	cluster_0	-0.452	-0.541
35	Papua Tengah	cluster_0	-0.492	-1.036
36	Papua Pegunungan	cluster_0	-0.494	-1.440
37	Papua Selatan	cluster_0	-0.494	-1.177
38	Papua Barat Daya	cluster_0	-0.493	-1.257

Figure 7. Cluster Data View Results

Figure 10 shows the results of grouping online lending regions into three risk clusters. Cluster 0 (Low Risk) is dominated by provinces with relatively low outstanding and TWP90 values, indicating more stable online lending conditions and controlled default rates. Cluster 1 (Medium Risk) includes regions with higher outstanding and TWP90 values, reflecting increased loan exposure and potential default risk. Meanwhile, Cluster 2 (High Risk) is characterized by very high TWP90 levels despite relatively lower outstanding values, indicating a high default risk. This cluster division confirms the variation in online lending risk levels across provinces and serves as an important basis for region-based risk mapping and monitoring.

Attribute	cluster_0	cluster_1	cluster_2
Outstanding	-0.256	3.012	-0.323
TWP90	-0.229	1.415	3.546

Figure 8. Centroid Table

Figure 11 shows the centroid values resulting from the K-Means analysis with K=3, which represent the average characteristics of each cluster based on the outstanding and TWP90 variables in the form of standardized values (z-scores). Cluster_0 (Low Risk) has outstanding and TWP90 values below the average, indicating a region with relatively low loan levels and default risk. Cluster_1 (Medium Risk) is characterized by outstanding and TWP90 values above the average, reflecting a region with increased loan exposure and risk. Meanwhile, Cluster_2 (High Risk) shows a very high TWP90 value despite relatively low outstanding, indicating a high default risk. These centroid values are used as the basis for interpreting and labeling the risk levels of online loan regions.

Figure 9. Cluster Visualization Results

Based on the clustering map, the Low Risk category is dominated by most provinces in Indonesia, including Aceh, North Sumatra, West Sumatra, Riau, Riau Islands, Jambi, Bengkulu, Lampung, Central Java, DI Yogyakarta, most of Kalimantan, Sulawesi, Bali, East Nusa Tenggara, Maluku, and almost all of Papua. Medium Risk is seen in several provinces in Java and its surrounding areas, such as DKI Jakarta, West Java, East Java, and West Nusa Tenggara, which show relatively higher outstanding and TWP90 levels compared to other regions. Meanwhile, High Risk is concentrated in certain regions in eastern Indonesia, particularly East Nusa Tenggara (and other provinces according to the clustering results), which are characterized by very high 90-day default rates (TWP90) despite relatively lower outstanding values. This pattern indicates spatial differences in online loan risk between provinces that need to be considered in area-based supervision.

CONCLUSION

This study categorizes online lending regions in Indonesia in 2024 based on outstanding value and default rate (TWP90) using the K-Means Clustering algorithm. The analysis yields three regional risk categories: Low Risk, Medium Risk, and High Risk, demonstrating significantly different characteristics, thus supporting the research hypothesis. This mapping addresses the need for quantitative and spatial online lending risk analysis and reinforces previous theoretical and empirical findings that clustering methods are effective in identifying digital financial risk patterns. The academic contribution of this research lies in the application of K-Means in the context of region-based fintech risk management, while the practical results can inform the formulation of more precise supervisory policies.

Based on these findings, the recommended solution is the implementation of a region-based risk mitigation strategy, where regulators establish tiered supervision with a focus on stricter controls in medium- and high-risk regions through adjustments to credit disbursement limits and borrower evaluation standards. Online lending industry players need to integrate the results of the regional clustering into their risk management systems to adapt their offering strategies, interest

rates, and collection mechanisms. In addition, financial literacy programs targeted at high-risk areas need to be strengthened to reduce default rates, while further research is recommended to expand indicators and analysis methods so that online loan risk management policies can be formulated in a more comprehensive and sustainable manner.

REFERENCE

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2021). An extensive comparative study of cluster validity indices. *Pattern Recognition, 115,* 107870. https://doi.org/10.1016/j.patcog.2021.107870

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 54*(6), 627–635. https://doi.org/10.1057/palgrave.jors.2601545

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2020). *Multivariate data analysis* (8th ed.). Cengage Learning.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

Irawati, N., Prasetyo, E., & Hidayat, R. (2025). Penerapan algoritma K-Means clustering untuk pengelompokan wilayah berbasis indikator ekonomi. *Jurnal Ilmu Komputer dan Informatika, 10*(1), 45–54.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jain, A. K. (2020). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2019.09.011

Khandani, A. E., Kim, A. J., & Lo, A. W. (2021). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2021.03.001

Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis. *Knowledge-Based Systems, 56,* 1–13. https://doi.org/10.1016/j.knosys.2013.10.005

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update. *European Journal of Operational Research, 247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Li, Y., Wang, Y., & Zhao, Y. (2020). Credit risk assessment in peer-to-peer lending: A clustering-based approach. *Financial Innovation, 6*(1), 1–18. https://doi.org/10.1186/s40854-020-00190-3

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications: A decade review. *Expert Systems with Applications, 39*(12), 11303–11311. https://doi.org/10.1016/j.eswa.2012.02.063

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

OECD. (2020). *Consumer policy and fraud: Evidence-based policy responses.* OECD Publishing.

Otoritas Jasa Keuangan. (2021). *Statistik fintech lending Indonesia.* OJK.

Otoritas Jasa Keuangan. (2024). *Statistik fintech lending Indonesia.* OJK.

Rahman, M. A., Hasan, M. K., & Sarker, I. H. (2023). Machine learning-based regional financial risk profiling using clustering techniques. *Expert Systems with Applications, 213,* 118919. https://doi.org/10.1016/j.eswa.2022.118919

Singh, A., & Yadav, A. (2022). Evaluation of clustering techniques using Davies–Bouldin Index in financial risk analysis. *International Journal of Data Science and Analytics, 13*(2), 145–158. https://doi.org/10.1007/s41060-021-00290-4

Tang, H., & Liu, Y. (2020). Credit risk assessment of peer-to-peer lending using data mining techniques. *Journal of Risk and Financial Management, 13*(9), 207. https://doi.org/10.3390/jrfm13090207

Zhang, Y., & Chen, W. (2021). Regional credit risk classification based on unsupervised learning methods. *Journal of Risk and Financial Management, 14*(9), 421. https://doi.org/10.3390/jrfm14090421