Alacrity : Journal Of Education
e-ISSN : 2775-4138
Volume 5 Issue 3 Oktober 2025
The Alacrity : Journal Of education is published 4 times a year in **(February, June, October)**
Focus : Learning, Education, Including, Social, Curriculum, Management Science, Educational Philosophy And Educational Approaches.
LINK : http://lpppipublishing.com/index.php/alacrity

The Use of Elbow K-Means And K-Medoids in the Grouping of Provinces in Indonesia Based on the Indicators of the Effectiveness of the Authentication Taspen Application With DBI and Silhoutte Coefficients

Anisya Tasya[1], Gusmi Kholijah[2], Sarmada[3]
[1,2,3] *Universitas Jambi, Indonesia*

ARTICLE INFO
Article history:
Received
01 September 2025
Revised
25 September 2025
Accepted
10 Oktober 2025

ABSTRACT

This study aims to classify Indonesian provinces based on the effectiveness of using the Taspen Authentication Application and to compare the performance of the K-Means and K-Medoids clustering algorithms. The research employed a quantitative approach using secondary data derived from the Taspen Authentication metrics, which include ten variables such as user, session, retention, login effectiveness, churn rate, and conversion rate. The data from 38 provinces were analyzed using cluster analysis. The optimal number of clusters was determined using the Elbow Method, and validation was performed with the Davies-Bouldin Index (DBI). The results indicate that the K-Means algorithm provides better clustering performance, with a DBI value of 1.752 and a Silhouette Coefficient of 0.2850. The findings reveal that 50% (19 provinces) demonstrate high effectiveness, 13.16% (5 provinces) moderate effectiveness, and 36.84% (14 provinces) low effectiveness in using the application. These results can serve as a basis for PT TASPEN and policymakers to develop region-specific strategies, including enhanced socialization, training, and infrastructure support to improve the overall effectiveness of the Taspen Authentication Application across Indonesia.

Keywords
Taspen Authentication Application, K-Means, K-Medoids, Davies-Bouldin Index, Elbow Method

Corresponding Author :
anisyatasya8@gmail.com

INTRODUCTION

PT TASPEN (Savings and Pension Insurance) is an institution that provides social security and pension programs in Indonesia that manages various benefits such as old age pension, disability pension, and widow/widower's pension (Taspen, 2018). Based on Law No. 11 of 1969, retirees managed by PT Taspen include civil servants (PNS), members of the TNI, Polri, and several other professional categories who have entered retirement. On November 28, 2022, PT Taspen launched the Taspen Authentication Application to improve the security and efficiency of pension

service management (Taspen, 2022). However, based on data from the 2023 Taspen Authentication Application, there are inconsistencies in the implementation in the field, where participants have not fully authenticated online. In Jambi Province, the effectiveness of using the application reaches 60–70% with an ineffectiveness of 30–40%, as conveyed by the Head of Service of PT Taspen Jambi Solichah, S (2023) that "The use of the authentication taspen application has not been fully effective in accordance with the data on the status of blocking from the receipt of pension payments continues to occur."

According to the Taspen Report (2023), the difference in the effectiveness of application use between provinces shows the need for socialization based on regional characteristics. The approach of grouping provinces based on application effectiveness can be done using Cluster analysis, which is a method that groups objects into several groups with similar characteristics (Hair et al., 2019; Backhaus et al., 2023). One of the methods used is k-means and k-medoids. K-means choose the center of the Cluster based on the average, but sensitive to outliers, whereas k-medoids are more robust because they use medoids as the center of the cluster (Christopher, 2006; Flowrensia, 2010). The determination of the optimal number of clusters can be done using the elbow method, which plots the value of the objective function against various k-values to find a balance point between the complexity and quality of the clustering (Muller et al., 2016).

Validation of grouping results was carried out using the Davies Bouldien Index (DBI) which is included in the internal validation category (Sa'adah, 2021; Mustika et al., 2021). Some indicators of the effectiveness of the Taspen Authentication application include user factors, sessions, retention, login effectiveness, and conversion rate (Zufwari Fadli, 2016). Based on previous research by Pratiwi (2016), Astria (2019), Agustin and Sirait (2021), and Marlina et al. (2018), the comparison of k-means and k-medoids methods showed results that varied depending on the characteristics of the data. Therefore, this study aims to compare the two methods using the Elbow method to group provinces in Indonesia based on indicators of the effectiveness of using the Authentication Taspen Application.

RESEARCH METHOD

This study uses secondary data obtained from the Taspen Authentication application metrics, focusing on application usage effectiveness indicators which include ten variables, namely *user*, *session*, *session interval*, *retention*, login effectiveness, *churn rate*, *new users*, *error count*, *conversion rate*, and *response time*. The object of the study covers 38 provinces in Indonesia, so the analysis is

carried out on application usage data in all regions. The data used is ratio-scale and processed using the Cluster analysis approach, with steps including problem formulation, data standardization, assumption testing (KMO test and intervariable correlation), determination of similarity distances between objects using Euclidean distances, and determination of the optimal number of *clusters* using *the elbow* method. Next, grouping was carried out using the K-Means and K-Medoids algorithms, then the results were compared to see the effectiveness of each method. The process ended with the interpretation and validation *of the cluster* using the Davies-Bouldin Index (DBI) value to determine the best clustering results and draw conclusions based on the grouping pattern of indicators of the effectiveness of the Taspen Authentication application in each province.

RESULT AND DISCUSSION

Normality Test

Table 1.
Normality Test Results

	Kolmogorov-Smirnov[a]			*Shapiro-Wilk*		
	Statistic	df	Sig.	Statistic	df	Sig.
User	.105	38	.200	.970	38	.420
Session	.098	38	.200*	.974	38	.503
Session Interval	.130	38	.130	.950	38	.008
Retention	.173	38	.006	.925	38	.014
Efektivitas Login	.200	38	<,001	.919	38	.009
Churn Rate	.141	38	.054	.957	38	.049
New Users	.149	38	.032	.933	38	.025
Error Count	.125	38	.144	.921	38	.011
Conversion Rate	.141	38	.054	.951	38	.009
Response Time	.217	38	<,001	.904	38	.003

Based on the results obtained, it can be seen that only *the user* and *session* variables showed *p-value* results greater than 0.05 in the *Kolmogorov-Smirnov* and *Shapiro-Wilk tests*, which indicates that these variables are normally

distributed. On the other hand, most variables, such as *retention*, *login effectiveness*, and *response time*, have a p-value of less than 0.05, so it can be concluded that these variables are not normally distributed.

The factor extraction process is based on the results of the KMO and Bartlett tests (Tables 2 and 3) which show the data is suitable for further analysis using factor analysis. The two new factors are then used as the basis for the cluster analysis process, both with k-means and k-medoids algorithms. Thus, the provincial grouping is carried out based on the two main factors from the results of the factor analysis, no longer on the initial ten variables.

Cluster Analysis Assumptions

Table 2.
SME Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0.93

[1] 0.93

From the above results, the KMO value of 0.93 shows that the data in this study is very good and feasible to be used in further analysis. With this value, it can be continued to carry out factor analysis, which is the next step in this research process.

Bartlett's Test of Sphericity

Table 3. Barlett Test

Bartlett's Test of Sphericity	*Approx. Chi-Square*	689.930
	Df	28
	Sig.	<,001

The above results show that *the Bartlett's Test* is very significant ($p < 0.001$ is smaller than 0.05, so it can be concluded that there is a correlation between the indicators. Thus, data on indicators of the effectiveness of the use of authentication taspen applications can be further analyzed using multivariate analysis methods, one of which is *cluster* analysis. These results reinforce that the grouping of provinces based on authentication indicators can be carried out validly, because the indicators are interrelated and are not independent of each other.

Tabel 4.
Rotated Component Matrix

	Component 1	*Component 2*	*Outcome Variables*
z-score (x_1)	-0.7539	-0.0783	V1
z-score (x_2)	0.1512	-0.9835	V2
z-score (x_3)	0.0506	-0.9209	V2

z-score (x_4)	-0.4409	-0.8381	V2
z-score (x_5)	-0.9901	-0.0528	V1
z-score (x_6)	-0.9854	-0.0068	V1
z-score (x_7)	-0.9764	-0.0594	V1
z-score (x_8)	-0.9644	-0.0195	V1
z-score (x_9)	-0.9813	-0.0832	V1
z-score (x_{10})	-0.9779	-0.1041	V1

Table 5.
Key Component Analysis Data

Objek	V1	V2
1	-1.010278	-1.184720
2	-1.794631	-1.332422
3	0.528889	0.454765
4	-0.239199	-0.290549
5	-0.330727	-0.088348
6	-0.329915	-0.004947
7	-0.327267	-0.022495
8	-0.324199	-0.135558
9	-0.328570	0.011138
10	-0.328843	-0.005167

After the factor analysis with PCA, the correlation test of each variable was again carried out using the following hypothesis:

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

H_0 rejected when $r \geq 0{,}5$ In other words, there is a correlation between variables. Instead, H_0 accepted when $r < 0{,}5$ In other words, there is no correlation between variables.

The correlation value of each PCA result variable using R *software* can be seen in the following correlation matrix:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 1{,}700114 \times 10^{-16} \\ 1{,}700114 \times 10^{-16} & 1 \end{bmatrix}$$

Based on the matrix, the correlation value between variable 1 and variable 2 can be obtained by$\boldsymbol{R}$ $1{,}700114 \times 10^{-16}$ where this value is less than 0.5 so it is accepted which means that there is no strong correlation between the variables of the data used. After the correlation test was carried out, the KMO value test was adjusted again in Table 6. H_0 The KMO value of 0.93 indicates that the data

in this study is very good and feasible to be used in further analysis. With this value, it can be continued to carry out factor analysis, which is the next step in this research process.

Determining the Optimal Number *of Clusters*

Table 6.
SSE Results for Each Number *of Clusters*

Cluster	1	81.991
	2	24.327
	3	12.302
	4	7.888
	5	6.706
	6	6.000
	7	5.500
	8	5.200
	9	5.000
	10	4.900
Missing		.000

The above results show that the value of SSE decreases as the number of *clusters* increases. However, of concern is the point where the decline in SSE values starts to slow down this is the elbow point that indicates *the optimal number of clusters. In the table, there is a change in the pattern around* the 3rd cluster. This is where *elbows* appear visually. Before this point, the decline in SSE was relatively large. After this point, the decrease in the value of SSE is smaller and more stable, so the *optimal cluster* is in *cluster* 3.

Table 7.
Dissimilarity *Result* for Each Number *of Clusters*

Cluster	1	81.991
	2	24.326
	3	12.126
	4	7.728
	5	6.030
	6	5.035
	7	4.104
	8	3.435
	9	3.139
	10	2.788
Missing		.000

The above results show that the total value *of the dissimilarity* decreases as the number of *clusters* increases. However, of concern is the point at which the decline in the total *value of dissimilarity* begins to slow down this is called *the*

elbow point, which indicates the optimal number of *clusters*. From the data pattern above, it can be seen that a significant decrease occurred until around k=3 where the *decrease in dissimilarity* was quite large from k=1 to k=3. After k=3, the decrease *in dissimilarity* becomes very small and almost flattened. Meanwhile, after that, the decline began to be small, which shows that the optimal number of clusters is around *cluster* 3.

Optimal *Cluster* Results Analysis

Table 8.
Final Cluster Centers* for *K-Means

Variabel	*Cluster 1*	*Cluster 2*	*Cluster 3*
User	63.98	77.06	80.00
Session	44.90	52.60	60.90
Session Interval	44.88	53.34	36.60
Retention	54.41	66.93	66.00
Efektivitas Login	53.24	69.45	70.50
Churn Rate	53.91	64.39	62.18
New Users	46.81	63.31	64.50
Error Count	53.91	67.14	61.75
Conversion Rate	50.81	66.36	65.25
Response Time	58.90	71.10	77.50

Table 12 is the final *medoid* resulting from the iteration of *the k-medoids* algorithm. *This medoid* was chosen because it has the closest distance to all objects in the *cluster*, making it the most ideal representation to describe the characteristics of the *cluster*. All calculations were performed using the R software, which is presented in Table 12 as follows:

Table 9.
Final Cluster Centers for K-Medoids

Variabel	**Cluster 1**	**Cluster 2**	**Cluster 3**
User	57.15	75.84	82.10
Session	40.30	54.70	60.90
Session Interval	47.15	53.07	36.60
Retention	49.63	68.22	66.00
Efektivitas Login	51.83	69.00	70.50
Churn Rate	49.97	64.70	62.18
New Users	40.33	64.03	64.50
Error Count	46.83	67.47	61.75

Conversion Rate	48.47	66.62	65.25
Response Time	53.00	71.70	77.50

From the table above, it can be analyzed that each *cluster* has different characteristics. For example, *cluster* 1 in *k-means* shows the lowest *value of user, session, retention, and* other metrics compared to *other clusters, for example* users *around 64,* session *45,* retention *54,* login effectiveness *53, and* response time *59. K-medoid cluster* 1 also showed similar characteristics with *user* values of 57, session 40, retention of 49, and *login effectiveness* of 52, indicating a group of users with lower activity and engagement.

Cluster 1 consists of provinces with high average values of effectiveness indicators, such as number of active users, retention rate, and login effectiveness. Provinces in this *cluster* have made optimal use of the authentication app and tend to have a good level of technology adoption.

Cluster 2, the province in this group shows fairly good application usage, but there are still some constraints in terms of user retention or authentication frequency.

Cluster 3 contains provinces with relatively low average indicators, such as low number of new users, high *churn rate*, or many *errors* in applications. This indicates the need for special attention and increased socialization or training in the use of applications in these provinces.

Grouping with *K-Means* and *K-Medoids*

Table 10. Initial Centroid

Centroid	V1	V2
C1	-1.010278	-1.184720
C2	-1.794631	-1.332422
C3	0.528889	0.454765

Based on Table 13 C1 is the first *centroid* by taking the 1st object as the center of the *cluster*, C2 is the second *centroid* by taking the 2nd object as the center of the *cluster*. While C3 is the third *centroid* with the 3rd object as the center of the *cluster*.

Calculation of the distance between *the centroid* and the data object

The measure of the distance between the initial centroids symbolized as C1, C2, and C3 is on the data object using the Euclidean distance measure. The manual calculation of the distance between the first centroid (C1) to the data object is as follows: (more on Appendix)

Alacrity : Journal Of Education
Volume 5 Issue 3 Oktober 2025
Page : 1129-1143

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2}$$

$$d_{1,c1} = \sqrt{\sum_{k=1}^{2} (x_{1k} - c_{1k})^2}$$

$$= \sqrt{(x_{11} - c_{11})^2 + (x_{12} - c_{12})^2}$$

$$= \sqrt{\left((0.1786) - (-1.010278)\right)^2 + \left((-0.2987) - (-1.184720)\right)^2}$$

$$= 1{,}4827$$

$$d_{2,c1} = \sqrt{\sum_{k=1}^{2} (x_{2k} - c_{1k})^2}$$

$$= \sqrt{(x_{21} - c_{11})^2 + (x_{22} - c_{12})^2}$$

$$= \sqrt{\left((0.9452) - (-1.010278)\right)^2 + \left((1.2420) - (-1.184720)\right)^2}$$

$$= 3{,}1170$$

$$\vdots$$

$$d_{38,c1} \sqrt{\sum_{k=1}^{2} (x_{38k} - c_{1k})^2}$$

$$= \sqrt{(x_{381} - c_{11})^2 + (x_{382} - c_{12})^2}$$

$$= \sqrt{\left((-1.7908) - (-1.010278)\right)^2 + \left((-1.1399) - (-1.184720)\right)^2}$$

$$= 0{,}7818$$

The manual calculation of the distance between *the second centroid* (C2) to the data object is as follows: (more on that in the Appendix)

$$d_{1,c2} = \sqrt{\sum_{k=1}^{2} (x_{1k} - c_{2k})^2}$$

$$= \sqrt{(x_{11} - c_{21})^2 + (x_{12} - c_{22})^2}$$

$$= \sqrt{\left((-0.1786) - (-1.794631)\right)^2 + \left((-0.2987) - (-1.332422)\right)^2}$$

$$= 2{,}2277$$

$$d_{2,c2} = \sqrt{\sum_{k=1}^{2} (x_{2k} - c_{2k})^2}$$

$$= \sqrt{(x_{21} - c_{21})^2 + (x_{22} - c_{22})^2}$$

$$= \sqrt{\left((0.9452) - (-1.794631)\right)^2 + \left((1.2420) - (-1.332422)\right)^2}$$

$$= 3{,}7601$$

Alacrity : Journal Of Education
Volume 5 Issue 3 Oktober 2025
Page : 1129-1143

$$\vdots$$

$$d_{38,c2}\sqrt{\sum_{k=1}^{2}(x_{38k}-c_{2k})^2}$$

$$=\sqrt{(x_{381}-c_{21})^2+(x_{382}-c_{22})^2}$$

$$=\sqrt{\left((-1.7908)-(-1.794631)\right)^2+\left((-1.1399)-(-1.332422)\right)^2}$$

$$=0{,}1926$$

The manual calculation of the distance between *the third centroid* (C3) to the data object is as follows: (more on Appendix)

$$d_{1,c3}=\sqrt{\sum_{k=1}^{2}(x_{1k}-c_{3k})^2}$$

$$=\sqrt{(x_{11}-c_{31})^2+(x_{12}-c_{32})^2}$$

$$=\sqrt{\left((0.1786)-(0.528889)\right)^2+\left((-0.2987)-(0.454765)\right)^2}$$

$$=0{,}8309$$

$$d_{2,c3}=\sqrt{\sum_{k=1}^{2}(x_{2k}-c_{3k})^2}$$

$$=\sqrt{(x_{21}-c_{31})^2+(x_{22}-c_{32})^2}$$

$$=\sqrt{\left((0.9452)-(0.528889)\right)^2+\left((1.2420)-(0.454765)\right)^2}$$

$$=0{,}8905$$

$$\vdots$$

$$d_{38,c3}\sqrt{\sum_{k=1}^{2}(x_{38k}-c_{3k})^2}$$

$$=\sqrt{(x_{381}-c_{31})^2+(x_{382}-c_{32})^2}$$

$$=\sqrt{\left((-0.1786)-(0.528889)\right)^2+\left((-1.1399)-(0.454765)\right)^2}$$

$$=1{,}7446$$

Table 11. *Centroid* Iteration 1

Centroid	V1	V2
C1	-0,7094	-0,9247
C2	-1,4231	-1,0530
C3	0,5687	0,5168

Table 12. *Centroid* Iteration 2

Centroid	V1	V2
C1	0,3947	0,9619

C2	1,0480	0,5345
C3	1,9745	2,5864

Based on the results of the calculation of iteration 2 in Table 18, it can be seen that the grouping results in the new iteration have experienced data transfer from *the cluster* results obtained compared to the results of the previous grouping. So that a new *centroid* value is obtained from taking the average position of the data in each *cluster* for iteration 3 presented in Table 19.

Table 13. *Centroid* Iteration 3

Centroid	V1	V2
C1	0,1871	0,6864
C2	0,6151	0,1960
C3	1,6142	2,0770

Cluster 3 is the *cluster* with the highest positive and average value. This indicates that the data in this *cluster* has a higher value than the total average for both variables. In other words, *cluster* 3 is categorized as "high". So that the provinces included in *cluster* 3 are the provinces with the most effective or high effectiveness of using the authentication taspen application. There are 19 members in *cluster* 3, namely North Sumatra, DKI Jakarta, West Java, Central Java, Yogyakarta, East Java, West Kalimantan, East Kalimantan, South Kalimantan, Central Kalimantan, North Kalimantan, Bali, West Nusa Tenggara, Gorontalo, West Sulawesi, Central Sulawesi, North Sulawesi, Southeast Sulawesi and South Sulawesi.

Cluster 1 shows that the data in this *cluster* has a medium amount compared to the total average. *Cluster* 1 is categorized as "Intermediate". So that the provinces included in *cluster* 1 are provinces with the effectiveness of using the authentication taspen application with a medium effective level. There are 5 members in *cluster* 1, namely Bengkulu, Lampung, East Nusa Tenggara, North Maluku and Maluku.

Cluster 2 shows that the data on *this cluster* has a value lower than the average of the total which is categorized as "low". So that the provinces included in *cluster* 2 are the provinces with the least effective or low effectiveness of using the authentication application taspen. The provinces included in *cluster* 2 are 14 provinces including Banda Aceh, South Sumatra, West Sumatra, Riau, Riau Islands, Jambi, Bangka Belitung, Banten, West Papua, Papua, South Papua, Central Papua, Mountainous Papua and Southwest Papua.

Alacrity : Journal Of Education
Volume 5 Issue 3 Oktober 2025
Page : 1129-1143

The results of the grouping *of the k-means* algorithm using R *software* obtained the following results:

Clustering Vector .

[1] 2 3 2 2 1 2 2 2 1 2 2 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 1 1 2 2 2 2 2 2

Davies Bouldin Index

Table 14. Value R_{ij}

Centroid	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**
Cluster 1	0	0,782	1,246
Cluster 2	0,782	0	2,005
Cluster 3	1,246	2,005	0

[1] 2.717

After obtaining the value of each R_{ij}*cluster* resulting from the *k-means* algorithm in Table 29, the maximum value of each cluster against the other clusters presented in Table 15 is then sought.R_{ij}

Table 15. Maximum Value R_{ij}

Cluster	**Maksimum R_{ij}**
Cluster 1	1,246
Cluster 2	2,005
Cluster 3	2,005

After obtaining the maximum value of each number of R_{ij}*clusters* used based on the results of the *k-means* algorithm, then the DBI value can be calculated by calculating the average value of each maximum value.R_{ij}

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}(R_{i,j})$$
$$= \frac{1}{3}(1{,}246 + 2{,}005 + 2{,}005)$$
$$= 1{,}752$$

a. *Davies Bouldin Index* Algorithm results *K-Medoids*

Table 16. Value R_{ij}

Centroid	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**
Cluster 1	0	2,314	0,815
Cluster 2	2,314	0	3,072
Cluster 3	0,815	3,072	0

[1] 1,789

After obtaining the value of each R_{ij}*cluster* resulting from the *k-means* algorithm in Table 31, the maximum value of each R_{ij}*cluster* against *the other clusters presented in Table 32 is then sought.*

Alacrity : Journal Of Education
Volume 5 Issue 3 Oktober 2025
Page : 1129-1143

Table 17. Maximum Value R_{ij}

Cluster	Maksimum R_{ij}
Cluster 1	2,314
Cluster 2	3,072
Cluster 3	3,072

After obtaining the maximum value of each number of R_{ij}*clusters* used based on the results of *the k-medoids* algorithm, then the DBI value can be calculated by calculating the average value of each maximum value.R_{ij}

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}(R_{i,j})$$
$$= \frac{1}{3}(2{,}314 + 3{,}072 + 3{,}072)$$
$$= 2{,}819$$

Based on the calculation, the DBI value for each method is obtained from the manual *cluster* result and the output R result can be seen in the following table:

Coupistin *Silhouettes*

Table 18.
Cluster Evaluation Results *Using* Silhoutte *Coefficient*

No.	Algoritma	Koefisien *Silhoutte*
1	*K-Means*	0,2850
2	*K-Medoids*	0,2073

Based on Table 18, it can be seen that the results of *cluster evaluation using* the Silhoutte *coefficient* obtained the highest value in the manual *cluster* results, namely the *k-means* algorithm with a value of 0.2850 which is closest to 1. So, it can be concluded that the best grouping results are the cluster results using the *k-means algorithm*.

CONCLUSION

The validation results using the Davies-Bouldin Index (DBI) showed that the best grouping was obtained with the K-Means algorithm with a DBI value of 1.752 and a Silhouette Coefficient of 0.2850, which means that K-Means provides more optimal clustering results. Based on these results, 50% or 19 provinces in Indonesia have been optimal in utilizing the Taspen Authentication application, 13.16% or five provinces are in the medium category, and 36.84% or 14 provinces are relatively low in the effectiveness of using the application. These results can be the basis for PT Taspen (Persero) and policy makers to determine different intervention strategies according to the

level of effectiveness of each *cluster*, such as increasing socialization, training, or infrastructure support in low-category provinces so that the use of the Taspen Authentication application can increase evenly throughout Indonesia.

REFERENCES

Ali Muhidin. (2009). *Analisis Korelasi Regresi dan Jalur dalam Penelitian*. Bandung: CV Pustaka Setia.

Asra, Abuzar. (2017). *Analisis Multivariabel: Suatu Pengantar*. Bogor: In Media.

Astria, C. (2019). *Penerapan K-Medoid Pada Rumah Tangga Yang Memiliki Sumber*, 3(1), 604–609. https://doi.org/10.30865/komik.v3i1.1667

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2023). *Multivariate Data Analysis*. Springer.

Christopher, M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Everitt, B., & Hothorn, T. (2011). *Cluster Analysis* (5th ed.). Chichester: John Wiley & Sons.

Flowrensia, Y. (2010). *Perbandingan Penggerombolan K-Means dan K-Medoid pada Data yang Mengandung Pencilan*. Bogor: Institut Pertanian Bogor.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). New York: Pearson Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). *Multivariate Data Analysis* (8th ed.). New York: Pearson Prentice Hall.

Mahmudi. (2005). *Manajemen Kinerja Sektor Publik*. Yogyakarta: UPP-AMP YKPM.

Marlina, M., Putri, E., Fernando, A., & Ramadhan, R. (2018). *Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak*, 4(2), 64. https://doi.org/10.24014/coreit.v4i2.4498

Muller, A., & Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media.

Mustika, D., et al. (2021). *Analisis Validasi Internal Clustering Menggunakan Davies-Bouldin Index*.

Narimawati, U. (2008). *Metodologi Penelitian Kualitatif dan Kuantitatif: Teori dan Aplikasi*. Bandung: Agung Media.

Pratiwi, D., & Agustin, E. P. (2016). *Penerapan K-Means Clustering untuk Memprediksi Minat Nasabah pada PT Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih*. Palembang: Universitas Bina Darma.

Sa'adah, N. (2021). *Analisis Validasi Clustering Menggunakan Davies-Bouldin Index*.

Santoso, S. (2015). *SPSS 20: Pengolahan Data Statistik di Era Informasi*. Jakarta: PT Alex Media Komputindo.

Sihombing. (2022). *Metode Analisis Multivariat*. Jakarta: Universitas Mercu Buana.

Suprayogi. (2013). *Penelitian Metode Analisis Statistika*. Jakarta: PT Rajagrafindo Persada.

Taspen. (2018). *Laporan Tahunan PT Taspen (Persero) Tahun 2018*. Jakarta.

Taspen. (2019). *Laporan Tahunan PT Taspen (Persero) Tahun 2019*. Jakarta.

Taspen. (2022). *Peluncuran Aplikasi Taspen Otentikasi*. Jakarta.

Taspen. (2023). *Laporan Tahunan PT Taspen (Persero) Tahun 2023*. Jakarta.

Wananda, J., & Nugraha, N. (2023). *Implementasi K-Means Clustering Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan*. Yogyakarta: Universitas Islam Indonesia.

Wibowo, E. (2023). *Analisis Data Multivariat*. Yogyakarta: Universitas Pembangunan Nasional Veteran Yogyakarta.

Zufwari, F. (2016). *Faktor-Faktor Keefektifan Aplikasi Mobile*. Jakarta: ND