ORIGINAL ARTICLE

Evaluation of Clustering Algorithm for *Diabetes mellitus* Intensive Care Unit Patients' Dataset Using K-Means and DBSCAN Techniques

Nur Jihan Mohd Zukhi[1], Mohd Zulfakar Mazlan[2], Farhah Muhammad[1], Mastura Mohd Sopian[1], Fatanah Mohamad Suhaimi[1,3]

[1] Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia
[2] Department of Anaesthesiology and Intensive Care, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, 16150 Kota Bharu, Kelantan, Malaysia
[3] Dental Simulation and Virtual Learning Research Excellence Consortium, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia

ABSTRACT

Introduction: Insulin infusion therapy is applied for all Intensive Care Unit (ICU) patients who have hyperglycaemia regardless of diabetic history or status. Unfortunately, improper glycaemic control and patient dynamics may cause hypoglycaemia episodes. The purpose of this study was to categorise ICU patients by considering potential variables that may affect blood glucose for the effectiveness of insulin infusion therapy. **Materials and methods:** Thirty-three *diabetes mellitus* ICU patients who received insulin infusion therapy and had at least 12 blood glucose readings throughout their ICU stay were selected for this study. Variables of a fraction of inspired oxygen (FiO_2) and mean arterial pressure (MAP) were chosen for clustering this cohort using K-means and density-based spatial clustering of applications with noise (DBSCAN) techniques. **Results:** he outcomes showed K-means produced four clusters for each variable, while DBSCAN provides unstandardised clusters, nine for FiO_2 and four for MAP. K-means is more ap-propriate for this cohort, validated by Silhouette analysis and clinically applicable cluster numbers for both variables. The clusters obtained from K-means for FiO2 were > 0.5, > 0.4, > 0.3, and ≤ 0.3, while for MAP were > 98.5 mmHg, 86 mmHg, > 75 mmHg, and ≤ 75 mmHg. **Conclusion:** The formation of these clusters is expected to improve the effectiveness of the insulin infusion protocol while addressing personalised treatment for the categorised patients. *Malaysian Journal of Medicine and Health Sciences* (2024) 20(SUPP8): 81-85. doi:10.47836/mjmhs20.s8.11

Keywords: K-means clustering, DBSCAN clustering, FiO_2, MAP, *Diabetes mellitus* ICU patients

Corresponding Author:
Fatanah Mohamad Suhaimi, PhD
Email: fatanah.suhaimi@usm.my
Tel : +604-5622561

INTRODUCTION

Patients react differently to the treatment they receive in the intensive care unit (ICU) due to differences in physiology. Commonly, most ICU patients may experience stress-induced-hyperglycemia (SIH) during their stay in the ICU due to the stress induced by the disease. SIH happened to the ICU patient with or without diabetes prior to the ICU admission. SIH occurs because the disease affects the insulin sensitivity of the cells and diminishes insulin secretion as stress hormones, including adrenaline, and noradrenaline, are released. These conditions elevate the glucose level in the blood and cause hyperglycemia. Thus, it is crucial to control the blood glucose level among ICU patients to avoid more complications leading to ICU mortality (1–4).

Insulin infusion therapy is used to regulate blood glucose levels in ICU patients using the sliding scale approach, especially in Malaysia's ICUs (5). Elliot Joslin invented the sliding scale technique in 1934 (6,7). Nonetheless, the current sliding scale only considers the blood glucose levels in determining the insulin dose, and it is a one-size-fits-all procedure for insulin therapy which is used for ICU patients with and without diabetes history. Moreover, the main drawback of the sliding scale is that high hypoglycemia events occurred during the treatment, which also led to mortality in the ICU (6,8). Thus, categorising the patients based on their diabetic background and introducing some clinical variables may help to optimise insulin infusion therapy's effectiveness.

Hence, this study introduces new potential variables to be considered for administering insulin, which are fraction

of inspired oxygen (FiO_2) and mean arterial pressure (MAP), which are personalised for *diabetes mellitus* (DM) ICU patients. ICU patients frequently experience acute respiratory failure and sepsis during their stay in the ICU. Lung injury is one of the acute respiratory failures that necessitate oxygen support during their stay (9–11). The high quantities of oxygen will influence glucose metabolism and insulin sensitivity in patients, resulting in hyperglycemia episodes. Additionally, sepsis can cause sepsis-induced hypotension, which decreases MAP levels and causes organ failure owing to hypoperfusion (12,13). The body's stress response will be triggered by the high levels of FiO_2 and hypotension, which result in the release of stress hormones. Hence, the blood glucose levels will increase as stress hormones are released in the body, which leads to SIH among ICU patients.

The classification of FiO_2 and MAP are essential in designing the new insulin infusion protocol to provide efficient insulin treatment. Hence, the clustering technique should be done on FiO_2 and MAP to obtain the optimum cluster. There are many clustering approaches available, such as K-means, hierarchical, and density-based spatial clustering of applications with noise (DBSCAN) clustering (14,15). This study employed K-means and DBSCAN techniques to cluster FiO_2 and MAP, and to evaluate appropriate techniques applicable to the DM cohort in this study.

K-means clustering is the simplest clustering approach and produces excellent outcomes (14,16,17). The k-value of this method is obtained from an elbow-shape graph, which indicates the cluster number. Meanwhile, the DBSCAN technique provides optimal results for large datasets (18–21). Before executing the DBSCAN clustering, the neighbour number (minpts) and Euclidean distance (ε) threshold must be determined (22,23). DBSCAN implemented a knee-shape graph to obtain the Euclidean distance (k-distance).

MATERIALS AND METHODS

This study included 33 ICU patients with *diabetes mellitus* (DM) from Hospital Universiti Sains Malaysia (HUSM), who underwent insulin treatment and had more than 12 blood glucose readings while in the ICU. The Universiti Sains Malaysia Human Research Ethics Committee (HREC) approved this study with an ethics code of USM/JEPeM/16100402. MATLAB (version R2020a) was implemented for data processing, analysis, and clustering.

This study included 42% female patients with a median age of 56 years old, as indicated in Table I. The cohort of this study showed the median blood glucose was 10.4 mmol/L, which is considered hyperglycemia. Furthermore, this cohort had a FiO_2 level median of 0.4, which is slightly higher than the lung injury prediction score ($FiO_2 > 0.35$) (9–11). However, the median MAP of this cohort was 82.7 mmHg, which indicates the patients are not within the sepsis category as the guideline for sepsis-induced hypotension is MAP < 70 mmHg (13).

Table I: Demographic of the patients.

Characteristics	Value*
Number of patients	33
Percentage of female (%)	42
Age (years old)	56 [48 – 61]
Blood glucose (mmol/L)	10.4 [8.3 – 12.8]
Fraction of inspired oxygen, FiO_2	0.4 [0.3 – 0.5]
Mean Arterial Pressure, MAP (mmHg)	82.7 [75.0 – 91.0]

* Median [IQR (25^{th} – 75^{th})] where appropriate

K-means provide the elbow-shape graphs for FiO_2 and MAP, as shown in Fig. 1(A) and Fig. 1(B), respectively. The y-axis and x-axis of the elbow-shape graphs were used to represent the within-cluster sum of squares (WCSS) and k value, respectively. The value of k is derived based on a significant decrease observed in the elbow graph (24). According to the significant decrease in these graphs, both variables gave four k values, indicating that four clusters were obtained from DM patients. These cluster numbers were then utilised to categorise this cohort's FiO_2 and MAP values.

Fig. 1(C) and Fig. 1(D) depict the knee-shape from DBSCAN clustering for FiO_2 and MAP. It is important to make a sensible choice for the minimum point selection in the study since there are no established criteria for estimating the DBSCAN parameter (25). Therefore, different minimum points of four and 10, which have been used in this study to identify the most appropriate point for the cohort. For the minimum point inside the Euclidean distance, the number four was chosen as it resulted in a significant number of clusters, as shown in Fig. 2. FiO_2 had an epsilon value of 0.05 from the Euclidean distance at the y-axis (k-distance), while MAP had an epsilon value of 0.12. These epsilon values were utilised to display the FiO_2 and MAP clusters acquired from DBSCAN.

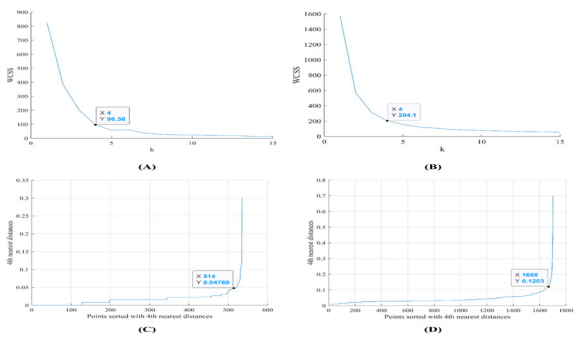

Fig. 1: Graphs of k-distance from K-means and elbow-shape graphs from DBSCAN. K-means: (A) FiO2 and (B) MAP, and DBSCAN: (C) FiO2 and (D) MAP.

Fig. 2 shows the flowchart of the DBSCAN technique that utilised different minimum points at the beginning

to decide which minimum point was more applicable to these data. The epsilon values and cluster numbers obtained differed when other minimum points were applied. As shown in Fig. 2(B), the MAP resulted in one cluster with a minimum point of ten, which indicates more clusters need to be formed. This caused the minimum point of ten to be rejected for FiO_2 and MAP in this study.

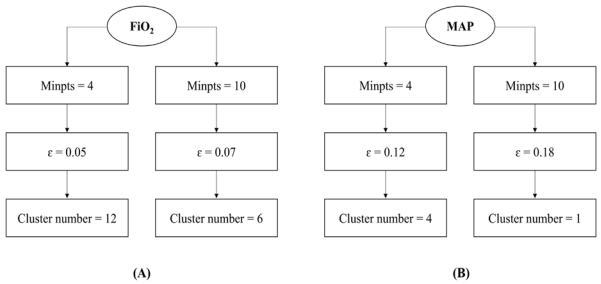

Fig. 2: Flowchart of DBSCAN method. (A) FiO2 and (B) MAP.

RESULTS

FiO_2 and MAP have been linked to patients' blood glucose levels in an indirect way, yet they are easily obtained at the patients' bedside. This study showed that FiO2 had a weak negative correlation with blood glucose ($r = -0.06$, p-value = 0.02), as well, MAP with blood glucose ($r = -0.09$, p-value < 0.05). Even though these variables had a weak correlation with blood glucose, they had a significant correlation with a p-value < 0.05.

Fig. 3 depicts the FiO_2 and MAP clusters produced from K-means and DBSCAN clustering in DM patients. The FiO_2 clusters obtained from K-means and DBSCAN are shown Fig. 3(A) and Fig. 3(C), respectively. K-means clustering provided four clusters, while DBSCAN clustering provided nine clusters, which were non-coherent clusters with some outliers. However, from Fig. 3(C), the DBSCAN provided an excessive number of clusters; some clusters had identical FiO_2 values but different blood glucose levels. For example, Cluster 4 and Cluster 5 in Fig. 3(C) had a FiO_2 value of 0.6, but different blood glucose ranges of 4.8 – 13.6 mmol/L and 14.7 – 16.1 mmol/L, respectively. This condition defeated the objective of classifying the information into a functional group.

The clusters for MAP from the K-means and DBSCAN methods are visualised in Fig. 3(B) and Fig. 3(D), respectively. Both methods provided four clusters for MAP among DM ICU patients. However, DBSCAN provided one large cluster and three small clusters. There were also outliers marked in red due to the patients' variability, as shown in Fig. 3(D). The MAP cluster obtained from the DBSCAN technique also resulted in non-coherent clusters as it formed the cluster based on the dense data region. The large cluster from DBSCAN is the MAP within 63 – 110 mmHg. The smallest cluster, Cluster 2, has four members, as seen in Fig. 3(D). Cluster 2 is formed only between 10.0 – 11.3 mmol/L blood glucose readings.

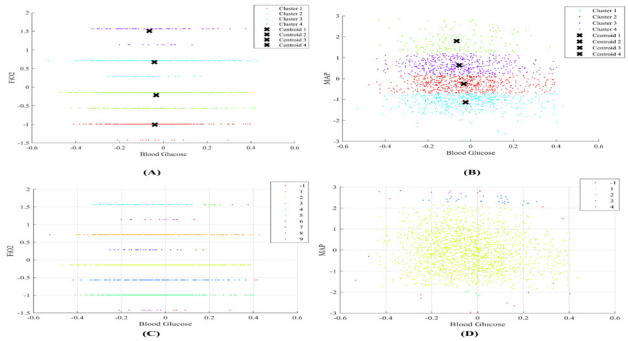

Fig. 3: Clusters obtained from K-means and DBSCAN techniques. K-means: (A) FiO2 and (B) MAP, and DBSCAN: (C) FiO2 and (D)MAP.

From this study, K-means is a more appropriate technique applicable to the DM ICU cohort. Hence, the range of FiO2 and MAP clusters derived from K-means are tabulated in Table II. The FiO2 range is FiO2 ≤ 0.3, FiO2 > 0.3, FiO2 > 0.4, and FiO2 > 0.5, while the MAP range is MAP > 98.5 mmHg, MAP > 86.0 mmHg, MAP > 75.0 mmHg, and MAP ≤ 75.0 mmHg.

Table II: The cluster ranges of FiO_2 and MAP in DM patients.

Variables	Cluster	Normalised data	Absolute value	Range
FiO_2	1	-1.42 to -0.99	0.25 to 0.3	≤ 0.3
	2	-0.45 to -0.14	0.35 to 0.4	> 0.3
	3	0.28 to 0.71	0.45 to 0.5	> 0.4
	4	1.14 to 1.57	0.55 to 0.6	> 0.5
MAP (mmHg)	1	1.19 to 2.84	99 to 118	> 98.5
	2	0.17 to 1.19	87 to 98.33	> 86.0
	3	-0.69 to 0.17	75.67 to 86	> 75.0
	4	-2.96 to -0.74	48.33 to 75	≤ 75.0

Silhouette analysis was done on the K-means to validate the separation distance between the clusters obtained from this research. The graphs of Silhouette for FiO_2 and MAP are visualised in Fig. 4. Silhouette's score for FiO_2 was 0.677, while MAP was 0.467, which is indicated as a straight line (red) in the graphs.

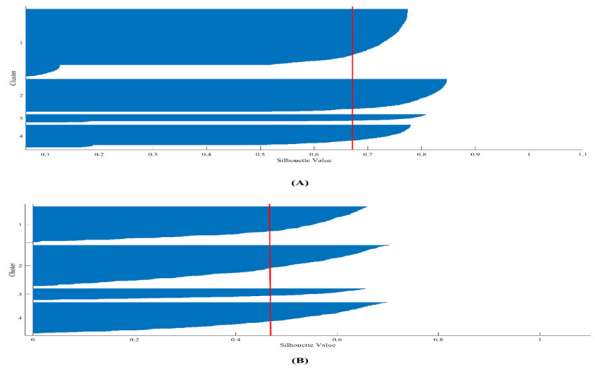

Fig. 4: Silhouette graphs. (A) FiO_2 and (B) MAP.

The thickness of the graphs for FiO2 fluctuated, which can be seen in Fig. 4(A). This shows that the cluster size

for FiO_2 is not similar in size, and the thinnest cluster among FiO_2 is Cluster 4. Meanwhile, MAP had quite similar thickness for three clusters, but it also had one thinnest cluster similar to FiO_2.

DISCUSSION

From this study, FiO_2 and MAP can be utilised to design novel insulin infusion therapy as these variables have a significant correlation with blood glucose. Hence, introducing these potential variables can improve the effectiveness and efficiency of the current sliding scale-based insulin infusion therapy. Besides, hypoglycemia events can be reduced during insulin treatment.

The findings showed that K-means is a more acceptable approach for the DM ICU cohort in this study, as K-means provided a standardised cluster for FiO_2 and MAP. Meanwhile, DBSCAN formed an excessive number of clusters and non-coherent clusters since it developed the cluster based on the data density within the Euclidean distance. This result makes it challenging to utilise the cluster acquired from the DBSCAN technique. Furthermore, this will increase the complexity of the protocol when adopting insulin infusion treatment since it has too many classifications. The main impetus for this study is to improve the effectiveness of the insulin infusion treatment by clustering the cohort and providing effective insulin administration.

The cluster of $FiO_2 > 0.3$ obtained from this study is closely equivalent to the lung injury prediction score ($FiO_2 > 0.35$) (11). This cluster will facilitate the clinical staff in deciding the insulin dose to be administered to the DM patients, which will improve the effectiveness of the insulin infusion treatment. Moreover, the MAP cluster of ≤ 75 mmHg from this study is approximately matched with the MAP score (< 70 mmHg) for sepsis-induced hypotension (13). The MAP ≤ 75 mmHg from this study may also be a reference for nurses and clinicians in implementing insulin infusion therapy for sepsis DM patients. Thus, different groups will receive appropriate insulin doses during their treatment in the ICU.

The choice of K-means as the better clustering approach in this study is validated by the Silhouette analysis results. Silhouette's score for FiO_2 and MAP was 0.677 and 0.467, respectively, and all clusters obtained surpassed Silhouette's score, indicating that all clusters are practically acceptable. Furthermore, the K-means clusters are well separated, with one relatively smallest cluster for both variables.

CONCLUSION

The clustering method can be used to classify data in order to assess the impact of potential variables in establishing novel insulin infusion protocols. The findings from this study showed that K-means clustering is ideal for the DM cohort of this dataset compared to the DBSCAN technique. K-means provided standardised clusters for FiO2 and MAP compared to DBSCAN. Furthermore, the potential variables of FiO2 and MAP can be considered in developing a novel personalised insulin infusion therapy to improve the effectiveness and efficiency of the current insulin infusion protocol.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support from the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme (FRGS), with reference code number FRGS/1/2020/STG01/USM/02/3 and account code number 203.CIPPT.6711841.

REFERENCES

1. Hui M, Kumar A, Adams GG. Protocol-directed insulin infusion sliding scales improve perioperative hyperglycaemia in critical care. Perioper Med. 2012;1(1):7. doi:10.1186/2047-0525-1-7.
2. Siegelaar SE, Hermanides J, Oudemans-van Straaten HM, van der Voort PHJ, Bosman RJ, Zandstra DF, et al. Mean glucose during ICU admission is related to mortality by a U-shaped curve in surgical and medical patients: A retrospective cohort study. Crit Care. 2010;14(6). doi:10.1186/cc9369.
3. Olariu E, Pooley N, Danel A, Miret M, Preiser JC. A systematic scoping review on the consequences of stress-related hyperglycaemia. PLoS One. 2018;13(4):1–15. doi:10.1371/journal.pone.0194952.
4. Lazzeri C, Bonizzoli M, Cianchi G, Ciapetti M, Socci F, Peris A. The prognostic role of peak glycemia and glucose variability in trauma: a single-center investigation. Acta Diabetol. 2020;57(8):931–935. doi:10.1007/s00592-020-01493-w
5. Ministry of Health Malaysia. Practical Guide to Inpatient Glycaemic Care. 2020. 35–39 p. Available from internet: https://mems.my/wp-content/uploads/2020/07/PRACTICAL-GUIDE-TO-INPATIENT-GLYCAEMIC-CARE_V2-2020.pdf
6. Guillermo EU, Andres P, Dawn S. Sliding Scale Insulin Use: Myth or Insanity? Am J Med. 2007;120(7):563–567. doi: 10.1016/j.amjmed.2006.05.070.
7. Joslin EP. Diabetic Manual For The Mutual Use of Doctor and Patient. Lea, Febiger, editors. Philadelphia; 1937.
8. Kushimoto S, Abe T, Ogura H, Shiraishi A, Saitoh D, Fujishima S, et al. Impact of blood glucose abnormalities on outcomes and disease severity in patients with severe sepsis: An analysis from a multicenter, prospective survey of severe sepsis. PLoS One. 2020;15(3):1–15. doi:10.1371/journal.pone.0229919.
9. Senapathi TGA, Wiryana M, Widnyana IMG,

Ryalino C, Har-Tono B. Lung injury prediction score as a predictor of acute respiratory distress syndrome in intensive care unit. Crit Care Shock. 2020;23(3):136–141. Available from: https://criticalcareshock.com/2020/06/11/lung-injury-prediction-score-as-a-predictor-of-acute-respiratory-distress-syndrome-in-intensive-care-unit/
10. Trillo-Alvarez C, Cartin-Ceba R, Kor DJ, Kojicic M, Kashyap R, Thakur S, et al. Acute lung injury prediction score: Derivation and validation in a population-based sample. Eur Respir J. 2011;37(3):604–609. doi:10.1183/09031936.00036810.
11. Gajic O, Dabbagh O, Park PK, Adesanya A, Chang SY, Hou P, et al. Early identification of patients at risk of acute lung injury: Evaluation of lung injury prediction score in a multicenter cohort study. Am J Respir Crit Care Med. 2011;183(4):462–470. doi:10.1164/rccm.201004-0549OC.
12. Leone M, Asfar P, Radermacher P, Vincent JL, Martin C. Optimizing mean arterial pressure in septic shock: A critical reappraisal of the literature. Crit Care. 2015;19(1):1–7. doi:10.1186/s13054-015- 0794-z.
13. Dellinger RP, Levy M, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med. 2013;39(2):165–228. doi:10.1007/s00134-012-2769-8.
14. Nithya NS, Duraiswamy DK, Gomathy P. A Survey on Clustering Techniques in Medical Diagnosis. Int J Comput Sci Trends Technol. 2013;1(1):17–
22. Available from: https://www.ijcstjournal.org/volume-1/issue-2/IJCST-V1I2P4.pdf
15. Nedyalkova M, Madurga S, Simeonov V. Combinatorial k-means clustering as a machine learning tool applied to *diabetes mellitus* type 2. Int J Environ Res Public Health. 2021;18(4):1–10. doi:10.3390/ijerph18041919.
16. MacQueen, James and others. Some methods for classification and analysis of multivariate observations. Proc fifth Berkeley Symp Math Stat Probab. 1967;1(14):281–297. Available from: https://digitalassets.lib.berkeley.edu/math/ucb/text/math_s5_v1_article-17.pdf
17. Braiki M, Benzinou A, Nasreddine K, Hymery N. Automatic Human Dendritic Cells Segmentation Using K-Means Clustering and Chan-Vese Active Contour Model. Comput Methods Programs Biomed. 2020;195:1–16. doi:10.1016/j.cmpb.2020.105520.
18. Yang Y, Qian C, Li H, Gao Y, Wu J, Liu CJ, et al. An efficient DBSCAN optimized by arithmetic optimization algorithm with opposition-based learning. J Supercomput. 2022;78(18):19566–19604. doi: 10.1007/s11227-022-04634-w.
19. Tomar D, Agarwal S. A survey on data mining approaches for healthcare. Int J Bio-Science Bio-Technology. 2013;5(5):241–266. doi:10.14257/ijbsbt.2013.5.5.25.
20. Barhak J. Visualization and Pre-Processing of Intensive Care Unit Data Using Python Data Science Tools Visualization and Pre-Processing of Intensive Care Unit Data Using Python Data Science Tools. In: MODSIM World 2018. 2018. p. 1–11. Available from: https://www.modsimworld.org/papers/2018/MODSIM_2018_Barhak.pdf
21. Ester M, Kriegel H-P, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proc 2nd Int Conf Knowl Discov Data Min. 1996;226–231. Available from: https://cdn.aaai.org/KDD/1996/KDD96-037.pdf
22. Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Trans Database Syst. 2017;42(3):1–21. doi:10.1145/3068335.
23. Sharafoddini A. Toward Precision Medicine in Intensive Care: Leveraging Electronic Health Records and Patient Similarity. University of Waterloo. 2019. Available from: https://uwspace.uwaterloo.ca/handle/10012/14726
24. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster Integration
K- Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. IOP Conf Ser Mater Sci Eng. 2018;336. doi:10.1088/1757-899X/336/1/012017.
25. Lai W, Zhou M, Hu F, Bian KAI, Song QI. A New DBSCAN Parameters Determination Method Based on Improved MVO. IEEE Access. 2019;7:104085–104095. doi:10.1109/ACCESS.2019.2931334.