SINERGI Vol. 27, No. 3, October 2023: 433-442
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.2023.3.014

Service quality dealer identification: the optimization of K-Means clustering

Yolanda Enza Wella[1], Okfalisa Okfalisa[1,*], Fitri Insani[1], Faisal Saeed[2], Ab Razak Che Hussin[3]
[1]Informatics Engineering Department, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[2]School of Computing and Digital Technology, Birmingham City University, United Kingdom
[3]Information Systems Department, Universiti Teknologi Malaysia, Malaysia

Abstract
Service quality and customer satisfaction directly influence company branding, reputation and customer loyalty. As a liaison between producers and consumers, dealers must preserve valuable consumer relationships to increase customer satisfaction and adherence. Lack of comprehensive measurement and standardization regarding service quality emerges as a consideration issue towards the company service excellence. Therefore, identifying the service quality performance and grouping develops into valuable contributions in decision-making to control and enhance the company's intention. This study applies the K-Means Algorithm by optimizing the number of clusters in identifying dealer service quality performance. Hence, the ultimate service quality formation will be performed. The analysis found three dealer identification categories, including Cluster One, with 125 dealers grouped as good performance; Cluster Two, with 30 dealers grouped as very good performance; and Cluster Three, with 38 dealers grouped as not good performance. In order to evaluate the efficacy of optimum k value, the lists of testing approaches are conducted and compared, whereby Calinski-Harabasz, Elbow, Silhouette Score, and Davies-Bouldin Index (DBI) contribute in k=3. As a result, the optimum clusters are determined through the highest performance of k values as three. These three clusters have successfully identified the service quality level of dealers effectively and administered the company guidelines for corrective actions and improvements in customer service quality instead of the standardized normal distribution grouping calculation.

This is an open access article under the CC BY-SA license.

Keywords:
Algorithm Optimization;
Calinski-Harabasz;
Davies-Bouldin Index (DBI);
Elbow Method;
K-Mean Clustering;
Service Quality Identification;
Silhouette Score;

Article History:
Received: May 25, 2023
Revised: June 30, 2023
Accepted: July 19, 2023
Published: October 2, 2023

Corresponding Author:
Okfalisa Okfalisa
Informatics Engineering Department, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
Email: okfalisa@gmail.com

INTRODUCTION

Service quality is briefly defined as how a company meets or exceeds customer expectations [1]. Service quality also regards the service customers' experience usage value to advise the service managers to deeply focus on continuous quality management by facilitating user value creation [2]. Service quality hands over the company's suggestion to improve service quality and increase customer satisfaction [3]. Company Quality is closely related to customer satisfaction. To evaluate customer satisfaction, the company must have a performance evaluation system and performance measurement quality tools [4]. This tool provides quality analysis measurement and company performance. Thus, customer satisfaction, company reputation and trust as the key antecedents of customer loyalty and purchase intention can be controlled and enhanced [5]. Customer loyalty was the most important positive consequence of customer satisfaction [6]. The above findings reveal a significant positive relationship between customer trust and loyalty.

Meanwhile, customer satisfaction and perception of service quality highly contribute to indistinguishable loyalty predictors, directly influencing perceived value, corporate image,

company reputation, and customer loyalty [7]. Besides, customer satisfaction had a significant and sizeably positive effect on return intention [8].

Dealers are companies or individuals engaged in the liaison between producers and consumers. In this study, the intended dealer is limited to a place to sell vehicles and provide consumer services such as maintenance and providers of vehicle products. To be sustained, the dealers need to maintain relationships with consumers; thus, it impacts the customer satisfaction demand. The main dealer controlling the service and maintenance system always scrutinizes whether each branch dealer has met the consumer service standards and qualities. Therefore, this study tries to identify the service quality in maintaining customer satisfaction as one of the company services to ensure that consumers obtain similar treatments at every dealer without any differences.

Clustering, classification and prediction in various disciplines have developed and shown significant contributions to fields of knowledge, including medicine, finance, quality management, industry, technology, and even molecular biology and bioinformatics [9]. K-Means is one of the most popular unsupervised machine learning algorithms that solve clustering identification problems [10] and [11]. The K-Means, an iterative clustering analysis algorithm, effectively solves and determines the complicated number of clusters and initializes the central cluster sensitively [14] and [12]. K-Means is a simple clustering algorithm, easy to be implemented, relatively fast operation and efficient. Regarding on the centroid and connectivity models, Karthikeyan et al. (2020) found the high performed of K-Means in terms of execution time and memory for small and large dataset rather than agglomerative hierarchical [13]. Due to the low complexity of K-Means, Wiharto and Esti (2020) identified the affection of faster computing compared to Fuzzy CM especially for imaginary detection process [14]. Zhou and Yang (2020) resumed that for data sets with significant uneven distributions, *k*-means clustering is a better choice compared with FCM clustering [15].

Unfortunately, several weaknesses were found in the K-Mean localized optimization technique: sensitive selection of the cluster midpoints, thus impacting the high errors and poor clustering results. Previous studies have examined the optimizing number of clusters by K-Mean, thus improving the clustering performance. A noise-based K-Means clustering algorithm has been effectively proposed for the problems solving schema for K-Mean clustering limitation in difficulty determining the number of clustering and the initialization sensitivity of the clustering center [12]. Adopting the Elbow testing approach on K-Means contributes to the subjective and optimal estimating numbers of dataset analysis. The clear and unclear elbows' position on the line graph impacts the estimation of optimal numbers of clusters, whether in the high probability or fails to work properly [16]. Besides the Elbow testing approach, Sum Square of Error (SSE), Runtime, silhouette coefficient (SC), Calinski–Harabasz index (CHI), and Davies–Bouldin index (DBI) are commonly applied to evaluate the performance of K-Mean clustering, for examples Chang et al., (2020) analyze the university students' behavior based on a fusion K-Means clustering algorithm. They found the effect and operating efficiency of K-Mean clustering through the running time and values of SC, CHI, DBI, and SSE [17].

Moreover, the Davies Bouldin index has successfully identified the optimal number of clusters in measuring the performance of automatic data clustering using hybrid chaos game optimization with particle swarm optimization algorithm [18]. Uday et al. (2020) reveals k-means clustering algorithm can be significantly improved by using a better initialization technique such as Davies–Bouldin Index (DBI) [19]. To support this, Yuan (2021) successes improved K-means clustering algorithm for global earthquake catalogs and earthquake magnitude prediction. The sum of squares error, Davies–Bouldin index, Calinski–Harabasz index, and silhouette coefficient are applied to determine the number of clusters. [20]. Huiling (2022) have successfully determine the optimal number of clusters with Improved k value based on the Elbow Rule [21]. Viloria and Lezama (2019) enhanced the efficacy of the Automatic Clustering using Differential Evolution (ACDE) approach using the U Control Chart (UCC). The results show that the proposed method yields excellent performance compared to prior research for most datasets with optimal cluster number yet lowest DBI and Cosine Similarity (CS) measure. Thus, the advancement of this approach can determine the k activation threshold in ACDE that caused effective determination of the cluster number for k-means clustering [22]. Jun et. al (2020) applied the elbow method which is generally used to determine the best k value. The relationship curve between SSE and k is angled, and the value of k that corresponds to this angle becomes the actual cluster numbcser of the data [23]. These above studies trigger the application of K-Means for the classification of dealer service identification case with deep analysis evaluation viz., Calinski-Harabasz, Elbow, Silhouette Score,

and DBI in order to enhance the efficacy and performance of K-Means clustering.

METHOD

Data Collection

Data was collected through thorough literature studies and several interviews at dealers' companies under the Network Operational Standards section by a Network Caring Development Officer and the head of the assessor's team. The interviewees agree that identifying service quality significantly improves the company's quality and brand vision. Therefore, starting in 2019, the company has been conducting the performance measurement qualification using a separate section of services assessment.

1. Call Survey (CS). The assessment uses telephone calls to appraise the service providers compared to actual service users [24] and [25].
2. Mystery Call (MC) assesses the customer service staff's work and services, starting from answering the customer calls up to ending calls, viz., dealer operation, hospitality, product knowledge, and selling skills [26] and [27].
3. Mystery Shoppers (MS), commonly called anonymous, silent, or secret, visit service points or shops, pretend to be regular customers, observe the service delivery process, and, immediately after the service interaction, record their observations on various aspects of the service experience on a detailed questionnaire, such as exterior, interior, material promotions, parking area, attribute interior, unit display, riding test unit, negotiation and dealing, customer facilities, ending buying, payment, customer services uniform, security, greeter, sales counter, booking services, service advisor area, front desk area, waiting room, mechanic area, final confirmation, front desk officer, admin Customer Relationship Management (CRM), service adviser, cashier, mechanics, and parts counter [13, 28–30].
4. Result in the audit of Network operations standards (NOS), as an assessment technique that evaluates and monitors the entire aspects of customer service quality carried out by a team of assessors [31] and [32].

The service assessment above is then manually calculated using the company standard formula. Hence, it transforms into five service quality grading result classification using the standard normal distribution scale viz., Bronze (0-49.9%), Silver (50-69.9%), Gold A (70-79.9%), Gold B (80-89.9%), and Platinum (90-100%). The measurement of dealer service quality is then performed and mapped per service quality assessment section with no resume calculation for the entire sections, for example Dealer 1 is measured as Platinum for Call Survey (CS), Gold B for Mystery Shoppers (MS), Gold A for Mystery Call (MC), and Silver for NOS.

Herein, we resumed the above assessment results and transformed them into one dataset to identify the quality of dealer service in the data mining platform with selected attributes, including dealer name, CS, MC, MS, and NOS values from semesters 1 to 4. The list of Knowledge Discovery Data (KDD) stages is conducted, viz. selection, pre-processing, and transformation and [9]. As a result, several necessary features with various data types from the interviews and services assessments are obtained and cleaned from 205 into 193 data. Then, this data is ready to be analyzed with K-Mean clustering mining. The stages process of K-Mean clustering mining can be depicted in Figure 1.

K-Means Clustering

The k-Means algorithm is a simple iterative clustering algorithm that uses the metric distance and the k classes provided by the data set, for next to calculate the average distance and equip the initial center of the class described by the center of mass [25, 27, 33]. However, several drawbacks of K-Means are found, especially in data categorical, whereby the most frequent value represents the k-mode centroid and needs to be initially determined by the user. Determining the optimal number of clusters is challenging, especially for data sets with less prior knowledge. A reasonable percentage of the partition clustering algorithm must define the number of clusters as the input parameter before training randomly to find the optimized grouping results [16, 17, 23], and [13]. Besides, this algorithm is highly sensitive to outliers. The K-Means steps are as follows [29].

1. Determining the value of k.
2. Calculating the distance from the data to the centroid using the Euclidean Distance formula below.

$$d(x_i, \mu_i) = \sqrt{(xi - \mu i)^2} \quad (1)$$

3. Grouping the data based on minimum distance.
4. Updating the centroid value with the formula is as follows.

$$C_k = \frac{1}{nk} \sum d_i \quad (2)$$

Where:
nk = the amount of data in the cluster
di = the sum of the values included in each cluster

5. Repeating the iteration until the cluster is stable.

Next, K-Mean mining has put a value using several testing approaches, including Elbow, Silhouette, Calinski-Harabasz Index, and Davies-Bouldin Index (DBI).

Elbow Method

The Elbow Method is one of the most commonly used methods to distinguish the optimal number of clusters through the elbow points identification on the curve visualization [16]. Herein, SSE for each cluster is calculated and compared to optimize the number of clusters [31]. The elbow method goes through the graph of the excellent k value with the position on the elbow along with the SSE (less than 1). The best k-cluster results will be the basis for data grouping. The smaller the SSE's value and the angled graph, the better the cluster results [23].

$$SSE = \sum_{i=1}^{n} (d_i)^2 \tag{3}$$

Where:
d = distance between data and cluster center

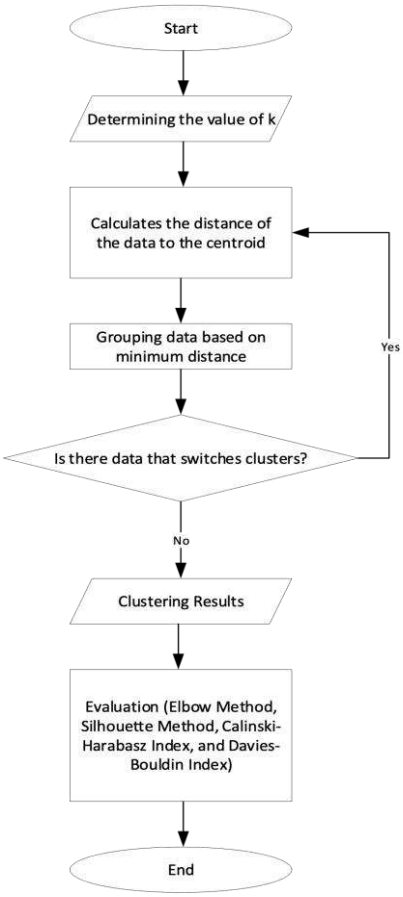

Figure 1. Flow Process K-Means Mining

Silhouette Method

The Silhouette approach calculates the silhouette coefficient numbers combine with the separation and cohesion. A higher Silhouette coefficient is a better cluster provided [32].

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b(i) - a(i)}{max\{a(i), b(i)\}}\right) \tag{4}$$

Where:
a(*i*) = the average distance of sample *i* to other samples in the cluster
b(*i*) = minimum sample distance from sample *i* to other clusters

Calinski-Harabasz Index (CHI)

The Calinski-Harabasz Index (CHI) evaluates cluster validity based on the Inter-Cluster Sum of Squares (BSS) and Within-groups sum of squares (WSS) calculations. CHI measures the separation ratio based on the maximum distance between the centroids and the compactness based on the sum of the distances between each data and the centroids. A compact and well-separated cluster configuration is expected to have high inter-cluster variance and relatively low intra-cluster variance [34].

$$CH(K) = \frac{B(K)(N-K)}{W(K)(K-1)} \tag{5}$$

Where:

$$B(K) = \left(\sum_{k=1}^{K} a_k \left\|\bar{x}_k - \bar{x}\right\|^2\right) \tag{6}$$

$$W(K) = \left(\sum_{k=1}^{K} \sum_{C(j)=K} \left\|x_j - \bar{x}_k\right\|^2\right) \tag{7}$$

Where:
K as the appropriate number of clusters, B(K) as inter-cluster divergence, also called inter-cluster covariance, and W(K) as intra-cluster divergence, also called intra-cluster covariance.

Davies-Bouldin Index (DBI)

One of the clustering validation techniques is the Davies Bouldin Index technique. This technique will identify the cohesion matrix (closeness of one group) and the separation matrix (differences between groups). The smaller, the better the resulting Davies Bouldin Index (DBI) value [17].

$$DBI = \frac{1}{K}\sum_{i}^{k} max_{i \neq j} R(i,j) \tag{9}$$

Where:
K = existing clusters
Ri,j = ratio between clusters i and j

max = find the ratio between the largest cluster.

RESULTS AND DISCUSSION

K-Means Clustering

Following the pre-processing stages in K-Mean clustering, the 205 data is reduced into 193 raw datasets; thus, it was filtered into nine features and values calculation by service assessment, including CS1 (First semester Calls Survey), CS2 (Second semester Calls Survey), MC1 (First quarter Mystery Call), MC2 (Second quarter Mystery Call), MC3 (Third quarter Mystery Call), MC4 (Fourth quarter Mystery Call), MS1 (First quarter Mystery Shopping), MS2 (Second quarter Mystery Shopping), and NOS (Result audit of NOS). A detailed sample of raw data can be depicted in Table 1. As a result of the K-Mean algorithm tracking process and calculation at (1) and (2), Table 2 is captured to show the clustering identification process from k=1 to 10.

Evaluation of Optimum K-Values

Ensuring the formula of the equation for Elbow, Silhouette, and HCI, the maximum index values are described in Table 3.

Table 1. The Sample of Raw Data Set Values

Nama Dealer	CS1	CS2	MC1	MC2	MC3	MC4	MS1	MS2	NOS
Dealer 1	0,9	0,9	0,2	0,7	0,7	0,5	0,7	0,7	0,6
Dealer 2	0,9	0,9	0,5	0,7	0,6	0,9	0,7	0,7	0,7
Dealer 3	0,8	0,9	0,7	0,6	0,7	0,5	0,7	0,7	0,6
Dealer 4	0,8	0,9	0,6	0,6	0,7	0,8	0,7	0,7	0,6
Dealer 5	0,8	0,8	0,3	0,6	0,8	0,7	0,7	0,7	0,6
…	...	...	...	...	...	...	...	...	...
…	...	...	...	...	...	...	...	...	...
Dealer 191	0,0	0,0	0,6	0,1	0,1	0,4	0,0	0,5	0,3
Dealer 192	0,0	0,0	0,6	0,2	0,6	0,7	0,0	0,5	0,3
Dealer 193	0,0	0,0	0,4	0,7	0,5	0,2	0,0	0,5	0,3

Table 2. Cluster Identification Based on K-Value

Nama Dealer	CS 1	CS 2	MC 1	MC 2	MC 3	MC 4	MS 1	MS 2	NO S	k= 2	k= 3	k= 4	k= 5	k= 6	k= 7	k= 8	k= 9	k=1 0
Dealer 1	0,9	0,9	0,2	0,7	0,7	0,5	0,7	0,7	0,6	1	1	1	0	0	6	4	0	2
Dealer 2	0,9	0,9	0,5	0,7	0,6	0,9	0,7	0,7	0,7	1	1	1	1	2	0	0	0	1
Dealer 3	0,8	0,9	0,7	0,6	0,7	0,5	0,7	0,7	0,6	1	1	1	1	2	0	3	0	1
Dealer 4	0,8	0,9	0,6	0,6	0,7	0,8	0,7	0,7	0,6	1	1	1	1	2	0	0	0	1
Dealer 5	0,8	0,8	0,3	0,6	0,8	0,7	0,7	0,7	0,6	1	1	1	1	0	6	4	0	2
…	...	...	...	...	...	...	...	...	...	…	...	...	...	...	...	...	...	...
…	...	...	...	...	...	...	...	...	...	…	...	...	...	...	...	...	...	...
Dealer 191	0,0	0,0	0,6	0,1	0,1	0,4	0,0	0,5	0,3	0	2	3	4	5	4	1	4	6
Dealer 192	0,0	0,0	0,6	0,2	0,6	0,7	0,0	0,5	0,3	0	2	3	4	5	5	1	5	6
Dealer 193	0,0	0,0	0,4	0,7	0,5	0,2	0,0	0,5	0,3	0	2	3	4	5	5	1	5	6

Table 3. The Comparison of k Evaluation

Number of Clusters	SSE	CHI	Silhouette coefficient	DBI
2	145,12	86.06	0,325	1,405
3	100,04	85.35	0,366	1,038
4	76,44	75.76	0,229	1,459
5	65,88	69.37	0,244	1,320
6	58,73	67.55	0,265	1,180
7	51,74	64.58	0,270	1,211
8	47,09	60.22	0,285	1,302
9	44,45	56.25	0,240	1,351
10	41,89	53.48	0,214	1,478

Table 3 explains that the SSE value of the Elbow Graph for the third cluster is the most significant, with a k value of 100.04. The CHI index defines the optimum value of 85.35 at the third cluster. Meanwhile, the Silhouette coefficient analyzes the optimum k=3 at 0.366. These real results are also supported by DBI's excellent value at k=3 with a value of 1.038. In a nutshell, these evaluations confirm the optimum cluster values at k=3.

Clustering Result

Proceeding the K-Mean calculation with optimum identification of k value at 3, Table 4 and Figure 2 are defined. Herein, dealers' data are grouped into three clusters: cluster 1 for 38 data, cluster 2 and 3 for 125 and 30 data, respectively. The mapping cluster analysis for clusters 1, 2 and 3 are explained in Table 5. Whereby cluster 1 is good, cluster 2 is very good, and cluster 3 is not a good performance. The mean analysis of the Dealer s' performance per variable can be depicted in Figure 3.

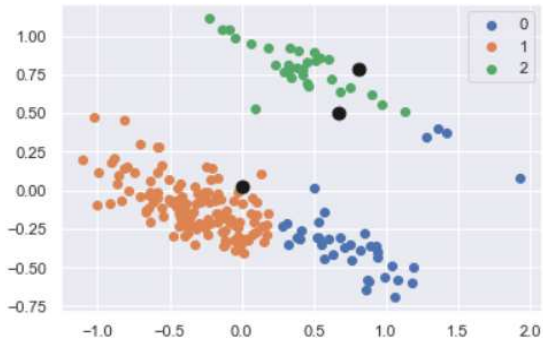

Figure 2. K-Means Clustering Map for k=3

Table 4. Cluster Identification for k=3

Dealer Name	CS1	CS2	MC1	MC2	MC3	MC4	MS1	MS2	NOS	CLUSTER
Dealer 1	0,9	0,9	0,2	0,7	0,7	0,5	0,7	0,7	0,6	C2
Dealer 2	0,9	0,9	0,5	0,7	0,6	0,9	0,7	0,7	0,7	C2
Dealer 3	0,8	0,9	0,7	0,6	0,7	0,5	0,7	0,7	0,6	C2
Dealer 4	0,8	0,9	0,6	0,6	0,7	0,8	0,7	0,7	0,6	C2
Dealer 5	0,8	0,8	0,3	0,6	0,8	0,7	0,7	0,7	0,6	C2
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
Dealer 191	0,0	0,0	0,6	0,1	0,1	0,4	0,0	0,5	0,3	C3
Dealer 192	0,0	0,0	0,6	0,2	0,6	0,7	0,0	0,5	0,3	C3
Dealer 193	0,0	0,0	0,4	0,7	0,5	0,2	0,0	0,5	0,3	C3

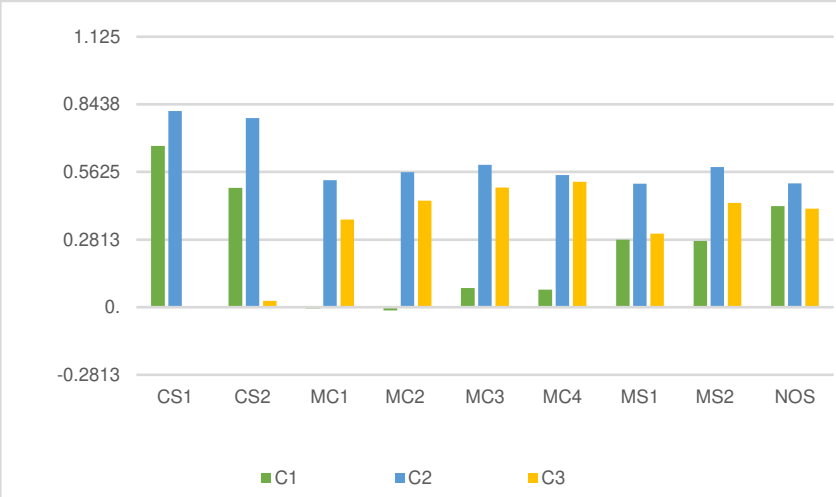

Figure 3. Dealer Performance Analysis per Variables

Table 5. Cluster Mapping Analysis for k=3

Dealer Name	CS1	CS2	MC1	MC2	MC3	MC4	MS1	MS2	NOS	CLUSTER
Dealer 1	0,9	0,9	0,2	0,7	0,7	0,5	0,7	0,7	0,6	1
Dealer 2	0,9	0,9	0,5	0,7	0,6	0,9	0,7	0,7	0,7	1
Dealer 3	0,8	0,9	0,7	0,6	0,7	0,5	0,7	0,7	0,6	1
Dealer 4	0,8	0,9	0,6	0,6	0,7	0,7	0,7	0,7	0,6	1
Dealer 5	0,8	0,8	0,3	0,6	0,8	0,7	0,7	0,7	0,6	1
…	...	...	...	...	...	...	...	...	...	…
…	...	...	...	...	...	...	...	...	...	…
Dealer 191	0	0	0,6	0,1	0,1	0,4	0	0,5	0,3	2
Dealer 192	0	0	0,6	0,2	0,6	0,7	0	0,5	0,3	2
Dealer 193	0	0	0,4	0,7	0,5	0,2	0	0,5	0,3	2

Figure 3 describes that, generally, the performance of CS1 and CS2 point out at Cluster 1 and 2 (Good and very good performance). This indicates that the customers are satisfied with the service providers' assistance with products, features, promotions, communities, vehicle services explanation, and the flexibility of the payment. Meanwhile, the performance of MC, MS, and NOS is still in the balance between good/very good and not good. This specifies that the dealers should pay more attention to the service quality at dealer operation, hospitality, product knowledge, and selling skill. Herein, this study reveals the dealers' clustering analysis by considering CS1, CS2, MC, MS, and NOS performance variables as the analysis methods of service quality measurement at points of sale or customer services. Besides, the study found that MS or Mystery Shoppers and NOS (Network operations standards) audit performs a similar cluster result, as presented in Figure 3. This bears out the statement of [24][26] that MS is one of the most widely used tools to monitor the quality of service and personal selling and an effective way of testing service provision and predicting customer satisfaction and sales performance. MS technique reports back to organizations on the communicative competencies of call-takers; their report is based on a seriously constrained notion of their regular activity on the phone and the range of contingencies they must deal with [26][27].

Meanwhile, comparing to the previous company measurement with five grading result calculation, this K-means classification with three optimum classes proposes high valuable data distribution and successfully performed measurement per dealer's service assessment variable and general performance. Thus, the quality measurement can be enhanced and more accurate.

CONCLUSION

This research has successfully identified the dealer service quality. Thus, the company can maintain customer satisfaction by intensifying the dealer service performance. The optimum number of quality performance clusters is proposed by adopting the K-Means algorithm with four evaluation techniques: Elbow Method, Silhouette Score, Davies-Bouldin Index (DBI), and the Calinski-Harabasz index. As a result, k values at three are the well-separated and precise distribution shapes in grouping the dealer quality service performance into cluster one for good, cluster two and three for very good and not good performance, respectively. This dealer grouping proposes novel and auspicious service quality performance schema comprehensively instead of mathematical normal distribution standard. Besides the dealer quality service clustering, this result also shows that Mystery Shoppers (MS) is an effective tool in monitoring and testing service provision and predicting customer satisfaction and sales performance. Therefore, these proposed results can be used as a guideline for the management level in improving performance measurement tools and evaluation schema, dealer performance and sales service quality.

ACKNOWLEDGMENT

This research was supported by Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia and Honda Motorbike Dealer at Pekanbaru, Indonesia, for data collection. We also thank our colleagues Lestari Handayani and Yusra from Universitas Islam Negeri Sultan Syarif Kasim Riau, who provided insights and expertise that greatly assisted and enriched this research.

REFERENCES

[1] M. Pakurár, H. Haddad, J. Nagy, J. Popp, and J. Oláh, “The service quality dimensions that affect customer satisfaction in the Jordanian banking sector,” *Sustainability (Switzerland)*, vol. 11, no. 4, Feb. 2019, doi: 10.3390/su11041113.

[2] G. Medberg and C. Grönroos, “Value-in-use and service quality: do customers see a difference?” *Journal of Service Theory and Practice*, vol. 30, no. 4–5, pp. 507–529, Nov. 2020, doi: 10.1108/JSTP-09-2019-0207.

[3] G. Chandirasekaran, P. Mannadhan, and N. N. Abdullah, “Business Analytics and Supply Chain Performance: Partial Least Squares-Structural Equation Modeling (PLS-SEM) Approach,” *International journal of management and business research*, vol. 9, no. 1, pp. 91–96, 2019.

[4] D. Sumrit and K. Sowijit, “Winning customer satisfaction toward omnichannel logistics service quality based on an integrated importance-performance analysis and three-factor theory: Insight from Thailand,” *Asia Pacific Management Review*, 2023, doi: 10.1016/j.apmrv.2023.03.003.

[5] T. Islam *et al.*, “The impact of corporate social responsibility on customer loyalty: The mediating role of corporate reputation, customer satisfaction, and trust,” *Sustainable Production and Consumption*, vol. 25, pp. 123–135, Jan. 2021, doi: 10.1016/j.spc.2020.07.019.

[6] R. P. Rajendran and J. Suresh, “Customer satisfaction index as a performance evaluation metric: A study on indian e-banking industry,” *Article in International Journal of Business*, vol. 22, no. 3, p. 2017, 2017.

[7] P. Özkan, S. Süer, İ. K. Keser, and İ. D. Kocakoç, “The effect of service quality and customer satisfaction on customer loyalty: The mediation of perceived value of services, corporate image, and corporate reputation,” *International Journal of Bank Marketing*, vol. 38, no. 2, pp. 384–405, Apr. 2020, doi: 10.1108/IJBM-03-2019-0096.

[8] A. A. Salamah, Shahizan,H., Ali, A., Walaa A, Z., Mohammad, A., et al.,“Customer retention through service quality and satisfaction: using hybrid SEM-neural network analysis approach,” *Heliyon*, vol. 8, no. 9, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10570.

[9] Okfalisa, Budianita, E., Yuliani, R. , Suanmali, L., Megawati, Rusnedy, H., and Saktioto, “Forecasting company financial distress: C4.5 and adaboost adoption,” *Engineering and Applied Science Research*, vol. 49, no. 3, pp. 300–307, 2022, doi: 10.14456/easr.2022.31.

[10] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, “A novel K-means clustering algorithm with a noise algorithm for capturing urban hotspots,” *Applied Sciences (Switzerland)*, vol. 11, no. 23, Dec. 2021, doi: 10.3390/app112311202.

[11] K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[12] F. H. Awad and M. M. Hamad, “Improved k-Means Clustering Algorithm for Big Data Based on Distributed SmartphoneNeural Engine Processor,” *Electronics (Basel)*, vol. 11, no. 6, 2022, doi: 10.3390/electronics11060883.

[13] K. B, D. J. George, G. Manikandan, and T. Thomas, "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 1600–1604, May 2020, doi: 10.30534/ijeter/2020/20852020.

[14] W. Wiharto and E. Suryani, "The comparison of clustering algorithms K-means and fuzzy C-means for segmentation retinal blood vessels," *Acta Informatica Medica*, vol. 28, no. 1, pp. 42–47, Mar. 2020, doi: 10.5455/AIM.2020.28.42-47.

[15] K. Zhou and S. Yang, "Effect of cluster size distribution on clustering: a comparative study of k-means and fuzzy c-means clustering," *Pattern Analysis and Applications*, vol. 23, no. 1, pp. 455–466, 2020, doi: 10.1007/s10044-019-00783-6.

[16] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13638-021-01910-w.

[17] W. Chang *et al.*, "Analysis of university students' behavior based on a fusion K-means clustering algorithm," *Applied Sciences (Switzerland)*, vol. 10, no. 18, Sep. 2020, doi: 10.3390/APP10186566.

[18] M. W. Ouertani, G. Manita, and O. Korbaa, "Automatic Data Clustering Using Hybrid Chaos Game Optimization with Particle Swarm Optimization Algorithm," in *Procedia Computer Science*, Elsevier B.V, 2022, pp. 2677–2687, doi: 10.1016/j.procs.2022.09.326.

[19] G. Uday Kiran and D. Vasumathi, "Disease Detection using Enhanced K-Means Clustering and Davies-Bould in Index in Big Data-Safe Engineering Application," *Journal of Green Engineering (JGE)*, vol. 10, no. 12, pp. 13089–13106, 2020.

[20] R. Yuan, "An improved K-means clustering algorithm for global earthquake catalogs and earthquake magnitude prediction," *Journal of Seismology*, vol. 25, no. 3, pp. 1005–1020, 2021, doi: 10.1007/s10950-021-09999-8.

[21] H. Zhao, "Design and Implementation of an Improved K-Means Clustering Algorithm," *Mobile Information Systems*, vol. 2022, p. 6041484, 2022, doi: 10.1155/2022/6041484.

[22] A. Viloria and O. B. P. Lezama, "Improvements for determining the number of clusters in k-means for innovation databases in SMEs," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 1201–1206. doi: 10.1016/j.procs.2019.04.172.

[23] J. Wu, Li Shi, Wen-Pin Lin, Sang-Bing Tsai, Yuanyuan Li, Liping Yang, and Guangshu Xu, "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm," *Mathematical Problems in Engineering*, vol. 2020, p. 8884227, 2020, doi: 10.1155/2020/8884227.

[24] G. Blessing and M. Natter, "Do Mystery Shoppers Really Predict Customer Satisfaction and Sales Performance?" *Journal of Retailing*, vol. 95, no. 3, pp. 47–62, Sep. 2019, doi: 10.1016/j.jretai.2019.04.001.

[25] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J (Basel)*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.

[26] E. Stokoe, R. O. Sikveland, S. Albert, M. Hamann, and W. Housley, "Can humans simulate talking like other humans? Comparing simulated clients to real customers in service inquiries," *Discourse Studies*, vol. 22, no. 1, pp. 87–109, Feb. 2020, doi: 10.1177/1461445619887537.

[27] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8. MDPI AG, pp. 1–12, Aug. 01, 2020. doi: 10.3390/electronics9081295.

[28] C. J. Berg *et al.*, "Exploring the Point-of-Sale Among Vape Shops Across the US: Audits Integrating a Mystery Shopper Approach," *Nicotine & Tobacco Research*, 2020, doi: 10.1093/ntr/ntaa041/5801058.

[29] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.

[30] O. Okfalisa, A. Angraini, S. Novi, H. Rusnedy, L. Handayani, and M. Mustakim, "Identification of the distribution village maturation: Village classification using Density-based spatial clustering of applications with noise," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 3, pp. 133–141, Jul. 2021, doi: 10.14710/jtsiskom.2021.13998.

[31] A. Et-Taleby, M. Boussetta, and M. Benslimane, "Faults detection for photovoltaic field based on k-means, elbow, and average silhouette techniques through the segmentation of a thermal image,"

International Journal of Photoenergy, vol. 2020, 2020, doi: 10.1155/2020/6617597.

[32] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.

[33] S. Saifullah, "K-Means Clustering for Egg Embryo's Detection Based-On Statistical Feature Extraction Approach of Candling Eggs Image," *SINERGI*, vol. 25, no. 1, p. 43, Nov. 2020, doi: 10.22441/sinergi.2021.1.006.

[34] S. P. Lima and M. D. Cruz, "A genetic algorithm using Calinski-Harabasz index for automatic clustering problem," *Revista Brasileira de Computação Aplicada*, vol. 12, no. 3, pp. 97–106, Sep. 2020, doi: 10.5335/rbca.v12i3.11117.