JOIV : Int. J. Inform. Visualization, 7(3) - September 2023 871-877

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Comparison between K-Means and K-Means++ Clustering Models Using Singular Value Decomposition (SVD) in Menu Engineering

Dwi Hosanna Bangkalang [a], Nina Setiyawati [b,*], Hindriyanto Dwi Purnomo [b]

[a] *Department of Information System, Faculty of Information Technology, UKSW, Salatiga, Indonesia*
[b] *Department of Informatics Engineering, Faculty of Information Technology, UKSW, Salatiga, Indonesia*
*Corresponding author: *nina.setiyawati@uksw.edu*

***Abstract*— The menu is one of the most fundamental aspects of business continuity in the culinary industry. One of the tools that can be used for menu analysis is menu engineering. Menu engineering is an analytical tool that assists restaurants, companies, and small and medium-sized enterprises (SMEs) in assessing and making decisions on marketing strategies, menu design, and sales so that it can produce maximum profit. In this study, several menu engineering models were proposed, and the performance of these models was analyzed. This study used a dataset from the Point of Sales (POS) application in an SME engaged in the culinary field. This research consists of three stages. First, pre-processing the data, comparing the models, and evaluating the models using the Davies Bouldin index. At the model comparison stage, four models are being compared: K-Means, K-Means++, K-Means using Singular Value Decomposition (SVD), and K-Means++ using SVD. SVD is used in the dataset transformation process. K-Means and K-Means++ algorithms are used for grouping menu items. The experiments show that the K-Means++ model with SVD produced the most optimal cluster in this research. The model produced an average cluster distance value of 0.002; the smallest Davies-Bouldin Index (DBI) value is 0.141. Therefore, using the K-Means++ model with SVD in menu engineering analysis produces clusters containing menu items with high similarity and significant distance between groups. The results obtained from the proposed model can be used as a basis for strategic decision-making of managing price, marketing strategy, etc., for SMEs, especially in the culinary business.**

***Keywords*— Menu engineering; K-Means; K-Means++; Singular Value Decomposition Davies-Bouldin Index; SME; cluster.**

Manuscript received 27 Jul. 2022; revised 24 Jan. 2023; accepted 26 Mar. 2023. Date of publication 10 Sep. 2023.

I. INTRODUCTION

The menu is one of the most fundamental aspects of business continuity in the culinary industry [1], [2]. A well-designed and managed menu could generate greater profit for the business and provide product information to consumers [2], [3]. The menu belongs to the marketing tools that require a design and strategy regarding cost and price structure [4]. In the Restaurant Revenue Management (RRM) framework, one of the tools that can be used to perform menu management and analysis is the Menu Engineering model [5].

Menu engineering is an analytical technique that helps companies or SMEs evaluate and make decisions regarding marketing strategies, menu design, and sales [5]–[8]. Menu Engineering is a menu analysis model introduced by Kasavana and Smith [9]. The main idea of the analysis model is that the higher the item's contribution margin, the more the item needs to be sold [9]. The analysis model groups the menu items into four categories: Star, Plowhorse, Puzzle, and Dog. The menu engineering matrix [9], [10] can be seen in Fig. 1.

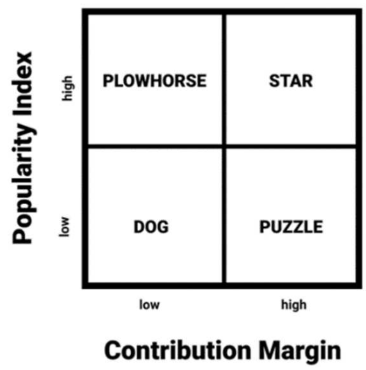

Fig. 1 Menu Engineering Matrix

The analysis technique in the engineering menu focuses on three elements [11]: customer demand, menu mix analysis, and item contribution margin. Customer demand is the total menu sold, while the menu mix is the level of popularity of each menu item, and the contribution margin is the difference between the selling price of the menu and the cost of goods sold. There are several studies on menu engineering. Tom and Annaraud [12] proposed nine-quadrant menu engineering by applying fuzzy multi-criteria decision-making (FMCDM). The study grouped menu items using the elements of sales volume, cost of goods sold, and contribution margin. The research results stated that using FMCDM in the engineering nine quadrant menu increased the average menu sales. Setiyawati and Bangkalang [13] proposed an engineering menu model by adding categorical and revenue variables as additional analysis variables. The study developed four clusters, each with a different popularity and contribution margin.

The cluster output is used to create strategic recommendations for the menu to increase sales. Setiyawati [1] also proposed an engineering menu model using the K-Nearest Neighbors Algorithm. The findings of this study suggested a new approach to engineering menu analysis using a classification of menu items according to the characteristics of the menu mix and item contribution margin. Each menu item analyzed was classified into the Star, Plowhorse, Puzzle, and Dog item groups. Each item group can be given strategic recommendations in managing prices, menu marketing, etc. Similar to previous studies, in this study, menu engineering modeling was carried out to produce menu category groupings of items that have high similarity.

In this research, several menu engineering models are proposed, and the performance of these models is analyzed. The models are based on the K-Means and K-Means++ algorithms. K-Means is a method of distance-based clusters [14], [15] and center-based clusters [16]–[18], so the calculation of the closest distance and the determination of the initial centroid will affect the cluster formed [17], [19]. One of the drawbacks of K-Means is the random determination of the centroid initials. Therefore, D^2 weighting on K-Means++ [20] is proposed as a probabilistic method for determining the initials of the centroid cluster.

Based on previous research, the data transformation process is one of the determining processes in producing optimal clusters with high member similarity. In order to increase the variance of cluster members with similar characteristics, the dataset transformation process in the proposed models is optimized using Singular Value Decomposition (SVD) [21]. Compared to previous studies' proposed menu engineering process model, the data transformation still produces a wide distribution of data. Therefore, the proposed use of SVD in the normalization process in this study is expected to produce more optimal values in data pre-processing.

II. Material and Method

This study used a dataset from the Point of Sales (POS) application in an SME engaged in the culinary field. Data from the database was extracted to obtain a dataset according to the required variables. The research steps used are described in Fig. 2.

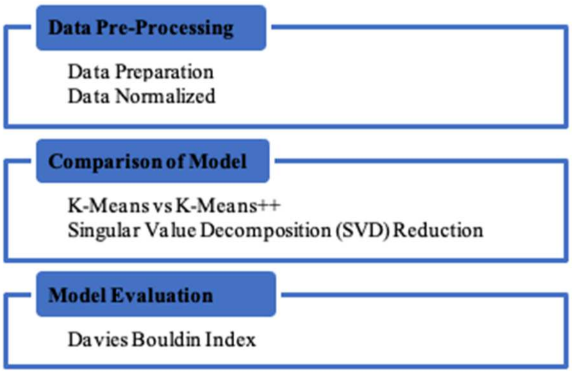

Fig. 2 Research Step

A. Data Pre-Processing

The data used in this study consisted of 73,000 transactions involving 120 menu items. The transaction details were then extracted using SQL queries to acquire three variable values, namely popularity (1), contribution margin (2), and revenue (3), with the following formula.

$$MM = \frac{number\ of\ the\ item\ sold}{total\ number\ of\ the\ item\ sold} \times 100\% \quad (1)$$

$$CM = Item\ price - food\ cost \quad (2)$$

$$R = number\ sold\ \times\ sell\ price \quad (3)$$

Where:
MM = *Popularity*
CM = *Contribution Margin*
R = *Revenue*

Before the data is clustered because distance-based clusters are sensitive to the scale variance of dataset items, normalization is needed [22]. Normalization is a pre-model data where attribute transformation is carried out into a specific value scale [22], [23]. The purpose of data normalization is to get standardized features, free from redundancy and noisy objects [22], [24], prevent bias in the model building, and get relevant output features [25]. The normalization technique used in this research is Z-Score Normalization. It can be formulated as in equation (4).

$$xb = \frac{xa - \bar{x}}{\sigma} \quad (4)$$

Where:
xb = *new value*
xa = *old value*
$\bar{x}$ = *average*
σ = *standard deviation*

Z-Score Normalization is a normalization technique suitable for use if the minimum and maximum values of the attribute are unknown [23]. In other words, this technique transforms the normal variance to a standard scale [22] by looking at the distribution of the mean and the distribution of the item data set [26], [27]. The distribution and value of the data used in this research after it has been standardized are illustrated in Fig. 3. The data have a minimum value of -8,724 and a maximum of 3,437.

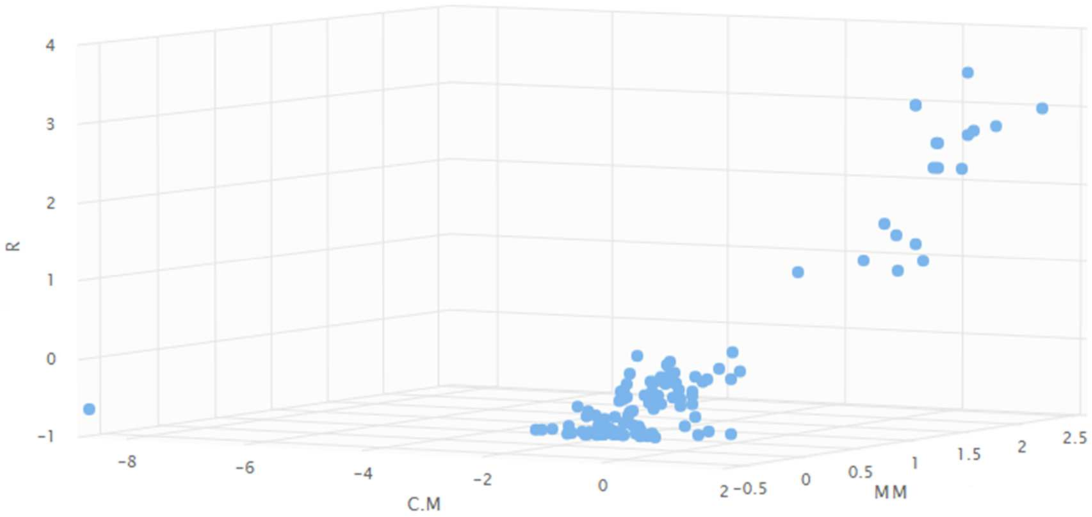

Fig. 3 Plot distribution of Z-Score Normalization items

Matrix decomposition [28] and the transformation of new features generated from original data [21], [29], [30] require the Singular Value Decomposition (SVD) method. This method has a good mechanism for producing similar value characteristics [21] and can improve the performance and normalization of the same information value [29]. The results of the normalization of variables and dimension reduction using SVD are shown in Table I.

TABLE I
THE VARIABLES VALUE AFTER NORMALIZATION AND SVD TRANSFORMATION

Item ID	MM	CM	R	SVD_1	SVD_2	SVD_3
1	-0,482	-0,353	-0,506	-0,048	-0,023	0,002
2	-0,412	-0,940	-0,385	-0,045	-0,078	-0,025
3	-0,341	0,843	-0,196	-0,017	0,083	-0,029
4	-0,419	0,256	-0,284	-0,030	0,030	-0,037
5	-0,574	0,395	-0,551	-0,048	0,047	0,001
6	-0,646	-0,054	-0,646	-0,059	0,007	0,000
7	-0,553	0,245	-0,546	-0,048	0,033	0,003
8	-0,624	-0,951	-0,626	-0,066	-0,075	-0,016
9	-0,613	1,441	-0,614	-0,042	0,143	0,028
10	-0,574	1,142	-0,522	-0,039	0,115	0,005
11	-0,482	0,056	-0,405	-0,040	0,014	-0,022
12	-0,064	-0,054	0,144	0,003	-0,005	-0,067
13	-0,447	-0,698	-0,425	-0,046	-0,056	-0,019
14	-0,482	-0,502	-0,465	-0,048	-0,037	-0,014
15	-0,341	0,843	-0,305	-0,022	0,083	0,005
16	-0,155	-0,203	-0,094	-0,013	-0,016	-0,023
17	-0,090	-0,054	0,109	0,000	-0,005	-0,064
18	-0,301	0,096	-0,178	-0,021	0,014	-0,036
19	-0,489	-0,173	-0,473	-0,045	-0,007	-0,007
20	-0,164	0,245	-0,104	-0,010	0,025	-0,014
..	..	..	..	..	..	..
119	-0,123	-0,652	-0,149	-0,018	-0,057	-0,004
120	-0,139	-0,682	-0,223	-0,023	-0,059	0,014

B. K-Means Model

K-Means [31] is a data cluster method. The main idea of K-Means is to minimize the distance [32] mean squared between objects in the same cluster category [20], [33] to minimize variance within a cluster as well as maximize variance differences between clusters [34]. K-Means is sensitive to the initial determination of the starting point [19]. To resolve the drawback of K-Means, K-Means++ was used. In K-Means++, the initial cluster was determined by a simple probabilistic method to define the initial cluster [20]. The main difference between the K-Means and the K-Means++ is at the initial centroid determination stage [20]. In K-Means++, D^2 Weighting was used. The stages in K-Means++ are described as follows:

1) Determine the Number of Clusters K
2) Determine the initial centroid for each Kn data point. D2 weighting:
   a. Choose an initial centroid c1 uniformly at random from the dataset
   b. Choose the next centroid cn, selecting cn = x' ∈ X with probability formulated as follow (5):

$$\frac{D(x')^2}{\sum_{x \in X} D(x)^2} \tag{5}$$

   Where:
   $D\ (x')^2$: Distance of Euclidean Distance
   $\sum_{x \in X} D(x)^2$: Sum distance
   c. Repeat Step 2.b until a total of K centers were chosen
3) Calculate the distance of each data object on each Kn
4) Group the data according to the closest distance to the Kn cluster
5) Calculate the new centroid for each Kn data point
6) Repeat steps 3-6 until there is no movement in the Kn cluster members

The basic concept of Menu Engineering is a group of items formed in a 2x2 matrix according to the intersection of popularity and item contribution margin [9]. The engineering menu has four categories of group items: Star, Puzzle, Plowhorse, and Dog. The number of categories is used as the number of clusters K = 4. Next, the initial centroid calculation was carried out using the D^2 weighting step. The first step is randomly determining the initial centroid of D1 from the dataset.

Furthermore, for the initials D_2 to D_n, a mathematical calculation was used through proportional probability (5) on the square of the distance above the sum of the squared distances for the previous D_n [35]. The process of determining D^2 weighting will stop if the total initial centroid is as many as K that has been selected. The initial iteration process was carried out by calculating the distance of each dataset to the centroid K_n. Several methods of calculating distance and Euclidean distance are used in this research as it has good performance in terms of iterations, total squared errors, and the time required to build the model [36]. The Euclidean distance method is formulated as shown in equation (6):

$$D(i,k) = \sqrt{(X_{1i} - X_{1j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad (6)$$

Where:
$D(i,k)$: Distance data *i* to centroid *k*
X_{ki}: Data *i* to attribute data *k*
X_{kj}: Centroid *j* to attribute data *k*

Using the Euclidean distance method (6), the distance of each D(i,k) to the Kn centroid is calculated. D(i,k) was grouped according to the closest distance on the Kn centroid. Then a new centroid calculation for Cn was carried out from the average members in the Kn cluster with the following formula (7). The final stage of K-Means is to perform iteratively until no members in the K_n cluster migrate.

$$New\ C_n = \frac{\sum D(i,k)}{n} \quad (7)$$

III. Result and Discussion

A. K-Means Model Comparison

This study compares the menu engineering model based on the K-Means and K-Means++ which were optimized using Singular Value Decomposition (SVD). K-Means is an algorithm whose cluster members are formed from the closest distance between the data and the centroid point [37], which means it is susceptible to deviations from feature calculations [23], [24]. Therefore, the use of D^2 weighting for determining the initials of the centroid cluster and data transformation using SVD in the pre-clustering process can significantly affect the model. The scatter plot cluster formed by the models can be seen in Fig. 4a-d.

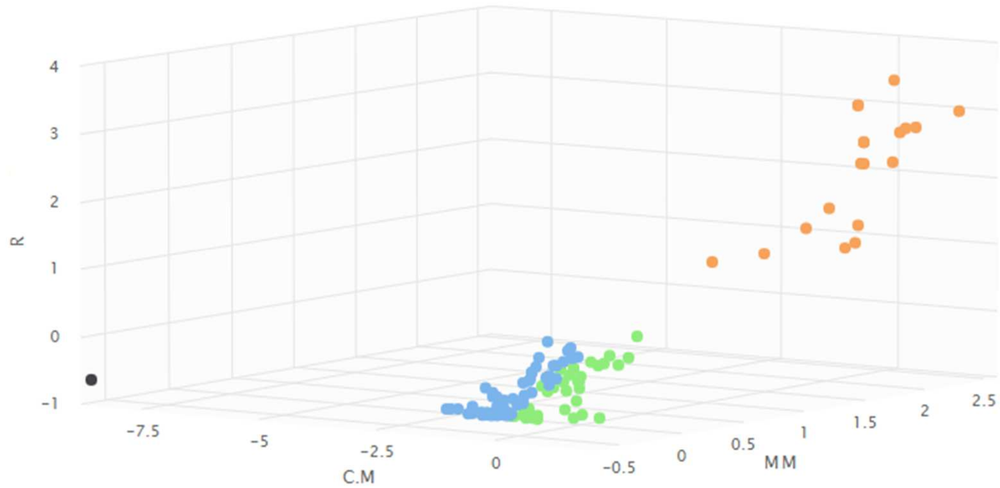

Fig. 4 (a) Scatter plot cluster K-Means

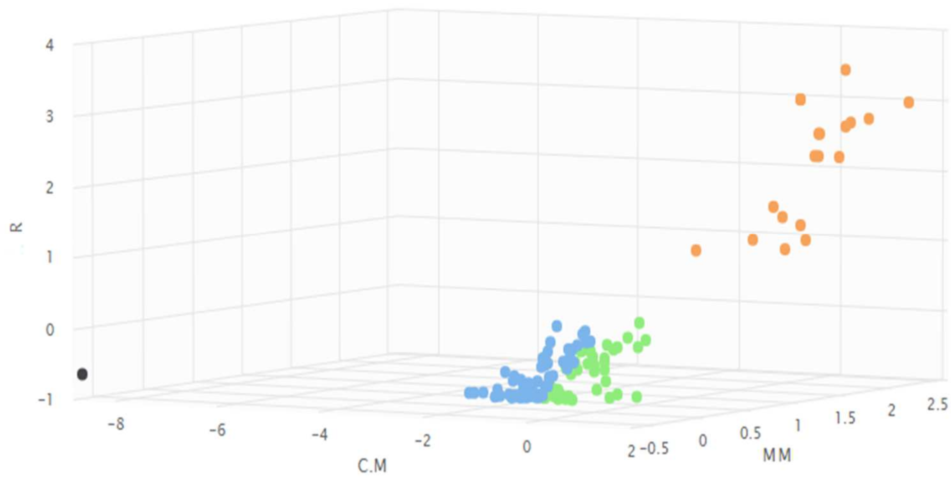

Fig. 4 (b) Scatter plot cluster K-Means++

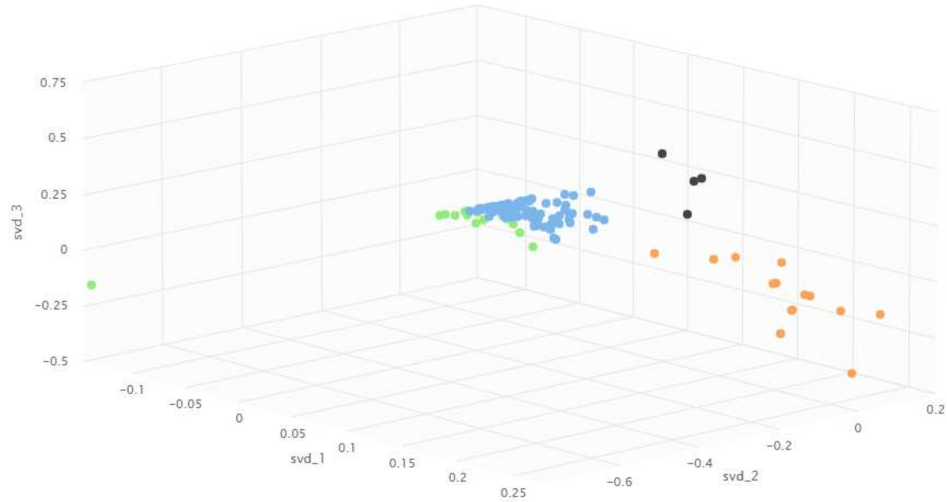

Fig. 4 (c) Scatter plot cluster K-Means using SVD

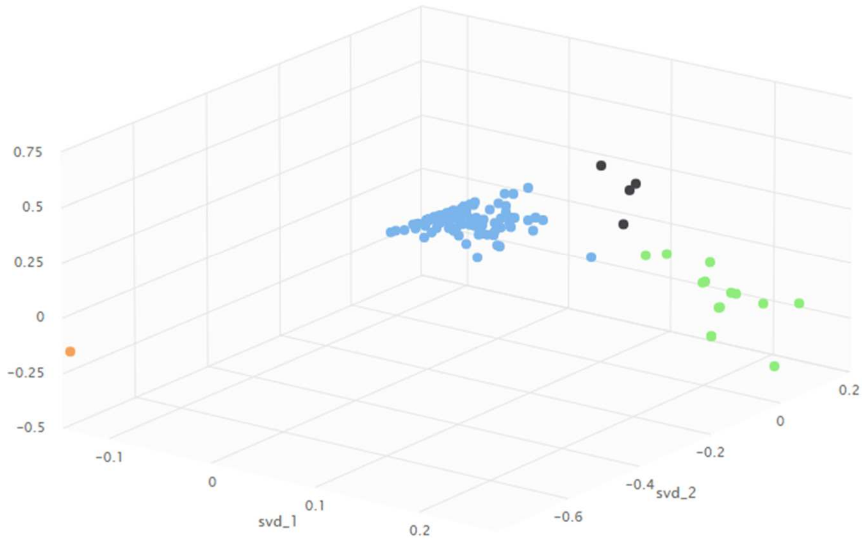

Fig. 4 (d) Scatter plot cluster K-Means++ using SVD

The scatter plot in Fig. 4 shows that the item group clusters are mapped into four models and four cluster groupings. The main difference in the distribution of the plot models using SVD showed in Figures 4(c) and 4(d), and the main difference in the distribution of the plot models not using SVD showed in Figures 4(a) and 4(b) is in the scale of feature transformation. The dimensions formed using SVD have a percentage threshold of 0.95, so the mean and variance values are closer to 0. This means that the data resulting from the SVD transformation was centered and not spread out [38]. As a result, clustering using the SVD transformation can produce cluster items with a high degree of similarity. The final centroid of the four clusters model is given in Table II. This final centroid becomes the model's cluster base center, as shown in Fig. 4.

TABLE II
FINAL CENTROID POINT

	K-Means			Kmeans & SVD			K-Means++			Kmeans++ & SVD		
Centroid	**MM**	**CM**	**R**	**SVD1**	**SVD2**	**SVD3**	**MM**	**CM**	**R**	**SVD1**	**SVD2**	**SVD3**
C_1	2.202	0.172	2.113	0.154	-0.096	0.409	2.202	0.172	2.113	0.154	-0.096	0.409
C_2	-0.371	0.609	-0.333	0.210	-0.006	-0.075	-0.371	0.609	-0.333	0.218	-0.005	-0.079
C_3	-0.438	-0.299	-0.434	-0.035	0.025	-0.002	-0.438	-0.299	-0.434	-0.034	0.012	-0.003
C_4	-0.648	-8.724	-0.655	-0.050	-0.114	-0.021	-0.648	-8.724	-0.655	-0.141	-0.786	-0.157

TABLE III
MODEL FRACTION

	K-Means		K-Means & SVD		K-Means++		Kmeans++ & SVD	
	Count	**Fraction**	**Count**	**Fraction**	**Count**	**Fraction**	**Count**	**Fraction**
Cluster 1	19	0.158	4	0.033	19	0.158	4	0.033
Cluster 2	39	0.325	15	0.125	39	0.325	14	0.117
Cluster 3	61	0.508	85	0.708	61	0.508	101	0.842
Cluster 4	1	0.008	15	0.125	1	0.008	1	0.008
Total	120		120		120		120	

The comparison of the cluster results using the K-Means and K-Means++ methods with and without SVD is shown in Table III. The table shows that cluster 1 has the same number of members and fractions for each pair of K-Means, and K-Means++ with and without the SVD. The number of members and fractions for K-Means, and K-Means++ with SVD are 19 and 0.158, respectively, while the number of members and fractions for K-Means, and K-Means++ without SVD are 4 and 0.033, respectively. Cluster 2 has the greatest number of members for K-Means, and K-Means++ without SVD while cluster 3 has the greatest number of member for K-Means, and K-Means++ with SVD. Cluster 4 has the least number of members for K-Means, and K-Means++ without SVD. In addition, the K-Means and K-Means++ without the SVD method have the same number of members and fractions in each cluster.

B. Evaluation Model

In this study, the evaluation of the clustering model was carried out using the Davies-Bouldin Index (DBI). DBI is a method of evaluating cluster performance by looking at the maximum distance between clusters and, at the same time, minimizing the distance between members in the cluster [30], [39]. The main idea in the evaluation of DBI is that the smaller the DBI number, the more optimal the results of the cluster formed [30], [34], [40]. A good DBI number is close to 0 and is positive. The results of the evaluation can be seen in Table IV.

TABLE IV
PERFORMANCE EVALUATION OF THE CLUSTER MODEL

Mean Centroid Distance	**K-Means**	**K-Means & SVD**	**K-Means++**	**Kmeans++ & SVD**
Cluster 1	0.434	0.012	0.434	0.003
Cluster 2	0.098	0.003	0.098	0.005
Cluster 3	0.073	0.003	0.073	0.001
Cluster 4	0.000	0.001	0.000	0.000
Mean	0.138	0.003	0.138	0.002
DBI	0.221	0.268	0.221	**0.141**

In Table IV, the average distance of the data to the centroid in the K-Means and K-Means++ models is the same. The DBI in the two models is also the same, so it can be concluded that the cluster formed using the two models has the same accuracy value, which is 0.221. The average distance in the model using SVD has a smaller value than the model without SVD, so it can be concluded that the cluster in the model using SVD produces low variance or has a high degree of similarity. However, in the K-Means model using SVD, the largest DBI is 0.268, which means it has a cluster accuracy level that is less than other models. Finally, the results of the K-Means++ DBI model using SVD resulted in the smallest number approaching 0, which is 0.141. So, it can be said that the cluster results using K-Means++ and SVD are more accurate models than the other three models.

IV. CONCLUSIONS

In this study, several menu engineering models are proposed. The models are based on K-Means and K-Means++ algorithms. As for the dataset transformation process, optimization was carried out using SVD. The experiment results show that optimizing the K-Means++ model with SVD resulted in the most optimal cluster. This is indicated by the average cluster distance value of 0.002 and the smallest DBI value of 0.141. Therefore, using the K-Means++ model with SVD in menu engineering analysis produces clusters containing menu items with high similarity and significant distance between groups.

Based on these results, the results of menu engineering with the proposed method can be used as suggestions and strategic recommendations for continuity in the culinary industry. For example, 4 item menus in cluster 1 have high popularity and margin similarities. The possible strategy that can be carried out in this cluster is to maintain the quality of the menu, visualization, and portions according to applicable recipe standards[12]. Thus, SMEs Managerial can carry out the result for further analysis, such as decision-making for making menu package strategies & menu prices, reshuffling menus that provide more revenue, and marketing strategy in menu sales to increase SMEs profits.

For further research, it is possible to compare the performance of the model proposed in research with similar studies. By getting the model with the best performance, the model can be implemented in management applications of menu engineering which can be used as a basis for strategic decision-making.

REFERENCES

[1] N. Setiyawati, "A Proposed Classification Method in Menu Engineering Using the K-Nearest Neighbors Algorithm," vol. 11, no. 4, 2021.

[2] H. B. J. Lai, M. R. Z. Abidin, M. Z. Hasni, M. S. A. Karim, and F. A. C. Ishak, "Key Activities of Menu Management and Analysis Performed by SME Restaurants in Malaysia," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 11, no. 4, p. 3, 2021, doi: 10.6007/IJARBSS/v11-i4/9184.

[3] J. M. Antun and C. M. Gustafson, "Menu analysis: Design, merchandising, and pricing strategies used by successful restaurants and private clubs," *J. Nutr. Recipe Menu Dev.*, vol. 3, no. 3–4, pp. 81–102, 2005, doi: 10.1300/J071v03n03_07.

[4] R. Linassi, A. Alberton, and S. V. Marinho, "Menu engineering and activity-based costing: An improved method of menu planning," *Int. J. Contemp. Hosp. Manag.*, vol. 28, no. 7, pp. 1417–1440, 2016, doi: 10.1108/IJCHM-09-2014-0438.

[5] B. M. Noone and T. A. Maier, "A decision framework for restaurant revenue management," *J. Revenue Pricing Manag.*, vol. 14, no. 4, pp. 231–244, 2015, doi: 10.1057/rpm.2015.15.

[6] S. Oktaviani, A. Sudono, and O. Sukirman, "Analisis Menu Engineering Pada Menu A ' la Carte di Grand Pasundan Convention Hotel Dalam Upaya Meningkatkan Keputusan Pembelian Melalui Suggestive Selling," vol. 4, 2008.

[7] N. A. S. Kadek, N. K. Bagiastuti, I. A. Elistyawati, and M. Sudiarta, "Menu Engineering on Main Course to Increase Sales," *Int. J. Glocal Tour.*, vol. 1, no. 1, pp. 51–60, 2020.

[8] H. Atkinson and P. Jones, "Menu Engineering: Managing the Foodsemice Micro-Marketing Mix," *J. Restaur. Foodserv. Mark.*, vol. 1, no. 1, pp. 37–55, 1993, doi: 10.1300/J061v01n01.

[9] M. L. Kasavana and D. Smith, *Menu Engineering: A Practical Guide to Menu Analysis.Lansing*. 1982.

[10] G. Rigas, "Menu Profitability Analysis Models: Linking Theory and Practice in the Greek Hospitality Context," no. December, pp. 1–62, 2018.

[11] D. V. Pavesic, "Cost/margin analysis: a third approach to menu pricing and design," *Int. J. Hosp. Manag.*, vol. 2, no. 3, pp. 127–134, 1983, doi: 10.1016/0278-4319(83)90033-6.

[12] M. Tom and K. Annaraud, "A fuzzy multi-criteria decision making model for menu engineering," *IEEE Int. Conf. Fuzzy Syst.*, 2017, doi: 10.1109/FUZZ-IEEE.2017.8015612.

[13] N. Setiyawati *et al.*, "Implementation of Two-Step Clustering Method in Menu," vol. 7, no. 2, 2020, doi: 10.25126/jtiik.202072012.

[14] S. S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," *J. Biometrika dan Kependud.*, vol. 9, no. 1, p. 62, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.

[15] M. Z. Rodriguez *et al.*, *Clustering algorithms: A comparative approach*, vol. 14, no. 1. 2019.

[16] Y. Singh and A. Mohan, "A Survey on Unsupervised Clustering Algorithm based on K-Means Clustering," *Int. J. Comput. Appl.*, vol. 156, no. 8, pp. 6–9, 2016, doi: 10.5120/ijca2016912481.

[17] K. Abhishekkumar, "Survey Report on K-Means Clustering Algorithm," *Int. J. Mod. Trends Eng. Res.*, vol. 4, no. 4, pp. 218–221, 2017, doi: 10.21884/ijmter.2017.4143.lgjzd.

[18] S. Shukla, "A Review ON K-means DATA Clustering Approach," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 17, pp. 1847–1860, 2014.

[19] A. . Jain, M. . Murty, and P. . Flynn, "Data Clustering: A Review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 265–323, 1999.

[20] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, vol. 07-09-Janu, pp. 1027–1035, 2007.

[21] H. Zarzour, Z. Al-Sharif, M. Al-Ayyoub, and Y. Jararweh, "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques," *2018 9th Int. Conf. Inf. Commun. Syst. ICICS 2018*, vol. 2018-Janua, pp. 102–106, 2018, doi: 10.1109/IACS.2018.8355449.

[22] U. Dauda and B. M. Ismail, "A study of normalization approach on K-means clustering algorithm," *Int. J. Appl. Math. Stat.*, vol. 45, no. 15, pp. 439–446, 2013.

[23] V. R. Patel and R. G. Mehta, "Performance analysis of MK-means clustering algorithm with normalization approach," *Proc. 2011 World Congr. Inf. Commun. Technol. WICT 2011*, pp. 974–979, 2011, doi: 10.1109/WICT.2011.6141380.

[24] D. Virmani, S. Taneja, and G. Malhotra, "Normalization based K means Clustering Algorithm," pp. 1–5, 2015.

[25] G. Aksu, C. O. Güzeller, and M. T. Eser, "The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model," *Int. J. Assess. Tools Educ.*, vol. 6, no. 2, pp. 170–192, 2019, doi: 10.21449/ijate.479404.

[26] A. Martinez-Millana *et al.*, "Optimisation of children z-score calculation based on new statistical techniques," *PLoS One*, vol. 13, no. 12, pp. 1–13, 2018, doi: 10.1371/journal.pone.0208362.

[27] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," 2018.

[28] N. Hidayati, M. Ihsan, and M. Danny, "Pengaruh Singular Value Decomposition Terhadap Metode – Metode Clustering," pp. 95–104, 2017.

[29] A. A. Mohamed, "An effective dimension reduction algorithm for clustering Arabic text," *Egypt. Informatics J.*, vol. 21, no. 1, pp. 1–5, 2020, doi: 10.1016/j.eij.2019.05.002.

[30] E. Muningsih, H. M. Nur, F. F. Dwi Imaniawan, Saifudin, V. R. Handayani, and F. Endiarto, "Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012101.

[31] J. Macqueen, "Some Methods For Classification And Analysis Of Multivariate Observations," 1967.

[32] D. S. Maylawati, T. Priatna, H. Sugilar, and M. A. Ramdhani, "Data science for digital culture improvement in higher education using K-means clustering and text analytics," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 4569–4580, 2020, doi: 10.11591/IJECE.V10I5.PP4569-4580.

[33] C. M. Fikri, F. E. M. Agustin, and F. Mintarsih, "Pengelompokan Kualitas Kerja Pegawai Menggunakan Algoritma K-Means++ Dan Cop-Kmeans Untuk Merencanakan Program Pemeliharaan Kesehatan Pegawai Di Pt. Pln P2B Jb Depok," *Pseudocode*, vol. 4, no. 1, pp. 9–17, 2017, doi: 10.33369/pseudocode.4.1.9-17.

[34] S. Sukamto, I. D. Id, and T. R. Angraini, "Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means," *JUITA J. Inform.*, vol. 6, no. 2, p. 137, 2018, doi: 10.30595/juita.v6i2.3172.

[35] B. Aubaidan, M. Mohd, M. Albared, and F. Author, "Comparative study of k-means and k-means++ clustering algorithms on crime domain," *J. Comput. Sci.*, vol. 10, no. 7, pp. 1197–1206, 2014, doi: 10.3844/jcssp.2014.1197.1206.

[36] Y. S.Thakare and S. B. Bagal, "Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics," *Int. J. Comput. Appl.*, vol. 110, no. 11, pp. 12–16, 2015, doi: 10.5120/19360-0929.

[37] H. L. Sari, D. Suranti, and L. N. Zulita, "Implementation of k-means clustering method for electronic learning model," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012021.

[38] I. A. Stmik and I. Gorontalo, "Penerapan Algoritma Singular Value Decomposition (Svd) Untuk Pengurangan Dimensi Pada High-Dimentional Biomedical Data Set," pp. 1–7.

[39] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1235, no. 1, pp. 6–12, 2019, doi: 10.1088/1742-6596/1235/1/012015.

[40] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.