J-Icon : Jurnal Informatika dan Komputer
Vol. 13 No. 1, March 2025, pp. 26~35
DOI: 10.35508/jicon.v13i1.20202

OPTIMIZING KINDERGARTEN SCHOOL SELECTION STATEGIES USING THE K-MEANS CLUSTERING ALGHORITHM

Nurkhasanah Fadhila Syifa[1*], Martanto[2], Arif Rinaldi Dikananda[3], and Dede Rohman[4]

[1,2,3,4]STMIK IKMI Cirebon, Jl. Perjuangan No.10b, Jawa Barat, Indonesia
Email[1*]: syifafadhila37@gmail.com
Email[2]: martantomusijo@gmail.com
Email[3]: rinaldi21crb@gmail.com
Email[4]: dede.rohman17@hotmail.com

ABSTRACT

This research aims to improve the kindergarten school location clustering model to support parents' school selection strategies. The main issue raised is the need to understand parents' preferences more deeply in choosing the right school for their children. To achieve this goal, the K-Means algorithm was applied and analyzed to cluster parents' data based on characteristics such as occupation, education, and residential location. This research utilizes a quantitative method with an exploratory descriptive approach. The results showed that the K-Means algorithm successfully formed two clusters with different characteristics. Cluster_0 includes groups with more centralized or close locations, education levels that tend to be low, and types of jobs that are at the lower middle economic level, while cluster_1 groups with more dispersed or distant locations, higher education levels, and jobs that are at higher economic levels. The quality of the resulting clusterization is considered quite good, with a Davies-Bouldin Index (DBI) value of 0.151. The findings of this study have practical implications for school administrators in developing more targeted service strategies that align with parents' needs. This research makes a significant contribution to the application of clustering techniques to support marketing strategies and decision-making in the early childhood education sector, while also assisting schools in enhancing the effectiveness of their services and marketing strategies. This research underscores the importance of using the K-Means algorithm to optimize strategic decision-making, particularly within the field of early childhood education.
Keywords: Kindergarten, Cluster Analysis, K-Means clustering.

1. INTRODUCTION

The rapid advancement of information technology has significantly impacted various aspects of human life, particularly in the field of education. Since the advent of computers, the exponential growth of digital information has accelerated dramatically. Digitalization has introduced new opportunities across numerous sectors, including early childhood education, which serves as a critical foundation for children's growth and development. Within this framework, information technology enables more efficient data management, thereby enhancing the capacity for informed and strategic decision-making [1][2].

Early childhood education, particularly kindergarten, is essential in establishing the moral, intellectual, and social foundations of children. The developmental period between the ages of 0-6 years, commonly referred to as the golden age, is characterized by heightened sensitivity to appropriate stimulation, which significantly impacts a child's long-term growth and development [3]. Despite its importance, parents often face considerable challenges in selecting an appropriate school for their children. Factors such as geographical location, financial considerations, accreditation status, available facilities, and safety measures are critical elements influencing their decision-making process [4].

Various studies have demonstrated that the K-Means algorithm is an effective method for data clustering. For instance, [5] applied this algorithm to analyse education completion rates across different regions, providing insights to support the formulation of more targeted education policies. Similarly, [6] employed K-Means to identify areas with the lowest education levels based on demographic data. However, these studies predominantly focus on secondary or higher levels of formal education and have not explored the preferences of parents in selecting early childhood education institutions. The K-Means algorithm was selected for this study due to its advantages in data clustering analysis, particularly in understanding parents' preferences when selecting schools. As a simple yet efficient clustering method, K-Means effectively groups data based on similarities such as parents' occupation, education level, and residential location. This algorithm is not only easy to implement but also capable of processing large datasets to uncover hidden patterns. By identifying distinct group characteristics, K-Means provides valuable strategic insights for

***) Corresponding Author**
Submitted : December 19, 2024
Accepted : January 18, 2025
Published : March 28, 2025

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

schools to better comprehend parents' needs. These insights enable the development of more targeted service and marketing strategies, offering an effective solution for schools to address parents' preferences.

Research on the application of clustering algorithms in the context of early childhood education remains limited. [7] demonstrated that the K-Means algorithm can be utilized to develop school promotion strategies based on data related to students' regional origins. However, their study did not specifically address parents' preferences in selecting kindergarten schools. This gap highlights the need for further research to provide educational institutions with deeper insights into parents' needs and expectations. This study aims to address this gap by applying the K-Means algorithm to cluster parental data based on employment characteristics, education level, and residential location. The aim of this study is to address the central research question: "How can the K-Means algorithm be utilized to identify parents' preferences in selecting kindergartens, and how can these insights enhance school service strategies?" Through an exploratory descriptive quantitative approach, this study provides a novel contribution to understanding parents' preference patterns. The findings are not only pertinent to the improvement of school service strategies but also serve as a foundation for the development of more targeted marketing strategies.

The study was conducted at Al Washliyah Cirebon Girang Kindergarten, a school facing challenges in understanding parents' needs. The school's administrative data, encompassing demographic and socio-economic information from 2019 to 2024, forms the foundation for analyzing parents' school selection preferences. This analysis aims to enhance the school's service quality and marketing strategies to better align with the needs of parents. In this study, the K-Means algorithm was employed to cluster data based on shared features and attributes, providing insights for improving service quality at Al Washliyah Cirebon Girang Kindergarten.

This study utilitiez the K-Means algorithm to cluster data on student parents based on the key factors they consider when selecting Al Washliyah Cirebon Girang Kindergarten. In choosing a school, particularly a kindergarten, parents typically evaluate several primary factors, including occupation, education level, and residential location. Employment status influences parents' financial capacity to pay tuition and cover additional costs, as well as their available time for dropping off and picking up their children. As a result, schools with convenient locations and flexible schedules are often prioritized. Moreover, the parents' educational background plays a significant role in their decision-making process. Parents with higher educational qualifications are more likely to choose schools that offer innovative learning programs, modern facilities, and a qualified teaching staff to nurture their children's potential. Residential location also factors into the decision, as parents generally prefer schools that are located near their home or workplace. A strategically placed school enhances mobility and increases convenience for both parents and children. These three factors occupation, education level, and location are interconnected and help parents ensure that their child's education aligns with their family's needs and socio-economic situation. The objective of this research is to identify groups of parents with different characteristics and needs, and develop a parent clustering model that can be used to enhance services at Al Washliyah Cirebon Girang Kindergarten [8]. By grouping parents based on their preferences and needs, this study aims to help kindergartens better understand the factors that parents consider when selecting a school for their children, and adapt services to better meet those needs [9].

An important aspect of an effective marketing strategy is the implementation of a good advertising system that targets the right places quickly and accurately. One data mining technique that can be used to determine marketing strategies is the clustering technique using the K-Means algorithm [10]. This research makes a significant contribution in the field of informatics by applying the K-Means clustering algorithm to group parent data in choosing a kindergarten. The novelty of this study lies in the application of the K-Means algorithm to support strategic decision-making within the early childhood education sector. The findings offer significant practical implications, particularly for schools aiming to enhance their services in accordance with the specific needs of different parent groups. Therefore, this research is not only academically valuable but also provides tangible benefits for the management of early childhood education institutions.

2. MATERIALS AND METHODS

This study employed a quantitative method with an exploratory descriptive approach. This method was selected to understand and explain the factors considered by parents in choosing Al Washliyah Cirebon Girang Kindergarten and to categorize these factors accordingly. The data collection technique used was documentation. The data consisted of primary data, including archives and administrative records of student admissions at Al Washliyah Kindergarten from 2019 to 2024. This data includes information such as the

parent identification number, student name, gender, place and date of birth, religion, parent/guardian name, parent/guardian's last level of education, and the student's home address.

In this study, the K-Means algorithm is used to cluster parent data based on employment characteristics, education level, and residential location. The algorithm was chosen due to its simplicity in cluster analysis and its effectiveness in grouping large datasets. This choice aligns with previous studies that have successfully applied the K-Means algorithm for various purposes, such as clustering student academic performance data [11] and new student admission data in secondary schools [12].

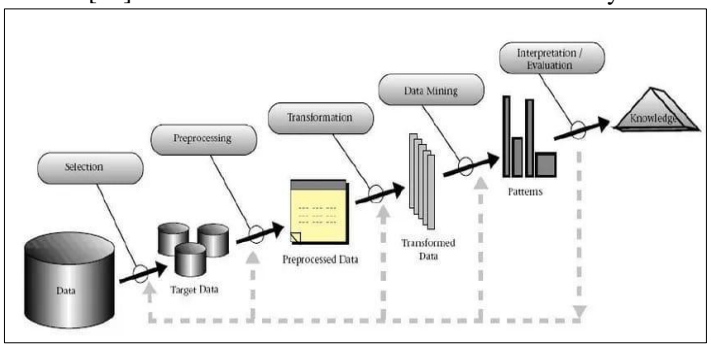

Figure 1. Research Methods

In this study using the K-Means clustering research method with a quantitative approach where the results will display statistical numbers in the form of graph visualizations, as for the explanation of the KDD stages in Figure 1, namely:

a. Data Selection
At this stage, relevant data for kindergarten school location selection will be selected. The information involves aspects of occupation, educational background, and address that parents consider in choosing a kindergarten school for their child. The importance of this data selection is that only data relevant to the research objectives will be processed.

b. Pre-Processing
The selected data was then cleaned and processed. These steps include: removing missing or inconsistent data, dealing with outlier data that affects the clustering results, and converting the data into a consistent format. Pre-processing ensures that the data is error-free and ready for processing in the next stage.

c. Data Transformation
The cleaned data is transformed into a format that is more suitable for the clustering process using the K-Means algorithm. Attributes such as occupation, education, and address were converted into numeric values to facilitate clustering. This transformation aligns with the findings by [13], which highlight the significance of data preprocessing before conducting further analysis.

d. Data Mining
The K-Means algorithm is applied to cluster parent data based on predetermined criteria. This algorithm groups the data into clusters with similar characteristics, considering attributes such as employment, education, and address. Research by [14] demonstrates that data mining techniques like K-Means can offer valuable strategic insights in the field of education.

e. Pattern Evaluation
After completing the clustering process, the identified patterns or clusters are evaluated. This evaluation is crucial for assessing the extent to which the clustering results align with the research objectives. The quality of the clusters is evaluated using the Davies-Bouldin Index (DBI). A low DBI value indicates good cluster quality, as demonstrated in the research by [5].

f. Knowledge
The final stage of the KDD process is to present the clustering results in a format that provides benefits to parents looking to select a kindergarten school for their children. The clustering results can be presented in a visual form, showing the location of kindergarten schools based on specific clusters. Each stage in the KDD process has an important role in ensuring that the resulting kindergarten school location clustering model can be effectively utilized.

This study utilizes RapidMiner software as the platform for data analysis. The selection of the K-Means algorithm over other clustering techniques is based on its simplicity, efficiency in handling large datasets, and ability to provide easily interpretable results. In comparison to algorithms such as K-Medoids or DBSCAN, K-Means is better suited for the size and distribution of the dataset employed in this research.

Furthermore, RapidMiner enhances the analysis process with its intuitive interface and integrated visualization tools.

The parameters used for the K-Means algorithm in this study include the number of clusters (k), which was set to 2 based on the evaluation of the DBI. Centroid initialization was performed randomly, and convergence criteria were established to halt iterations when the change in centroid position fell below a predefined threshold or when the maximum number of iterations was reached. This finding is consistent with previous studies, such as the research by [6], which identified that the K-Means algorithm is effective in analyzing demographic data. The use of RapidMiner ensures the reproducibility of the research, allowing competent readers to replicate the clustering process and obtain similar results. This approach is expected to offer valuable insights for schools in developing more targeted and relevant service strategies.

3. RESULTS AND DISCUSSION

The Data Selection process involves the selection and selection of data related to the factors considered by parents in choosing Al Washliyah Cirebon Girang Kindergarten. The data is then grouped into datasets that will be used in this study by applying the K-Means algorithm using Rapidminer software. The data is real data consisting of 98 records and 12 attributes and shown in Table 1.

Table 1. Dataset

No	NI	TM	NA	L/P	TTL	Agama	NO	Pekerjaan	Pend.	Alamat
1	20190223	15/04/2019	Syahdan T.F.A	L	Crb, 15-04-2014	Islam	Darta	Wiraswasta	SD	Dusun Astana
2	20190224	15/04/2019	Rizky A.H	L	Crb, 30-03-2014	Islam	Suhandi	Karyawan Swasta	SLTA	Dusun Astana
3	20190225	15/04/2019	Ghina S	P	Crb, 16-04-2014	Islam	Abdul	Wiraswasta	SLTA	Dusun Astana
…	…	…	…	…	…	…	…	…	…	…
97	20230318	04/04/2023	Muh Zidan	L	Crb, 28-08-2017	Islam	Taufik	Wiraswasta	SLTA	Perum Bukit Ciperna
98	20230319	04/04/2023	Almaira H.S	P	Crb, 01-10-2019	Islam	Eko	Karyawan Swasta	SLTA	Gg. Melati

Description:
NI : Nomor Induk
TM : Tanggal Masuk
NA : Nama Anak
NO : Nama Orang tua

The selection results from the dataset used are only 3 attributes that have the same value in each attribute as shown in Table 2.

Table 2. Data Attributes

No	Atribut	Type Data
9	Pekerjaan	Polynominal
10	Pendidikan	Polynominal
11	Alamat	Polynominal

Figure 2 depicts the data preprocessing steps undertaken to prepare the dataset for analysis using the K-Means algorithm. In the RapidMiner software, the Read Excel operator is employed to import the dataset. The initial step involves data selection using the Select Attributes operator, where only relevant attributes such as occupation, education, and address are chosen for analysis, while unrelated attributes like parent numbers, child names, and dates of birth are excluded.

Subsequently, data cleaning is performed to address missing or incomplete values and resolve inconsistencies in the dataset. Following this, data transformation is conducted by converting nominal attributes, such as occupation and education, into numerical formats using encoding techniques (Nominal to Numerical). This ensures the algorithm can accurately compute distances between data points. The data is also normalized to provide a uniform scale for all attributes, preventing those with larger value ranges from disproportionately influencing the clustering outcomes.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

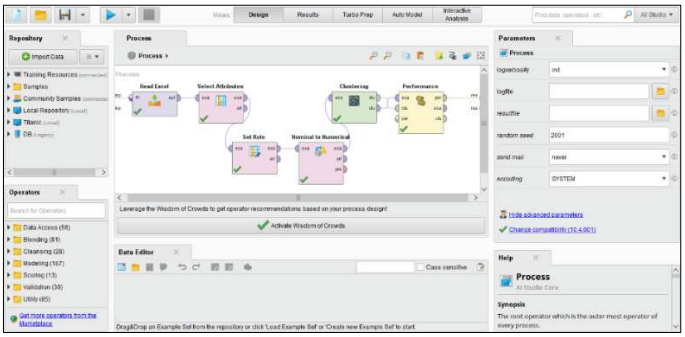

Figure 2. RapidMiner Operators

Finally, outliers are identified and managed to enhance the accuracy and effectiveness of the clustering process. By completing these pre-processing steps, the dataset is adequately prepared to generate relevant and precise clusters.

The data transformation stage using the Nominal to Numerical operator which works by replacing categorical values with numerical representations, thus facilitating calculation and analysis. The use of the Nominal to Numerical operator transforms attributes such as Occupation, Education, and Address into a numerical form that can be used in clustering analysis.

The Data Mining stage to identify groups of data with similar characteristics. This operator processes the input dataset and divides the data into the resulting clusters, based on the selected attributes.

The Performance operator is used to evaluate the effectiveness of the K-Means algorithm in dividing data into clusters, by calculating metrics such as DBI and the distance between clusters. The Performance operator facilitates the interpretation of the clustering results by displaying the results in visual and numerical formats, and preparing the data for the evaluation step.

The clustering results provide valuable insights for the management of Al Washliyah Cirebon Girang Kindergarten. By identifying the characteristics of each cluster, the school can: The school can develop targeted marketing strategies aimed at parents in Cluster 1 by emphasizing aspects such as quality education and advanced facilities. Simultaneously, efforts should be focused on ensuring accessibility and affordability to address the needs of parents in Cluster 0. Additionally, the low DBI value reflects a clear separation between clusters, ensuring reliable data groupings. These findings align with previous studies [7] that demonstrate the effectiveness of the K-Means algorithm in clustering demographic data for educational purposes (Figure 3).

Cluster Model

```
Cluster 0: 60 items
Cluster 1: 38 items
Total number of items: 98
```

Figure 3. Cluster Model Results

The results indicate that the K-Means algorithm effectively segmented the parental data into two primary clusters based on employment characteristics, educational background, and residential location. The transition from result evaluation to discussion was achieved by further analyzing the patterns identified within each cluster and their implications for school service strategies. The first cluster, consisting of 60 data points, comprises parents residing in more centralized areas, with lower educational attainment and occupations in the lower-middle economic bracket. In contrast, the second cluster, which includes 38 data points, is characterized by parents living in more dispersed locations, possessing higher educational levels, and engaged in higher-paying, more prestigious occupations (Figure 4).

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 18.352
Avg. within centroid distance_cluster_0: 13.898
Avg. within centroid distance_cluster_1: 25.385
Davies Bouldin: 0.151
```

Figure 4. Performance Vector

The evaluation of clustering quality using DBI yielded a value of 0.151, indicating that the formed clusters are of high quality. This value suggests that the data points within each cluster are well-centered,

with a distinct separation between clusters. To further explore these results, a pattern analysis was conducted to examine the specific preferences of each cluster. The analysis revealed that parents in the first cluster tended to prefer schools that were in proximity to their residential areas and offered more affordable tuition fees. In contrast, parents in the second cluster prioritized schools with modern facilities and innovative teaching methods.

This pattern offers valuable insights for schools in designing targeted service strategies. For instance, to appeal to parents in the first cluster, schools could highlight aspects such as convenience, accessibility, and affordability. In contrast, promotional strategies aimed at the second cluster could emphasize the quality of education, the facilities available, and innovative teaching methods.

These findings are consistent with previous research, such as the study by [7], which highlighted the significance of preference-based segmentation in enhancing the effectiveness of school marketing strategies. By leveraging the results of this clustering, schools can adopt a more targeted approach in attracting parents, while simultaneously improving the quality of services that cater to the specific needs of each group. Overall, this study demonstrates that the application of the K-Means algorithm can assist early childhood education institutions in gaining a deeper understanding of parents' preferences. This, in turn, provides a solid foundation for data-driven strategic decision-making, which can ultimately enhance the quality of early childhood education services.

Table 3 presents the data transformation process from nominal to numerical values conducted in this study. This transformation is designed to convert categorical data into numerical formats, enabling its use by the K-Means algorithm during the clustering process.

Table 3. Data Transformation and Cluster Distributions

Nominal Value	Numerical Value	Cluster	Description
Wiraswasta	0	Cluster_0	Close location, Low education
Karyawan Swasta	1	Cluster_1	Far location, High education
Buruh Harian Lepas	2	Cluster_0	Close location, Low education

This transformation enables the data to be processed quantitatively, allowing the K-Means algorithm to effectively identify patterns and generate optimal clusters. This study employed primary data from 98 samples, encompassing attributes such as occupation, education, and parental residence location. The nominal data was converted into numerical format using encoding techniques to facilitate processing by the K-Means algorithm. Following this transformation, the data was segmented into two primary clusters: Cluster_0 and Cluster_1. Cluster_0 comprises parents who generally reside near schools, have lower education levels, and hold lower-middle-income occupations. On the other hand, Cluster_1 consists of parents living in more dispersed areas, with higher education levels and occupations in higher economic brackets.

Table 4 presents the results of the cluster quality evaluation conducted in this study. The evaluation focuses on determining the average distance to the centroid for each cluster, providing insights into the compactness and separation of the clusters. This measure helps assess the quality of the clustering and how well the data points within each cluster are grouped.

Table 4. Evaluation of Cluster Quality

Jumlah Klaster (K)	Davies Bouldin Index (DBI)	Average Distance to The Centroid
2	0.151	Cluster_0: 13.898, Cluster_1: 25.385
3	0.184	More diffuse distribution without optimal focus

The evaluation results using DBI indicate that the two-cluster solution offers the best clustering quality, with a DBI value of 0.151. This value signifies a substantial inter-cluster distance and strong internal consistency within each cluster. In comparison to the alternatives of three or more clusters, the two-cluster configuration yielded a more centralized data distribution, aligning better with the research objectives.

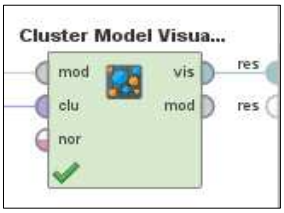

Figure 5. Operator Cluster Model Visualizer

Figure 5 shows the Cluster Model Visualizer operator in RapidMiner software. The cluster model visualizer has an important role in the process of evaluating and analyzing clustering models because it is able to provide a clear visual picture of the separation and distribution patterns of data in each cluster.

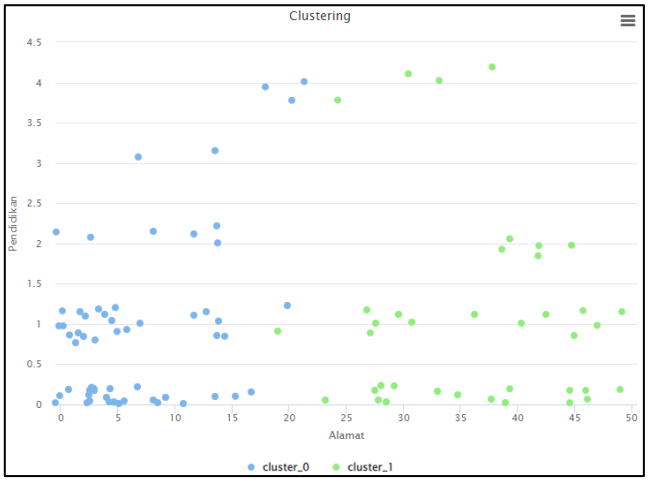

Figure 6. Scatter Plot

Figure 6 displays a Scatter Plot showing the clustering results using the K-Means algorithm, which visualizes the distribution of the two clusters based on the address attribute on the x-axis and education on the y-axis. The data in Cluster 0 is shown in blue and consists of 60 data points, while the data in Cluster 1 is shown in red and includes 38 data points. In Cluster 0, most of the data points are concentrated at lower address values, indicating a closer or more centralized location, with education levels tending to be lower. In contrast, in Cluster 1, the data points are scattered at higher address values, indicating a more distant or dispersed location, with relatively more varied and higher levels of education compared to Cluster 0. This visualization reinforces the clustering results, where Cluster 0 depicts groups with more centralized locations and lower levels of education, while Cluster 1 depicts groups with more distant locations and higher levels of education.

Table 5 presents the number of members and the corresponding values for each cluster derived from the cluster analysis. This information is instrumental in identifying specific patterns and characteristics of each cluster, offering a clearer understanding of the data distribution within the clusters.

Table 5. Value and Number of Members in each Cluster

Cluster	Number of Members of each Cluster	Values of each Cluster Member
Cluster_0	60	20190223 20190224 20190225 20190226 20190227 … 20210272 20220280 20220281 20220282 20220285 20220286 20220288 20220292

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

Cluster_1	38	20220292 20220296 20220297 20210273 20210274 20210275 20210276 20210277 … 20220279 20220283 20220284 20220287 20220289 20220290 20220291 20220293 20220294 20220295 20220297 … 20230319

Overall, this table highlights the unique characteristics of each cluster, providing valuable insights that can be utilized to formulate more targeted strategies or recommendations as required.

The DBI is employed to evaluate the quality of clustering by calculating the average ratio of intra-cluster distances to inter-cluster distances (Equation 1). A lower DBI value indicates better clustering results, reflecting more compact and well-separated clusters.

$$DBI = \frac{1}{N}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{S_i+S_j}{d_{i,j}}\right) \quad (1)$$

Where:

S_i (The average distance of all data points in cluster i to its centroid j)
S_j (The average distance of all data points in cluster j to its centroid j)
$d_{i,j}$ (The distance between the centroid of cluster i and the centroid of cluster j)
N (The number of clusters)
R_{ij} adalah distance of centroid cluster i and j

$$R_{ij=}\left(\frac{S_i+S_j}{d_{i,j}}\right) \quad (2)$$

Given (Equation 2):

S_i (The average distance of data points to the centroid of cluster 1) = 13.898
S_j (The average distance of data points to the centroid of cluster 2) = 25.385
$d_{i,j}$ (The distance between the centroids of cluster 1 and cluster 2) = 260.152

$$R_{i,j} = \left(\frac{13.898 + 25.385}{260.152}\right)$$
$$R_{i,j} = \left(\frac{39.283}{260.152}\right)$$
$$R_{i,j} = 0.151$$
$$DBI = \frac{1}{2}(0.151 + 0.151)$$
$$DBI = 0.151$$

DBI value for 2 clusters is 0.151, the smallest compared to other cluster configurations. This indicates that clustering with 2 clusters provides the most optimal results, as it reflects better cluster compactness and separation. The average distance of the data points to the centroid of each cluster is relatively small. A smaller DBI value, approaching 0, indicates better clustering results, as it reflects tighter cluster cohesion and greater separation between clusters.

This analysis shows that parents with lower educational backgrounds and irregular employment tend to choose schools that are easily accessible, while parents with higher education prefer schools that excel in terms of quality and facilities. Furthermore, the findings of this study are supported by [5], who found

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

that clustering can help identify educational needs suited to specific demographic groups. According to the research conducted by [7], the application of the K-Means algorithm is effective in clustering prospective students based on their region of origin, providing valuable insights for developing more targeted promotional strategies. This is further confirmed by [15], who asserted that a K-Means clustering-based approach can map the socio-economic needs in education.

This is related to research by [10] which shows that parents' occupation and education have an influence on school selection. The findings of this study have significant implications for the management of Al Washliyah Cirebon Girang Kindergarten. By understanding the preferences of each cluster of parents, schools can design more targeted marketing strategies and appropriate services to attract groups of parents. DBI with a relatively low value indicates that the preferences in each cluster are quite consistent so that they can implement a more focused strategy according to the needs of the group. Overall, the research shows that K-Means is effective in understanding educational needs based on demographic data, which can help educational institutions to improve services and enrollment strategies more precisely and measurably.

4. CONCLUSION AND SUGGESTIONS

Based on the research results regarding the application of the K-Means algorithm in clustering parents' data in choosing kindergarten schools, it can be concluded that in this clustering process, parents' data is organized into groups based on several relevant characteristics, such as occupation, education, and residential location. This implementation resulted in two distinct clusters, with each cluster exhibiting specific characteristics that reflect the parents' preference patterns. the application of this algorithm was successful in supporting the parents' school selection strategy. This is shown by DBI value of 0.151, indicating a clear separation between clusters as well as high consistency within each cluster. This analysis shows that the use of two clusters is the most effective in grouping parents based on relevant characteristics, so that the clustering results can be used as a basis for developing a more targeted and appropriate service strategy.

Despite the robustness of the K-Means algorithm, this study has some limitations. The clustering process is sensitive to outliers, which could potentially impact the results. Furthermore, the algorithm assumes spherical clusters, which may restrict its effectiveness when applied to datasets with irregular cluster shapes. Future research should consider exploring alternative algorithms, such as K-Medoids or DBSCAN, which can address the sensitivity to outliers and better capture clusters with varying shapes and densities. Expanding the dataset to include additional attributes, such as parents' income levels or children's prior education, could further refine the clustering process. Testing these approaches would provide deeper insights and help validate the findings of this study in a wider range of contexts.

REFERENCE

[1] M. B. Fajri and S. D. Purnamasari, "Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering," *J. Inf. Technol. Ampera*, vol. 3, no. 3, pp. 317–334, 2022, [Online]. Available: www.journal-computing.org/index.php/journal-ita/article/view/315.

[2] Y. Wahyudin and D. N. Rahayu, "Analisis Metode Pengembangan Sistem Informasi Berbasis Website: A Literatur Review," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 15, no. 3, pp. 26–40, 2020, doi: 10.35969/interkom.v15i3.74.

[3] L. F. Az Zahroh, N. Rahaningsih, and R. D. Dana, "Klasterisasi Data Kegemaran Membaca Menggunakan Algoritma K-Means Di Sma Al-Islam Cirebon," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2692–2698, 2024, doi: 10.36040/jati.v8i3.9543.

[4] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.

[5] A. T. Basalamah and R. Setyadi, "Penerapan Algoritma K-Means Clustering Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia," *J. Inform. dan Teknol. Komput.*, vol. 4, no. 2, pp. 114–121, 2023, [Online]. Available: ejurnalunsam.id/index.php/jicom/article/view/7893

[6] Nurahman and D. D. Aulia, "Klasterisasi Pendidikan Masyarakat untuk mengetahui Daerah dengan Pendidikan Terendah menggunakan Algoritma K-Means," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 5, no. 1, pp. 38–44, 2023.[Online]. Available: https://publikasiilmiah.unwahas.ac.id/JINRPL/article/view/7510.

[7] K. L. Putri, M. Agus Sunandar, and Y. Muhyidin, "Penerapan Algoritma K-Means untuk Penentuan Strategi Promosi(Studi Kasus: SMA PGRI 1 Purwakarta)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7,

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

no. 3, pp. 2044–2050, 2023, doi: 10.36040/jati.v7i3.7030.
[8] S. Haviyola, S. Susilawati, and M. Jajuli, "Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2786–2791, 2024, doi: 10.36040/jati.v7i4.7200.
[9] J. P. Tanjung, B. A. Wijaya, and M. Ridho, "Implementasi Algoritma K-Means Dalam Mengukur Tingkat Kepuasan Siswa Terhadap Pelayanan Pada SMA Swasta Bani Adam AS," *Data Sci. Indones.*, vol. 3, no. 1, pp. 1–11, 2023, doi: 10.47709/dsi.v3i1.2775.
[10] D. Nurdiansyah, S. Saidah, and N. Cahyani, "Clustering analysis for grouping sub-districts in Bojonegoro District with the K-Means method with a variety of approaches," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 1081–1092, 2023, doi: 10.30598/barekengvol17iss2pp1081-1092.
[11] P. S. Hasugian and J. S. Raphita, "Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 3, pp. 262–268, 2022.
[12] S. Suiroh, R. Astuti, and F. Muhamad Basysyar, "Implementasi Algoritma K-Means Pada Pengelompokan Data Penerimaan Peserta Didik Baru Di Smkn 1 Balongan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 1–8, 2024, doi: 10.36040/jati.v8i1.8335.
[13] Z. Miftah and F. Fahrurrozi, "Digitalisasi dan Disparitas Pendidikan di Sekolah Dasar," *Ibtida*, vol. 3, no. 02, pp. 149–163, 2022, doi: 10.37850/ibtida.v3i02.361.
[14] D. Fitamaya and M. N. Zulfahmi, "Analisis Pembelajaran Moral dan Agama Melalui Penerapan Loose Part," *J. Ris. dan Inov. Pembelajaran*, vol. 4, no. 1, pp. 193–207, 2024, doi: 10.51574/jrip.v4i1.1362.
[15] N. Yannuansa, M. Safa'udin, and M. I. Aziz, "Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Hasil Belajar Terhadap Pendapatan Orang Tua Pada Mata Pelajaran Produktif," *J. Tecnoscienza*, vol. 6, no. 1, pp. 43–55, 2021, doi: 10.51158/tecnoscienza.v6i1.530.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)