Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Clustering Analysis of Administrative Service Types Using K-Means (Study Case: Village bojongsalam)

Wafiq Azizah[1*], Ade Irma Purnamasari[2], Agus Bahtiar[3], Kaslani[4]

[1,2,3,4]STMIK IKMI Cirebon

wafiiqazizahh@gmail.com[1*],irma2974@yahoo.com[2], agusbahtiar038@gmail.com[3], kaslani123@gmail.com[4]

Abstract

Advances in information technology present significant opportunities for the improvement of public services, especially in relation to the administrative functions of Bojongsalam Village. Reliance on traditional methods often leads to inefficiencies and inaccuracies in administrative processes. This research uses the K-Means algorithm to categorize administrative service data based on service type, document number, printing date, and accompanying remarks. Utilizing the Knowledge Discovery in Databases (KDD) framework, the analysis includes data selection, pre-processing, transformation, and clustering analysis conducted through RapidMiner software. The dataset consisted of 718 administrative records that had undergone a rigorous cleaning process, including attribute normalization. The analysis resulted in an optimal Davies-Bouldin Index (DBI) value of -0.498 at K = 4, with each cluster representing a different service utilization pattern. The issuance of Family Cards (KK) and Birth Certificates showed higher demand compared to other available services. This classification promotes workload optimization, fair resource allocation, and formulation of effective operational strategies. The application of the K-Means algorithm demonstrated its effectiveness in data clustering and made a significant contribution to technology-based administrative management. The findings lay a basic framework for addressing the needs of the community in a timely manner.

Keywords: *K-Means Clustering Method, Classification, Governance Services, Bojongsalam Community, Effectiveness.*

1. Introduction

The development of information technology has changed various aspects of life, including the public service sector. In Kelurahan Bojongsalam, administrative services are still carried out manually, which causes slow processes and often long lines. The K-Means algorithm offers a solution to group types of services based on community demand patterns, so that the service process can be more efficient and targeted [1]

The application of K-Means algorithm in various sectors has shown positive results. For example, research by [2] proved that this algorithm is effective in grouping patients based on their medical history to facilitate the organization of health services. However, the application of K-Means in the public administration sector, especially at the neighborhood level, is still minimal. This suggests an opportunity to explore the potential of K-Means in improving the efficiency of public services.

This research aims to implement the K-Means algorithm to categorize the types of administrative services in Kelurahan Bojongsalam, based on community demand patterns. Hopefully, the results of this research can help speed up the service process, reduce waiting time, and increase community satisfaction. In addition, this research is also expected to be a reference in developing data-based public service models and computing technology in other areas [3], [1].

2. Literature

2.1. Data Mining

Data mining is the process of analyzing data to discover hidden patterns and useful information from large data sets. This process is often used to help make better decisions in various fields, such as business, education, and government. In the academic context, research shows that data mining can improve the effectiveness of data analysis, for example in the evaluation of student learning outcomes [4]. This process involves data exploration techniques, such as clustering, classification, and prediction, which are applied to complex and diverse data.

2.2. Clustering

Clustering is a technique in data mining that aims to group data into several clusters based on similar characteristics. This technique is useful for segmentation analysis, both in health, education, and business. Previous studies revealed that clustering can be used to categorize data such as employee performance, nutrition levels, or teachers' level of understanding of learning methods [5]. With this method, data with similar properties can be grouped for further analysis.

2.3. Algoritma K-Means

K-Means algorithm is one of the most popular clustering methods due to its simplicity and efficiency in processing large data. This algorithm works by dividing data into clusters based on the closeness of the distance between data. The implementation of K-Means has been applied in various studies, such as clustering unemployment based on age [1] and clustering stock items in retail stores [6]. The main advantage of this algorithm is its ability to handle large datasets with a good level of accuracy.

3. Research methods

The research method is a stage carried out to collect and analyze data systematically with the aim of achieving the expected target. This research uses quantitative methods with the K-Means Clustering algorithm approach. Data analysis techniques refer to the Knowledge Discovery in Database (KDD) framework to ensure the process is carried out in an organized manner.

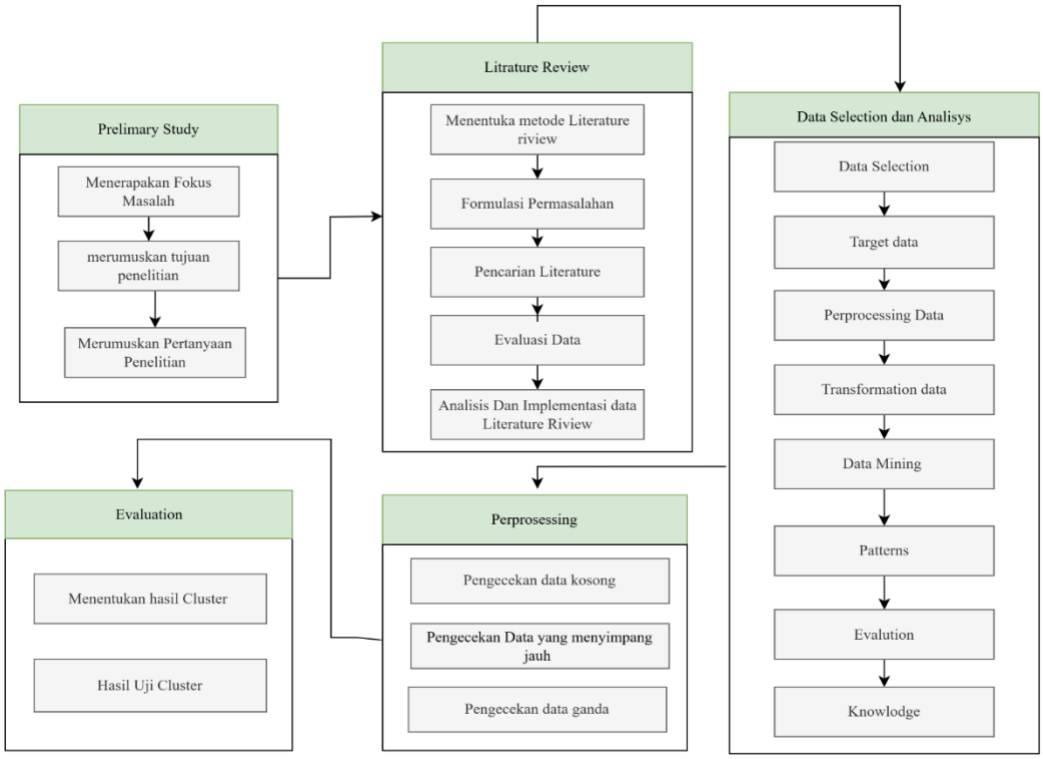

Figure 1: Research method.

Table 1: Research methods.

Stages	Activities	Activity Description
Prelimary Study	Defining the Problem Focus	Determining specific aspects or topics of administrative service types.
	Formulating Research Objectives	Establishing the main objectives of the research, such as understanding the types of services in Bojongsalam Village.
	Formulating Research Questions	Crafting specific and relevant research questions.
Literature Review	Determining Literature Review Methods	Selecting methods to review relevant literature.
	Problem Formulation	Identifying and formulating problems based on literature review.
	Literature Search	Searching for relevant articles and journals related to the use of the K-Means algorithm.
	Data Evaluation	Evaluating the quality and relevance of data from the reviewed literature.
	Data Analysis & Interpretation from Literature Review	Analyzing data to build a strong research foundation.
Data Selection	Identifying and Selecting Appropriate Data	The data used in this research includes administrative service data in Bojongsalam Village, such as service types, document numbers, print dates, and descriptions.

Data Preprocessing	Data Cleaning for Analysis	Removing missing or inconsistent attributes to ensure high data quality.
Data Transformation	Preparing Data for Analysis	Converting selected data to suit the data mining process.
Data Mining	Applying K-Means Clustering Algorithm	Performing data clustering using the K-Means algorithm with the help of RapidMiner software.
Evaluation	Evaluating Results	Assessing the relevance and accuracy of clustering results.
Perprossesing	Checking for Missing Data	Ensuring the dataset does not contain missing values that could interfere with the analysis. Missing data is handled through deletion or imputation (mean, median, mode).
	Checking for Outliers	Identifying outliers that may affect analysis results using the K-Means method.
	Checking for Duplicate Data	Removing duplicate entries to reduce bias in the analysis.
Evaluation	Determining Cluster Results	Identifying members and patterns of each cluster based on clustering results.
	Cluster Quality Assessment	Evaluating cluster quality using metrics such as the Davies-Bouldin Index (DBI).

4. Results and Discussion

4.1. Research Result

The data obtained are the results of the use of administrative services by the community in Bojongsalam Village. This data is then selected to obtain relevant attributes. In this study, the attributes selected include service type, document number, month, printing, description. After attribute selection, the data is grouped based on the description column, document number, printing date. From the data selection process, there are 718 user data that utilize administrative services, resulting in a dataset as in table 2.

Table 2: Data

NO	TYPE OF SERVICE	DOCUMENT NUMBER	PRINTING DATE	MONTHS	DESCRIPTION	YEAR
1	FAMILY CARDS	3204282403050817	02-01-2023	JANUARY	Not yet printed	2023
2	FAMILY CARDS	3204280609120119	02-01-2023	JANUARY	Not yet printed	2023
3	FAMILY CARDS	3204282605120015	02-01-2023	JANUARY	Not yet printed	2023
4	CHILD IDENTITY CARD	3204287001180001	10-01-2023	JANUARY	Not yet printed	2023
5	FAMILY CARDS	3204280401230010	10-01-2023	JANUARY	Not yet printed	2023
6	FAMILY CARDS	3204282003180015	10-01-2023	JANUARY	Not yet printed	2023
7	FAMILY CARDS	3204280401230010	10-01-2023	JANUARY	Not yet printed	2023
8	FAMILY CARDS	3204282508140022	10-01-2023	JANUARY	Not yet printed	2023
9	FAMILY CARDS	3204282211160020	10-01-2023	JANUARY	Not yet printed	2023
10	FAMILY CARDS	3204282903190012	10-01-2023	JANUARY	Not yet printed	2023
11	CHILD IDENTITY CARD	3204280905140001	20-01-2023	JANUARY	Not yet printed	2023
12	FAMILY CARDS	3204282508140022	20-01-2023	JANUARY	Not yet printed	2023
13	CHILD IDENTITY CARD	3204280905140001	20-01-2023	JANUARY	Not yet printed	2023
14	CHILD IDENTITY CARD	3204280905140001	20-01-2023	JANUARY	Not yet printed	2023
15	CHILD IDENTITY CARD	3204282708180001	20-01-2023	JANUARY	Not yet printed	2023
16	CHILD IDENTITY CARD	3204285403180004	20-01-2023	JANUARY	Not yet printed	2023
17	CHILD IDENTITY CARD	3204280905140001	20-01-2023	JANUARY	Not yet printed	2023
18	CHILD IDENTITY CARD	3204280905140001	24-01-2023	JANUARY	Not yet printed	2023
19	FAMILY CARDS	3204280401230010	26-01-2023	JANUARY	Not yet printed	2023
...						
...						
700	FAMILY CARDS	3204251505180013	29-12-2023	DECEMBER	already printed	2023
701	FAMILY CARDS	3204282505110927	29-12-2023	DECEMBER	already printed	2023
702	FAMILY CARDS	3204282712230007	29-12-2023	DECEMBER	already printed	2023
703	FAMILY CARDS	3204280508150010	29-12-2023	DECEMBER	already printed	2023
704	FAMILY CARDS	3204282505160018	29-12-2023	DECEMBER	already printed	2023
705	FAMILY CARDS	3204282003070033	29-12-2023	DECEMBER	already printed	2023
706	FAMILY CARDS	3204281010220005	29-12-2023	DECEMBER	already printed	2023
707	FAMILY CARDS	3204282112150013	29-12-2023	DECEMBER	already printed	2023
708	DEATH CERTIFICATE	3204-KM-28122023-0001	29-12-2023	DECEMBER	already printed	2023
709	FAMILY CARDS	3204281902140025	29-12-2023	DECEMBER	already printed	2023
710	DEATH CERTIFICATE	3204-KM-27122023-0004	29-12-2023	DECEMBER	already printed	2023
711	DEATH CERTIFICATE	3204-KM-28122023-0001	29-12-2023	DECEMBER	already printed	2023
712	BIRTH CERTIFICATE	3204-LT-29122023-0090	29-12-2023	DECEMBER	already printed	2023
713	FAMILY CARDS	3204282603050594	29-12-2023	DECEMBER	already printed	2023

4.2. Pre-Processing

The data cleaning process includes checking the dataset to ensure there is no empty data (Missing Value) and removing duplicate data (Double Data) contained in the database.

a. Empty Data Checking
Checking empty data (Missing Value) in the dataset by looking at the existing columns. From the results of this check, it can be seen that there is no empty data in this dataset.

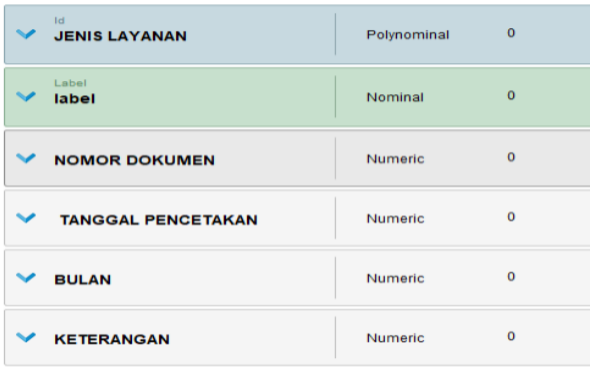

Figure 2: Display of empty data checking results

b. Checking Data that deviates far
Checking the minimum and maximum data values of each attribute in the dataset. The image shown shows that a ttributes such as service type, month, document number, printing date, and description have appropriate minimum and maximum values and can be used for further analysis.

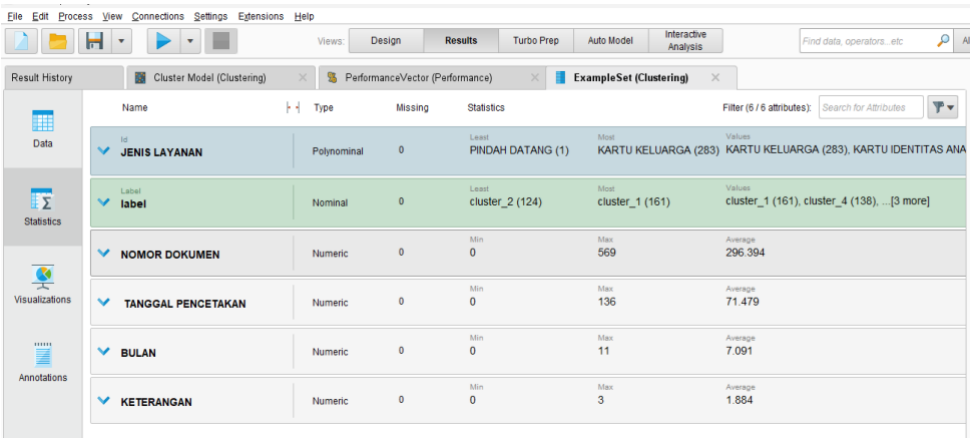

Figure 3: Result Display of Data Checking Results that Deviate Significantly

c. Double Data Checking
After the Delete Duplicates process has been run, it can be concluded that out of the total

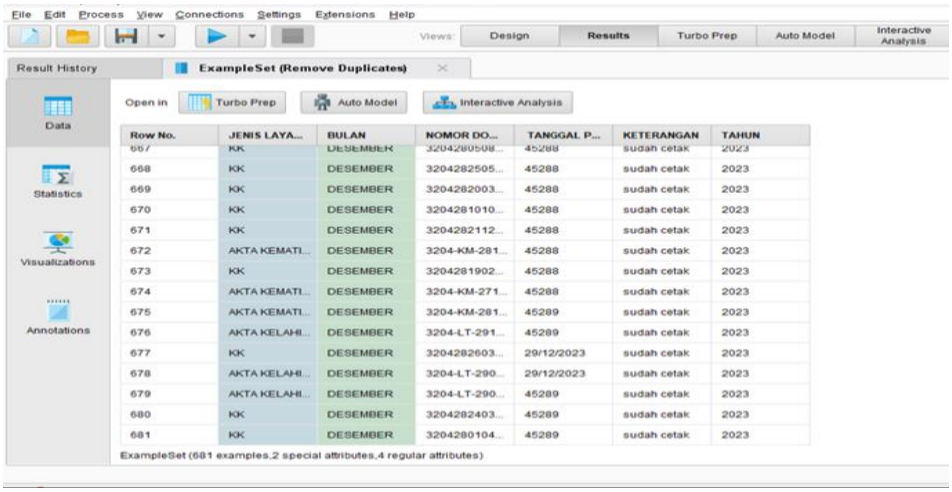

Figure 4: Result Display of Data Checking Results that Deviate Significantly

The next step after pre-processing is the attribute selection process to select relevant features from the dataset. The attribute selection process is done to select relevant features from the dataset. This attribute selection process is done using the Select Attributes Operator, which helps the author select important features.

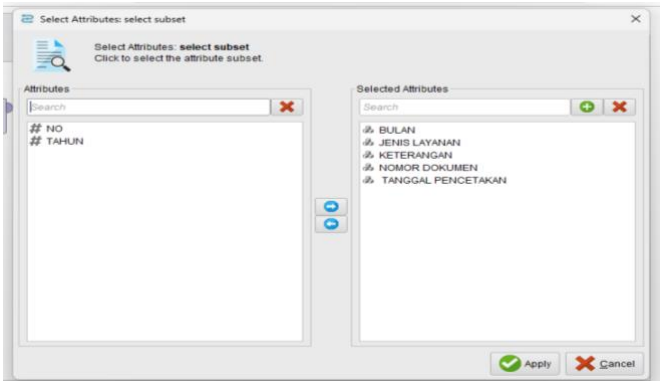

Figure 5: Display of Select Attributes Results

The Nominal to Numerical process is carried out to convert categorical attributes in administrative service data in Bojongsalam Village into numerical data. For example, categories such as “kk”, “KTP”, “Birth Certificate” are converted into numerical values to make it easier to process in further clustering analysis.

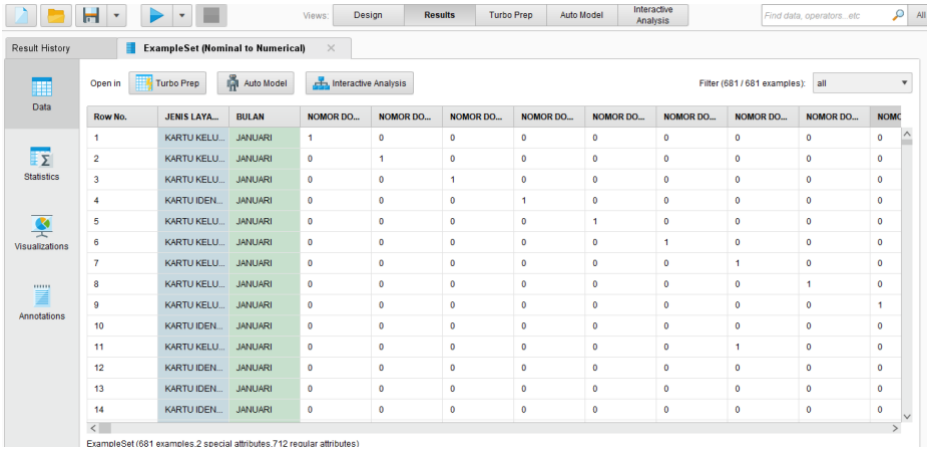

Figure 6: Result Display of Nominal to Numerical

This research process uses the K-Means Algorithm to group administrative service data in Bojongsalam Village based on relevant data characteristics. In this research, the author utilizes RapidMiner software to apply the K-Means Algorithm.

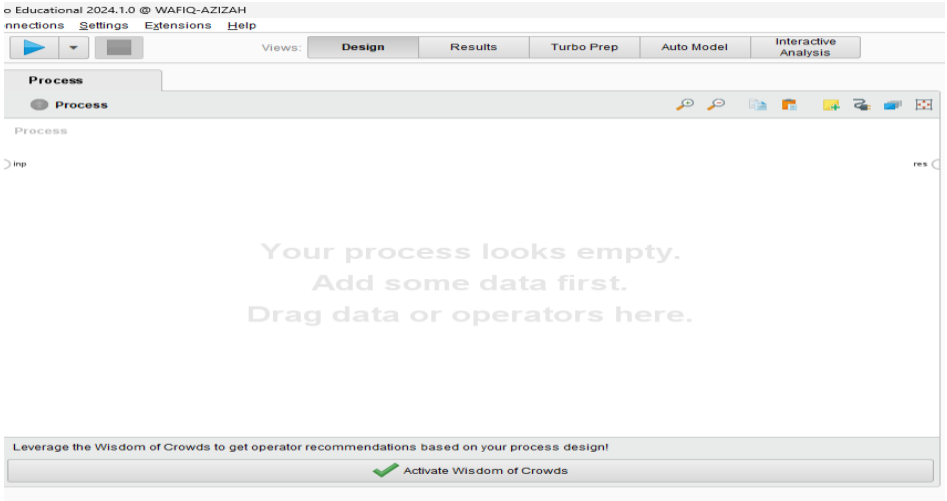

Figure 7: Rapidminer page

In Rapidminer, the process of creating a cluster model uses the K-Means operator. Furthermore, to generate model performance, the author conducted a test with cluster distance performance. This evaluation process is done by testing the model from cluster 2 to 10, to determine the optimal number of clusters.

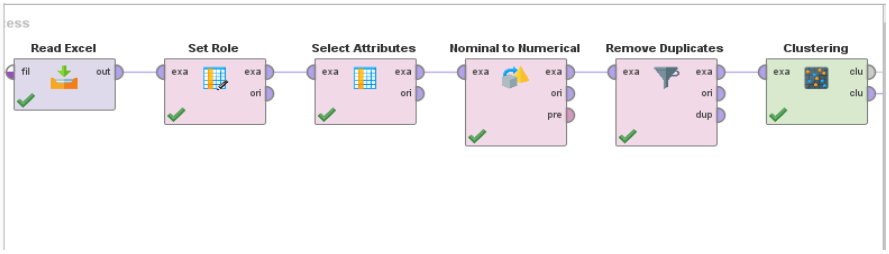

Figure 8: Cluster Model

4.3. Model Evaluation

Model evaluation in this study was carried out to find the optimal cluster in the modeling process. This process aims to get the best cluster by using the Davies Bouldin Index (DBI) evaluation, where the smallest DBI value shows the best evaluation results. The results of the applied K-Means model include testing from the value of K = 4 to K = 4 can be seen and evaluated the performance value.

Table 3: Cluster Test Results (K=2 - K 10)

Cluster	Cluster member	DBI value
2	Cluster 0: 362 items Cluster 1: 319 items Total number of items: 681	-0.503
3	Cluster 0: 230 items Cluster 1: 220 items Cluster 2: 231 items Total number of items: 681	-0.499
4	Cluster 0: 163 items Cluster 1: 184 items Cluster 2: 174 items Cluster 3: 160 items Total number of items: 681	-0.498
5	Cluster 0: 131 items Cluster 1: 161 items Cluster 2: 124 items Cluster 3: 127 items Cluster 4: 138 items Total number of items: 681	-0.520
6	luster 0: 140 items Cluster 1: 106 items Cluster 2: 115 items Cluster 3: 106 items Cluster 4: 107 items Cluster 5: 107 items Total number of items: 681	-0.518
7	Cluster 0: 104 items Cluster 1: 93 items Cluster 2: 97 items Cluster 3: 91 items Cluster 4: 106 items Cluster 5: 102 items Cluster 6: 88 items Total number of items: 681	-0.531
8	Cluster 0: 80 items Cluster 1: 91 items Cluster 2: 79 items Cluster 3: 84 items Cluster 4: 77 items Cluster 5: 90 items Cluster 6: 88 items Cluster 7: 92 items Total number of items: 681	-0.517
9	Cluster 0: 65 items Cluster 1: 82 items Cluster 2: 79 items Cluster 3: 73 items Cluster 4: 91 items Cluster 5: 90 items Cluster 6: 65 items Cluster 7: 65 items Cluster 8: 71 items Total number of items: 681	-0.520

10	Cluster 0: 76 items Cluster 1: 63 items Cluster 2: 89 items Cluster 3: 60 items Cluster 4: 69 items Cluster 5: 67 items Cluster 6: 68 items Cluster 7: 61 items Cluster 8: 68 items Cluster 9: 60 items Total number of items: 681	-0.516

So the optimal dbi value is at k = 4 with the following cluster members:

a. David Bouldin Index (DBI)
Based on table 4.2 shows that the best cluster value, which has the lowest DBI value or close to 0, is in cluster 4 with a DBI value of -0.498 which consists of Cluster_0: 163 items, Cluster_1: 1 184 items, Cluster _2: 174 items, Cluster_3: 160 items. The cluster model results can be seen in Figure 9.

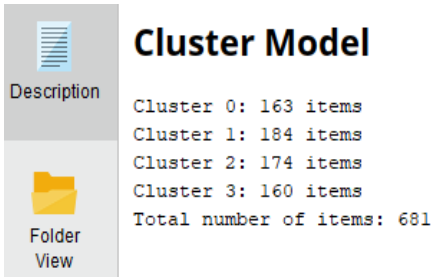

Figure 9: Description K-Means Cluster Model.

4.4. Scatter plot

The figure below illustrates the results of applying the K-Means clustering algorithm to administrative services in Bojongsalam Village. The horizontal axis represents the status of the service (not printed, printed, double), while the vertical axis depicts the different types of services (e.g. Family Card, Birth Certificate).

Data is categorized according to service status and type, with green indicating that the service has been printed. This configuration facilitates the identification of services that require priority, especially those with a large amount of data pending printing, thus improving the efficiency of administrative operations.

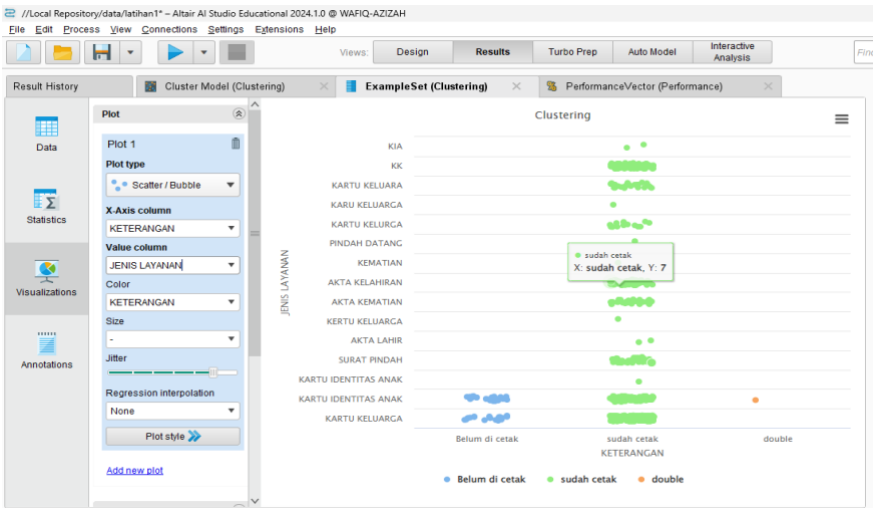

Figure 10: Scatter plot

4.5. Knowledge

At this stage will discuss the interpretation and evaluation of the results that have been obtained in the previous stages, here researchers will discuss in more detail the results of the DBI value of the cluster test value K = 2 to K = 10, and discuss each cluster based on characteristics.

The DBI results table can be seen below.

Table 4: DBI Result

Cluster	David Bouldin Index (DBI)
2	-0.503
3	-0.499
4	-0.498
5	-0.520
6	-0.518
7	-0.531
8	-0.517
9	-0.520
10	-0.516

5. Conclusions and Suggestions

5.1. Conclusions

This research proves that the K-Means algorithm effectively improves the efficiency of administrative services in Bojongsalam Village, with an optimal Davies-Bouldin Index (DBI) value of -0.498 at K=4. The algorithm is able to cluster data based on utilization patterns, identify community needs, and support workload distribution, allocation of additional resources, and automation of repetitive processes. The three main factors that affect efficiency are utilization frequency, service duration, and procedural complexity. Clustering also helps determine priority services, such as the issuance of Family Cards (KK) and ID cards, which require more attention, especially when workload is high. The findings provide strategic recommendations, including resource optimization, technology utilization, and staff competency improvement, which significantly improve operational efficiency and community satisfaction.

5.2. Suggestions

The suggestions given are:

1. Further research can explore the implementation of the K-Means algorithm with more varied parameters to improve operational efficiency, especially in administrative services with dynamic demand patterns.
2. It is recommended to utilize other algorithms or hybrid methods to compare analysis results, as well as evaluate factors such as frequency, duration, and complexity of procedures in more depth.
3. Future research could focus on developing automation systems based on clustering results, as well as prioritizing staff training and resource allocation for high-demand services.

References

[1] Andi Diah Kuswanto, Azumardi Nabil Fadhila, Paulus Tri Setiawan, Muhammad Kevin Setiawan, and Dody Renal Syahputra, "Penerapan K-Means Clustering Untuk Menentukan Jumlah Pengangguran Berdasarkan Umur," *Repeater Publ. Tek. Inform. dan Jar.*, vol. 2, no. 3, pp. 135–146, 2024, doi: 10.62951/repeater.v2i3.116.

[2] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.

[3] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, p. 13, 2021, doi: 10.30872/jim.v16i1.5182.

[4] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[5] G. B. Kaligis and S. Yulianto, "Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai," *IT-Explore J. Penerapan Teknol. Inf. dan Komun.*, vol. 1, no. 3, pp. 179–193, 2022, doi: 10.24246/itexplore.v1i3.2022.pp179-193.

[6] F. Rizki Ani, R. Kurniawan, and T. Suprapti, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Persediaan Stok Sepatu Di Toko Diomclothing," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3699–3707, 2024, doi: 10.36040/jati.v7i6.8233.